

Archimedean copula estimation using Bayesian splines smoothing techniques.

Philippe Lambert ^{a,*}

^a*Institut des sciences humaines et sociales,
Université de Liège, B-4000 Liège, Belgium*

Abstract

Copulas enable to specify multivariate distributions with given marginals. Various parametric proposals were made in the literature for these quantities, mainly in the bivariate case. They can be systematically derived from multivariate distributions with known marginals, yielding e.g. the normal and the Student copulas. Alternatively, one can restrict his/her interest to a sub-family of copulas named *Archimedean*. They are characterized by a strictly decreasing convex function on $(0,1)$ which tends to $+\infty$ at 0 (when *strict*) and which is 0 at 1. A ratio approximation of the generator and of its first derivative using B-splines is proposed and the associated parameters estimated using Markov chains Monte-Carlo methods. The estimation is reasonably quick. The fitted generator is smooth and parametric. The generated chain(s) can be used to build “credible envelopes” for the above ratio function and derived quantities such as Kendall’s tau, posterior predictive probabilities, etc. Parameters associated to parametric models for the marginals can be estimated jointly with the copula parameters. This is an interesting alternative to the popular two-steps procedure which assumes that the regression parameters are fixed known quantities when it comes to copula parameter(s) estimation. A simulation study is performed to evaluate the approach. The practical utility of the method is illustrated by a basic analysis of the dependence structure underlying the diastolic and the systolic blood pressures in male subjects.

Key words: Archimedean copula, Bayesian P-splines, Markov chains Monte carlo, Monotonicity and convexity constraints.

* Corresponding address: Institut des sciences humaines et sociales, Université de Liège, Boulevard du Rectorat 7 (Bat. B31), B-4000 Liège, Belgium. Tel.: +32-4-366.59.90 ; Fax: +32-4-366.47.51.

Email address: p.lambert@ulg.ac.be (Philippe Lambert).

1 Introduction

A n -variate copula $C(u_1, \dots, u_n)$ is a multivariate distribution with uniform margins on $(0, 1)$. A n -variate distribution function $H(y_1, \dots, y_n)$ with given univariate margins $F_i(y_i)$ can be derived from it:

$$H(y_1, \dots, y_n) = C(F_1(y_1), \dots, F_n(y_n)). \quad (1)$$

Conversely, any n -variate distribution function $H(y_1, \dots, y_n)$ can be written as in Equation (1) where $C(u_1, \dots, u_n)$ is a copula and $F_i(y_i)$ ($i = 1, \dots, n$) are the margins of H (Sklar, 1959). If the margins are continuous, then the copula is unique; otherwise, it is uniquely determined on $\text{Ran}(F_1) \times \dots \times \text{Ran}(F_n)$. Thus, the copula associated to H characterises its dependence structure.

Archimedean copulas (Genest and Mackay, 1986) is a subclass of copulas characterized by a *generator* $\varphi(t)$. It is a strictly decreasing and convex function on $(0, 1)$ such that $\varphi(1) = 0$, with the associated copula defined as

$$C(u_1, \dots, u_n) = \varphi^{-1}[\varphi(u_1) + \dots + \varphi(u_n)] \quad \forall (u_1, \dots, u_n) \in [0, 1]^n.$$

We shall restrict our attention to *strict* generators and, thus, assume that $\varphi(0^+) = +\infty$.

Many parametric proposals were made for that function (see e.g. Nelsen (1999)) usually involving one or two parameters to tune the strength of dependence. They differ in the induced dependence pattern.

In practice, one might decide to work with a specific generator. This could be motivated by an approximate prior knowledge of the dependence structure or by the robustness of the analysis conclusions to the generator choice. Otherwise, it is advisable to make the analysis for a library of generators and to proceed to a selection according to some criteria. Genest and Rivest (1993) proposed a graphical method comparing the distribution function $K(p)$ of a given bivariate Archimedean copula $C(u, v)$ with generator $\varphi(\cdot)$ to a non-parametric estimate of $K(p)$. The same authors proved that

$$K(p) = \Pr \{C(u, v) \leq p\} = p - \frac{\varphi(p)}{\varphi'(p^+)}.$$

Vandenhende and Lambert (2002b) consider different parametric generators in a longitudinal study and select the one yielding the smallest Akaike criterion for the fitted models. Goodness-of-fit procedures for copula models also exist: we refer to Genest et al. (2006) for a recent proposal and references on the topic. Alternatively, one could propose a very flexible specification for the

generator to make sure that the imposed structure is not too constraining. Vandenhende and Lambert (2005) approximate an unknown bivariate copula using a continuous piecewise log-linear combination of simple Archimedean generators.

Here, we consider a flexible B-splines specification for (a function of) the generator and show how the associated parameters can be estimated in a Bayesian framework. The B-splines specification is described in Section 2 and the inference strategy in Section 3. We show how smoothness can be required for the fitted generator in Section 4. The efficacy of the method is illustrated on simulated data in Section 5. A basic analysis of the dependence structure in systolic and diastolic blood pressures is proposed jointly with some marginal regression models in Section 6. We conclude the paper by a discussion.

2 B-splines specification of Archimedean copulas

Directly approximating the generator using polynomials is not a good strategy because $\varphi(u)$ is unbounded at 0^+ . Instead, we propose to build an approximation to the associated function $\lambda(t)$ where

$$\lambda(t) = \frac{\varphi(t)}{\varphi'(t)},$$

$$\varphi(t) = \varphi(t_0) \exp \left\{ \int_{t_0}^t \frac{ds}{\lambda(s)} \right\}.$$

Its properties can easily be derived from the requirements made on $\varphi(\cdot)$:

- (1) $\lambda(0) = \lambda(1) = 0$,
- (2) $\varphi(\cdot)$ is strictly decreasing on $(0, 1)$ provided that $\varphi'(\cdot) = \varphi(\cdot)/\lambda(\cdot)$ is negative, implying that $\lambda(\cdot)$ is negative on $(0, 1)$,
- (3) $\varphi(\cdot)$ is a convex function on $(0, 1)$ provided that

$$\varphi'' = \frac{\varphi'}{\lambda}(1 - \lambda')$$

is positive on $(0, 1)$, requiring that $\lambda'(\cdot) < 1$ on $(0, 1)$.

We consider a cubic B-splines (Boor, 1978) specification to approximate $\lambda(\cdot)$ on $[0, 1]$, i.e.

$$\tilde{\lambda}(t|\boldsymbol{\alpha}) = \mathbf{b}_t^T \boldsymbol{\alpha} \quad \forall t \in [0, 1]$$

where \mathbf{b}_t is the B-splines basis evaluated at t and associated to a pre-specified set of K knots on $[0, 1]$, and $\boldsymbol{\alpha} \in \mathbb{R}^{K+2}$. (The tilde symbol above a function

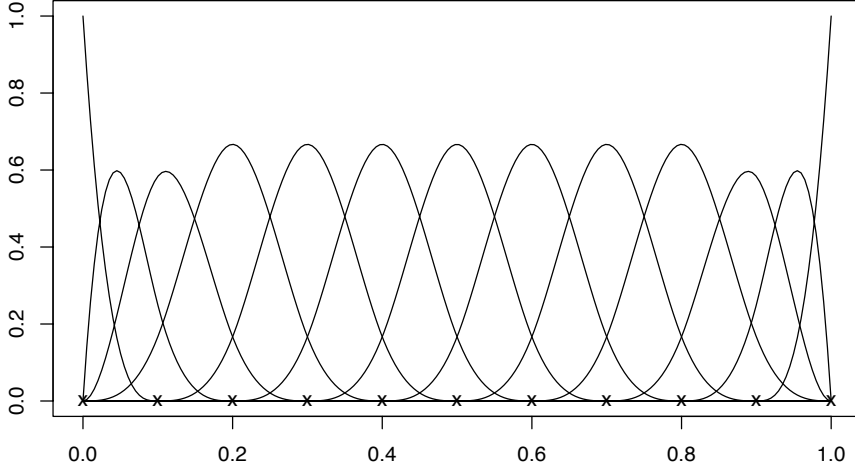


Fig. 1. Cubic B-splines on $[0, 1]$ for equidistant knots.

name will be used to indicate the associated B-spline approximation.) Such a basis is illustrated in Figure 1 for equidistant knots on $[0, 1]$. One can see that a cubic B-spline has a compact support spanned by 5 knots (possibly repeated).

Consider a set of K equidistant knots $\{\kappa_k = (k-1)/(K-1) : k = 1, \dots, K\}$ on $[0, 1]$. In practice, we recommend to take a reasonably large value for K (say between 20 and 30). If K is too large, our estimate of the generator will be too noisy (i.e. contaminated with irrelevant features of the dataset at hand). If it is too small, we shall potentially miss important features in the dependence structure. We propose to follow the proposal made by Eilers and Marx (1996) to keep a large number of knots counterbalanced by the introduction of a penalty in the inference procedure. Its specific implementation in a Bayesian framework is deferred to Section 3.

Consider the following notation, $\mathbf{b}'_t = \frac{d}{dt}\mathbf{b}_t$, for the first derivative of a B-spline basis. Then,

$$\tilde{\lambda}'(t|\boldsymbol{\alpha}) = \mathbf{b}'_t{}^T \boldsymbol{\alpha}$$

where $t \in [0, 1]$ and $\boldsymbol{\alpha} \in \mathbb{R}^{K+2}$.

These $K + 1$ components in $\boldsymbol{\alpha}$ cannot be freely chosen as we wish that the associated approximation

$$\tilde{\varphi}(t) = \tilde{\varphi}(t_0) \exp \left\{ \int_{t_0}^t \frac{ds}{\tilde{\lambda}(s)} \right\} \quad (2)$$

to the generator also checks the requirements made on any (strict) generator $\varphi(\cdot)$. As already examined at the beginning of the section, these requirements

translate into constraints on the lambda function, and thus here, on the components of $\boldsymbol{\alpha}$ involved in $\tilde{\lambda}(t|\boldsymbol{\alpha})$:

(1) We have

$$\tilde{\lambda}(0|\boldsymbol{\alpha}) = 0 \Leftrightarrow (1, 0, \dots, 0)^T \boldsymbol{\alpha} = \alpha_1 = 0.$$

Similarly,

$$\tilde{\lambda}(1|\boldsymbol{\alpha}) = 0 \Leftrightarrow (0, \dots, 0, 1)^T \boldsymbol{\alpha} = \alpha_{K+2} = 0.$$

(2) The monotonicity constraint on the generator forces

$$\tilde{\lambda}(t|\boldsymbol{\alpha}) = \mathbf{b}_t^T \boldsymbol{\alpha} < 0 \quad \forall t \in (0, 1).$$

Imposing that $\alpha_k < 0$ for $k \in \{2, \dots, K+1\}$ is a sufficient (but not necessary) condition.

(3) Finally, the convexity constraint on the generator imposes that a valid proposal for $\boldsymbol{\alpha}$ must check that

$$\tilde{\lambda}'(t|\boldsymbol{\alpha}) = \mathbf{b}_t'^T \boldsymbol{\alpha} < 1 \quad \forall t \in (0, 1).$$

Constraints 2 & 3 leave the parameter vector $\boldsymbol{\theta} = (\alpha_2, \dots, \alpha_{K+1})^T$ in a convex set Θ within \mathbb{R}^K .

3 Inference

We propose to make inference in a Bayesian framework. The following joint improper prior is first considered for the free B-splines (log-) parameters

$$p(\theta_1, \dots, \theta_K) = \begin{cases} 1 & \text{if } \boldsymbol{\theta} \in \Theta, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

A more elaborate prior will be proposed in Section 4.1. To simplify the presentation, we shall assume that we have paired data that can be seen as a random sample

$$\mathbf{y} = \{(u_i, v_i) : i = 1, \dots, n\}$$

with marginal (continuous) uniform distributions on $[0, 1]$.

If a B-splines model for the underlying (assumed) Archimedean copula is considered, see Section 2, then we obtain the following likelihood:

$$L(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^n \frac{\partial^2 \tilde{C}(u_i, v_i | \boldsymbol{\alpha})}{\partial u \partial v} = - \prod_{i=1}^n \frac{\tilde{\varphi}''(\tilde{C}_i | \boldsymbol{\alpha}) \tilde{\varphi}'(u_i | \boldsymbol{\alpha}) \tilde{\varphi}'(v_i | \boldsymbol{\alpha})}{\{\tilde{\varphi}'(\tilde{C}_i | \boldsymbol{\alpha})\}^3}$$

$$= - \prod_{i=1}^n [1 - \tilde{\lambda}'(\tilde{C}_i|\boldsymbol{\alpha})] \frac{\tilde{\lambda}(\tilde{C}_i|\boldsymbol{\alpha})}{\tilde{\lambda}(u_i|\boldsymbol{\alpha}) \tilde{\lambda}(v_i|\boldsymbol{\alpha})} \frac{\exp \left[\int_{v_i}^{u_i} \frac{ds}{\tilde{\lambda}(s|\boldsymbol{\alpha})} \right]}{\left\{ 1 + \exp \left[\int_{v_i}^{u_i} \frac{ds}{\tilde{\lambda}(s|\boldsymbol{\alpha})} \right] \right\}^2}$$

where $\boldsymbol{\alpha}^T = (0, \boldsymbol{\theta}^T, 0)$ and

$$\tilde{C}_i = \tilde{C}(u_i, v_i|\boldsymbol{\alpha}) = \tilde{\varphi}^{-1}(\tilde{\varphi}(u_i|\boldsymbol{\alpha}) + \tilde{\varphi}(v_i|\boldsymbol{\alpha})|\boldsymbol{\alpha}). \quad (4)$$

Several numerical difficulties arise when computing the likelihood. First, $\int_{v_i}^{u_i} \frac{ds}{\tilde{\lambda}(s|\boldsymbol{\alpha})}$ has to be evaluated for each pair of data. Directly using numerical quadrature methods (based on Newton-Cotes formulas) is not a good strategy as the integrand $1/\tilde{\lambda}(s|\boldsymbol{\alpha})$ has vertical asymptotes at 0 and 1. Instead, we found that making the complementary log-log change of variable $t = \log(-\log(s))$ stabilises the integrand. The integral becomes

$$\int_{\log(-\log(v_i))}^{\log(-\log(u_i))} \frac{\exp[1 - \exp(t)]}{\tilde{\lambda}\{\exp[1 - \exp(t)]|\boldsymbol{\alpha}\}} dt.$$

The integrand is computed once on a fine grid for a given $\boldsymbol{\alpha}$. The so-obtained values are used to compute the integral for all the pairs of data using a Newton-Cotes formula (e.g. the one corresponding to the trapeze method). These integrals are also necessary to compute the generator $\tilde{\varphi}(\cdot|\boldsymbol{\alpha})$, cf. Equation (2).

A second numerical difficulty is the computation of \tilde{C}_i . It requires the inversion of the function $\tilde{\varphi}(\cdot|\boldsymbol{\alpha})$, cf. Equation (4). It was done by evaluating the generator on a fine grid on $(0, 1)$ using the techniques just described: $\tilde{\varphi}^{-1}(\cdot|\boldsymbol{\alpha})$ can then be evaluated using interpolation.

Inference will be made using the posterior distribution for the B-splines parameters:

$$p(\theta_1, \dots, \theta_K|\mathbf{y}) \propto L(\boldsymbol{\theta}; \mathbf{y}) \times p(\theta_1, \dots, \theta_K).$$

It has the same support as the prior distribution.

We propose to explore that posterior distribution using Markov chains Monte-Carlo (McMC), see Gilks et al. (1996) for an excellent introduction. The Metropolis-Hastings algorithm will be used sequentially on the K components of $\boldsymbol{\theta}$. Some authors (see Carlin and Louis (2000), p. 159) suggest to name that particular procedure *univariate Metropolis*.

Assume that we start the chain at $\boldsymbol{\theta}^0 = (\theta_1^0, \dots, \theta_K^0)^T$. We propose to build the following chain at iteration m by updating the K components of $\boldsymbol{\theta}$ sequentially: to update the k th component,

- (1) Generate z such that $Z \sim N(0, 1)$ and build the proposal value

$$\boldsymbol{\xi}_k = (\theta_1^m \dots, \theta_{k-1}^m, \theta_k^{m-1} + \sigma_k z, \theta_{k+1}^{m-1}, \dots, \theta_K^{m-1})^T$$

for $\boldsymbol{\theta}$.

- (2) Denote the state of the chain after the update of the $(k-1)$ th component by

$$\boldsymbol{\zeta}_{k-1} = (\theta_1^m, \dots, \theta_{k-1}^m, \theta_k^{m-1}, \theta_{k+1}^{m-1}, \dots, \theta_K^{m-1})^T$$

with $\boldsymbol{\zeta}_0 = \boldsymbol{\theta}^{m-1}$.

Let

$$p = \frac{p(\boldsymbol{\xi}_k | \mathbf{y})}{p(\boldsymbol{\zeta}_{k-1} | \mathbf{y})}.$$

- If $p \geq 1$, accept the proposal, i.e. $\boldsymbol{\zeta}_k = \boldsymbol{\xi}_k$, and set $\theta_k^m = \theta_k^{m-1} + \sigma_k z$.
- If $p < 1$, accept the proposal with probability p and set

$$\theta_k^m = \begin{cases} \theta_k^{m-1} + \sigma_k z & \text{if accepted } (\boldsymbol{\zeta}_k = \boldsymbol{\xi}_k), \\ \theta_k^{m-1} & \text{otherwise } (\boldsymbol{\zeta}_k = \boldsymbol{\zeta}_{k-1}). \end{cases}$$

The theory ensures that after a sufficiently large number of iterations, say M , $\{\boldsymbol{\theta}^{M+1}, \boldsymbol{\theta}^{M+2}, \dots\}$ can be considered as a random sample from the posterior distribution. It is recommended in Gelman et al. (1996) to select the variances σ_k^2 's of the proposal densities to achieve acceptance rates in the range $(0.15, 0.40)$, as suggested by a careful study of the algorithm when the target distribution is multivariate normal. This can be done by increasing (decreasing) a variance when the observed acceptance rate is too small (large) for the concerned component.

To accelerate the inference procedure, one can use the following strategy close to the proposal in Lambert and Eilers (2005):

- (1) Run the univariate Metropolis algorithm for another extra few hundreds iterations, say M_1 , and use these to tune the proposal variances to make sure that the acceptance rate for each component is in a reasonable range.
- (2) Run the univariate Metropolis algorithm for a few hundreds iterations, say M_2 , using the updated proposal variances and make sure that the acceptance rates remain in the selected ranges for all the components.
- (3) Reparametrize the problem by applying a translation and a rotation to the initial parameter vector. More precisely, denote by S the empirical variance-covariance matrix of the parameters evaluated using the last M_2 iterations of the generated chain, and by $\bar{\boldsymbol{\theta}}$ the corresponding mean vector.

Reparametrize the posterior distribution using $\boldsymbol{\beta}$ where

$$\boldsymbol{\theta} = S^{\frac{1}{2}} \boldsymbol{\beta} + \bar{\boldsymbol{\theta}}.$$

Then, either use the univariate Metropolis algorithm described above on the reparametrized posterior distribution, or use the classical Metropolis algorithm with a vector proposal at each iteration instead of an iteration split into K component proposals.

Let us detail that vectorial approach. Denote by $\boldsymbol{\beta}^{m-1}$ the state of the chain at the $(m - 1)$ th iteration. Let

$$\begin{aligned}\boldsymbol{\xi} &= \boldsymbol{\beta}^{m-1} + \sigma \mathbf{z}, \\ p &= \frac{p(\boldsymbol{\xi}|\mathbf{y})}{p(\boldsymbol{\beta}^{m-1}|\mathbf{y})},\end{aligned}\tag{5}$$

where $\mathbf{Z} \sim N_K(0, I_K)$ and $\boldsymbol{\xi}$ is a proposal for the vector at iteration m .

- If $p \geq 1$, accept the proposal, i.e. set $\boldsymbol{\beta}^m = \boldsymbol{\xi}$.
- If $p < 1$, accept the proposal with probability p :

$$\boldsymbol{\beta}^m = \begin{cases} \boldsymbol{\xi} & \text{if accepted,} \\ \boldsymbol{\beta}^{m-1} & \text{otherwise.} \end{cases}$$

After a sufficiently large number of iterations, say M ,

$$\{S^{\frac{1}{2}}\boldsymbol{\beta}^{M+1} + \bar{\boldsymbol{\theta}}, S^{\frac{1}{2}}\boldsymbol{\beta}^{M+2} + \bar{\boldsymbol{\theta}}, \dots\}$$

can be considered as a random sample from the posterior distribution. In such high dimensional problem, Gelman et al. (1996) recommend to target an acceptance rate close to 25% which achieves the largest efficiency ($\approx 0.3/K$) of the Metropolis algorithm compared to independent samples when the target is multivariate normal. A useful guideline to choose σ in Equation (5) is to take $\sigma = 2.4/\sqrt{K}$ and to update it to approach the target acceptance rate.

4 Automatic smoothing of the copula estimator

In Section 2, we proposed to work with a large number of knots to build a B-splines approximation to the copula generator. That strategy can yield a noisy estimate revealing irrelevant (random) features from the dataset under study. A possible strategy proposed in the frequentist literature is to subtract a roughness penalty from the log-likelihood.

Schumaker (1981) proposes to use the penalty

$$\lambda \int |f''(x; \boldsymbol{\theta})|^2 dx$$

where $f(\cdot; \boldsymbol{\theta})$ is the curve model. The parameter λ is usually selected using cross-validation criteria or an information criterion like the AIC after defining the effective number of parameters as a function of λ .

O’Sullivan (1986) (see also Eilers and Marx (1996)) defines the penalty using differences of successive B-splines coefficients leading to

$$\text{penalty} = \lambda \sum_{j=r+1}^K (\Delta^r \theta_j)^2 = \lambda |D_r \boldsymbol{\theta}|^2 = \lambda \boldsymbol{\theta}^T P \boldsymbol{\theta}$$

where Δ^k denotes the difference operator of order k . Here, $P = D_r' D_r$, where D_r is the matrix representation of the difference operator Δ^r (Eilers and Marx, 1996).

4.1 Bayesian P-splines

In Bayesian terms, it is equivalent to introducing a prior distribution on the $(K - r)$ r th order differences of the B-splines coefficients, i.e.

$$\Delta^r \theta_j \sim N(0, \tau_\lambda^{-1})$$

(see Berry et al. (2002) for a Bayesian implementation of P-splines in normal regression models and Lambert and Eilers (2005); Lang and Brezger (2004); Brezger and Lang (2006) for a similar exercise in additive models). Consequently, we propose to multiply the improper prior in Equation (3) by

$$\tau_\lambda^{\frac{K-r}{2}} \exp \left\{ -\frac{1}{2} \tau_\lambda \boldsymbol{\theta}^T P \boldsymbol{\theta} \right\}.$$

The inverse variance τ_λ plays the role of λ in the frequentist approach. A vague prior distribution can be chosen for it, say, a gamma distribution $\mathcal{G}(a, b)$, i.e.

$$p(\tau_\lambda) = \frac{b^a}{\Gamma(a)} \tau_\lambda^{a-1} \exp(-b \tau_\lambda),$$

keeping in mind the potentially improper posterior that can arise when inappropriate choices are made (for a and b) to express our absence of knowledge on the penalty parameter (see Hobert and Casella (1996) for a discussion in hierarchical linear mixed models). Taking e.g. $a = b = .0001$ is usually suitable and expresses our ignorance, while yielding a proper posterior distribution for τ_λ (see the discussion and the simulations in Lang and Brezger (2004), and the recommendations in Brezger and Lang (2006)).

An initial condition $\boldsymbol{\theta}^0$ can be obtained for $\boldsymbol{\theta}$ using the non-parametric (NP) estimate $\tilde{\lambda}_{\text{NP}}(\cdot)$ proposed by Genest and Rivest (1993) for the lambda function

$\lambda(\cdot) = \varphi(\cdot)/\varphi'(\cdot)$. We propose to take for $\boldsymbol{\theta}^0$ the minimiser of

$$\sum_{v \in \mathcal{V}} \left\{ \tilde{\lambda}_{\text{NP}}(v) - \tilde{\lambda}(v|\boldsymbol{\alpha}) \right\}^2 + \nu \boldsymbol{\theta}^T P \boldsymbol{\theta}$$

with $\boldsymbol{\alpha}^T = (0, \boldsymbol{\theta}^T, 0)$ under the linear constraints $\boldsymbol{\theta} > 0$ and $\tilde{\lambda}'(v|\boldsymbol{\alpha}) < 1 \forall v \in (0, 1)$ where \mathcal{V} is chosen to be the set of values where the NP estimate changes. These constraints ensure that $\tilde{\varphi}(v|\boldsymbol{\alpha}^0)$ is a valid generator. The term $\nu \boldsymbol{\theta}^T P \boldsymbol{\theta}$ imposes smoothness to that function: the larger ν , the smoother the associated lambda function. This is a quadratic programming problem that can be solved numerically using e.g. the `quadprog` package developed in S by B.A. Turlach and ported to the R software (<http://cran.r-project.org>) by A. Weingessel.

A major advantage of the above Bayesian proposal over the frequentist penalized likelihood approach is the simultaneous estimation of the penalty parameter and of the other parameters, and, hence, the automatic accountability of the effects of the (usually) “imprecise” estimation of the penalty parameter on the dispersion of the other parameters. That imprecision is ignored in the frequentist approach leading to an underestimation of the standard errors of these other parameter estimators.

The Metropolis algorithm in Section 3 can easily be adapted by adding a Gibbs step for τ_λ to start each MCMC iteration. Indeed, one can easily show that the conditional posterior distribution of the penalty parameter, $p(\tau_\lambda|\boldsymbol{\theta}; \mathbf{y})$, is $\mathcal{G}(a^*, b^*)$ where

$$\begin{cases} a^* = a + \frac{K-r}{2} = a + \frac{\rho(P)}{2}, \\ b^* = b + \frac{1}{2} \boldsymbol{\theta}^T P \boldsymbol{\theta}, \end{cases}$$

with $\rho(P)$ denoting the rank of the penalty matrix P . Therefore, the Gibbs step at the m th iteration consists of a random generation of $\tau_\lambda^{(m)}$ using

$$\left(\tau_\lambda^{(m)} | \boldsymbol{\theta}^{(m-1)}; \mathbf{y} \right) \sim \mathcal{G} \left(a + \frac{\rho(P)}{2}, b + \frac{1}{2} \boldsymbol{\theta}^{(m-1)T} P \boldsymbol{\theta}^{(m-1)} \right).$$

Alternatively, one could work with the marginal posterior distribution for $\boldsymbol{\theta}$ obtained by integrating out the penalty parameter from the joint posterior:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) &= \int_0^\infty p(\boldsymbol{\theta}, \tau_\lambda|\mathbf{y}) d\tau_\lambda \\ &\propto \int_0^\infty L(\boldsymbol{\theta}|\mathbf{y}) \tau_\lambda^{a+\frac{\rho(P)}{2}-1} \exp \left\{ - \left(b + \frac{1}{2} \boldsymbol{\theta}^T P \boldsymbol{\theta} \right) \tau_\lambda \right\} d\tau_\lambda \end{aligned}$$

$$\propto \frac{L(\boldsymbol{\theta}|\mathbf{y})}{\left(b + \frac{1}{2}\boldsymbol{\theta}^T P \boldsymbol{\theta}\right)^{a + \frac{\rho(P)}{2}}}. \quad (6)$$

The Metropolis-Hastings algorithm proposed in Section 3 can be used to make inference from that marginal posterior distribution.

4.2 Bayesian P-splines with a mixture prior

The choice of the specific values for a & b in the penalty prior is usually not crucial and hardly affects the smooth of the approximating curve. However, in specific circumstances, the recommended choices can yield models that lack flexibility, see Jullion and Lambert (2007).

Therefore, we propose to use a mixture prior (see e.g. Bolstad (2004), Chap. 14) for the roughness penalty parameters. Let $\mathcal{Q} = \{a_q = 10^{-q} : q = 1, \dots, Q\}$ (say) be the set of values that we would like to evaluate for $a = b$ and denote by Q_q the q th prior model. A mixture prior for τ_λ giving an equal prior weight to the Q possibilities is

$$(\tau_\lambda|Q_q) \sim \mathcal{G}(a_q, b_q) \quad \text{with} \quad p(Q_q) = \frac{1}{Q}$$

where $p(Q_q)$ denotes the prior probability associated to the q th prior. The joint posterior distribution for $(\boldsymbol{\theta}, \tau_\lambda, Q_q)$ is

$$p(\boldsymbol{\theta}, \tau_\lambda, Q_q|\mathbf{y}) \propto L(\boldsymbol{\theta}; \mathbf{y}) p(\boldsymbol{\theta}|\tau_\lambda) p(\tau_\lambda|Q_q) p(Q_q).$$

The conditional posterior distributions, useful to set up the Gibbs sampler, are

$$\begin{aligned} p(\boldsymbol{\theta}|\tau_\lambda, Q_q; \mathbf{y}) &\equiv p(\boldsymbol{\theta}|\tau_\lambda; \mathbf{y}) \propto L(\boldsymbol{\theta}; \mathbf{y}) p(\boldsymbol{\theta}|\tau_\lambda), \\ (\tau_\lambda|\boldsymbol{\theta}, Q_q; \mathbf{y}) &\sim \mathcal{G}(a_q + 0.5 \rho(P), b_q + 0.5 \boldsymbol{\theta}^T P \boldsymbol{\theta}), \\ p(Q_q|\boldsymbol{\theta}, \tau_\lambda; \mathbf{y}) &\equiv p(Q_q|\tau_\lambda; \mathbf{y}) = \frac{p(\tau_\lambda|Q_q) p(Q_q)}{\sum_l p(\tau_\lambda|Q_l) p(Q_l)}. \end{aligned}$$

A marginal posterior for $\boldsymbol{\theta}$ can be derived from the joint:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \sum_{q=1}^Q \int_0^{+\infty} p(\boldsymbol{\theta}, \tau_\lambda, Q_q|\mathbf{y}) d\tau_\lambda$$

$$\propto L(\boldsymbol{\theta}; \mathbf{y}) \frac{1}{Q} \sum_{p=1}^Q \frac{\Gamma(a_q + 0.5 \rho(P)) b_q^{a_q}}{\Gamma(a_q) (b_q + 0.5 \boldsymbol{\theta}^T P \boldsymbol{\theta})^{a_q + 0.5 \rho(P)}}.$$

That expression is to be compared to Equation (6) where a single value was chosen for a & b . Like before, the Metropolis-Hastings algorithm proposed in Section 3 can be used to make inference from that marginal posterior distribution.

5 Illustration on simulated data

We propose to use simulations of paired data with uniform marginals and a dependence structure corresponding to two well-known Archimedean copula generators to illustrate the pertinence of the proposed inference procedure.

Let us shortly remind how such data can be generated when $\varphi(\cdot)$ is the generator:

- First generate two independent sets of n independent data, $\mathbf{u} = \{u_1, \dots, u_n\}$ and $\mathbf{t} = \{t_1, \dots, t_n\}$, with a uniform distribution on $(0, 1)$.
- Define

$$\begin{aligned} w_i &= \varphi'^{-1}\{\varphi'(u_i)/t_i\}, \\ v_i &= \varphi^{-1}\{\varphi(w_i) - \varphi(u_i)\}, \end{aligned}$$

for $i = 1, \dots, n$.

Then, as shown in Genest and Mackay (1986), $\{(u_i, v_i) : i = 1 \dots n\}$ is an independent random sample from a bivariate distribution with uniform marginals and an underlying Archimedean copula with generator $\varphi(\cdot)$.

The one-parameter families of Clayton's and Frank's generators were used to generate paired data with continuous uniform marginals. The copula parameters were chosen to correspond to a Kendall's tau equal to 0.3. One hundred datasets of size $N = 100$ were generated in both cases.

In all cases, inference was made using the strategy proposed in Sections 3 & 4 with $K = 20$ equidistant knots on $(0, 1)$. The non-parametric (NP) estimate (Genest and Rivest, 1993) of the copula generator was used to obtain a starting value for $\boldsymbol{\theta}^0$. A first chain of length 500 was run to help in specifying the variances σ_k^2 in the univariate Metropolis algorithm. Then, 500 extra iterations were run to evaluate the empirical variance-covariance matrix S used to reparametrize the posterior. Finally, two different strategies were considered to sample from the reparametrized posterior. The first one uses the univariate Metropolis algorithm to build a chain of length 1,000 with variances σ_k^2

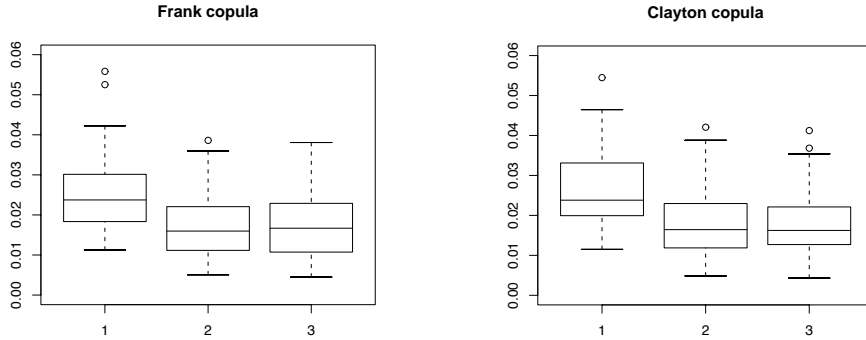


Fig. 2. Simulation study: boxplots of $\frac{1}{\text{card}(\mathcal{V})} \sum_{v \in \mathcal{V}} |\varphi(v) - \tilde{\varphi}(v)|$ for the Frank (left part) and Clayton (right part) copulas. The labels in abscissa indicate which estimate $\tilde{\varphi}(v)$ was used: 1: non-parametric estimate by Genest & Rivest. 2: posterior mean obtained with 1,000 iterations of the univariate Metropolis algorithm. 3: posterior mean obtained with 5,000 iterations of the classical (multivariate) Metropolis algorithm.

tuned with 200 preliminary iterations. The second one relies on 5,000 iterations of the classical (multivariate) Metropolis algorithm with multivariate normal proposals with variance-covariance matrix S (following 500 iterations for the burn-in). For each of the 100 simulated datasets, the distance

$$\frac{1}{\text{card}(\mathcal{V})} \sum_{v \in \mathcal{V}} |\varphi(v) - \tilde{\varphi}(v)|$$

was computed with $\tilde{\varphi}$ corresponding to the NP estimate of Genest & Rivest, to the posterior mean estimate obtained with the univariate Metropolis algorithm, and to the posterior mean estimate obtained with the classical (multivariate) Metropolis algorithm. \mathcal{V} was chosen to be the set of values where the NP estimate changes.

The boxplots of these distances are provided in Figure 2 for the three different estimates of the generator. The performances of the MCMC estimates are better than the non-parametric one. The MCMC estimates are equally performant (as it should be asymptotically). This suggests working with the classical (multivariate) Metropolis algorithm as it is much faster. Indeed, the univariate Metropolis algorithm requires K evaluations of the posterior per iteration while the former algorithm just evaluates the posterior once per iteration.

The posterior mean is not the only useful result of the MCMC procedure. The chains can be also used to compute credible regions for the generator or for any function of the generator like Kendall's tau, specific posterior predictive probabilities, etc. An example is proposed in Figure 3. The fitted lambda function $\tilde{\lambda}(u|\bar{\alpha}_{\text{post}}) = \tilde{\varphi}(u|\bar{\alpha}_{\text{post}})/\tilde{\varphi}'(u|\bar{\alpha}_{\text{post}})$ (evaluated at the estimated posterior mean $\bar{\alpha}_{\text{post}}^T = (0, \bar{\theta}_{\text{post}}^T, 0)$ of the spline parameters) is proposed (thick dashed

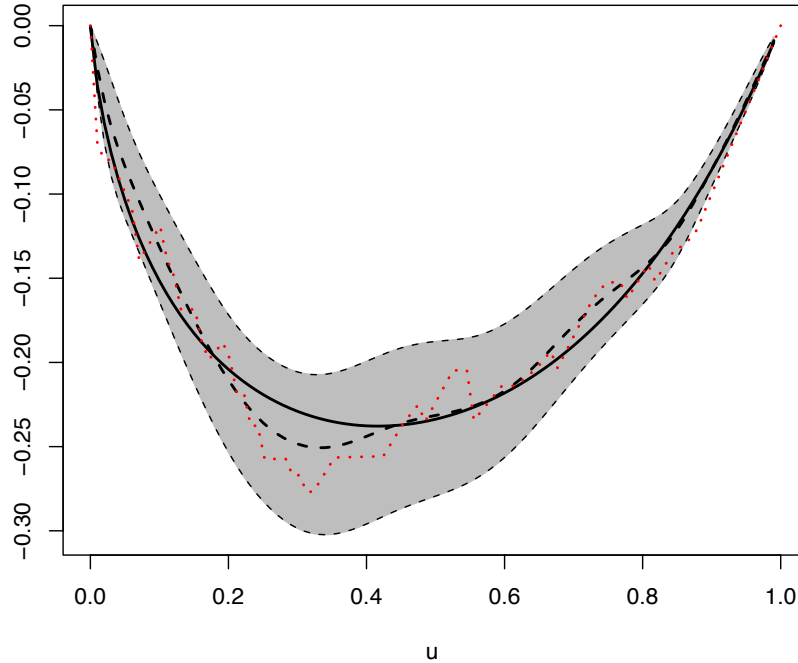


Fig. 3. Lambda function $\varphi(u)/\varphi'(u)$ associated to the Frank's generator (thick solid line ; Kendall's tau= 0.30), Genest & Rivest (1993) non-parametric estimate (dotted line), fitted lambda function $\tilde{\varphi}(u|\bar{\alpha}_{\text{post}})/\tilde{\varphi}'(u|\bar{\alpha}_{\text{post}})$ (dashed line) and associated 90% credible envelope (grey area).

lines) together with the 90% credible envelope (containing 90% of the 5,000 sampled $\tilde{\lambda}(\cdot|\alpha^m)$, $m = 1, \dots, 5,000$) for one of the 100 simulated samples: the envelope contains the target (thick solid line). One can see that the fitted function somehow tries to track the non-parametric estimate (dotted line) with the constraints that it should be smooth and computed from a valid generator (cf. end of Section 2).

The posterior probability $p(Q_q|\mathbf{y})$ associated to the q th prior for τ_λ with $\mathcal{Q} = \{a_q = b_q = 10^{-q} : q = 1, \dots, 6\}$ was estimated and found to be

q	1	2	3	4	5	6
$p(\widetilde{Q}_q \mathbf{y})$	0.00	0.00	0.32	0.57	0.10	0.01

showing the relative contribution of each prior to the mixture prior.

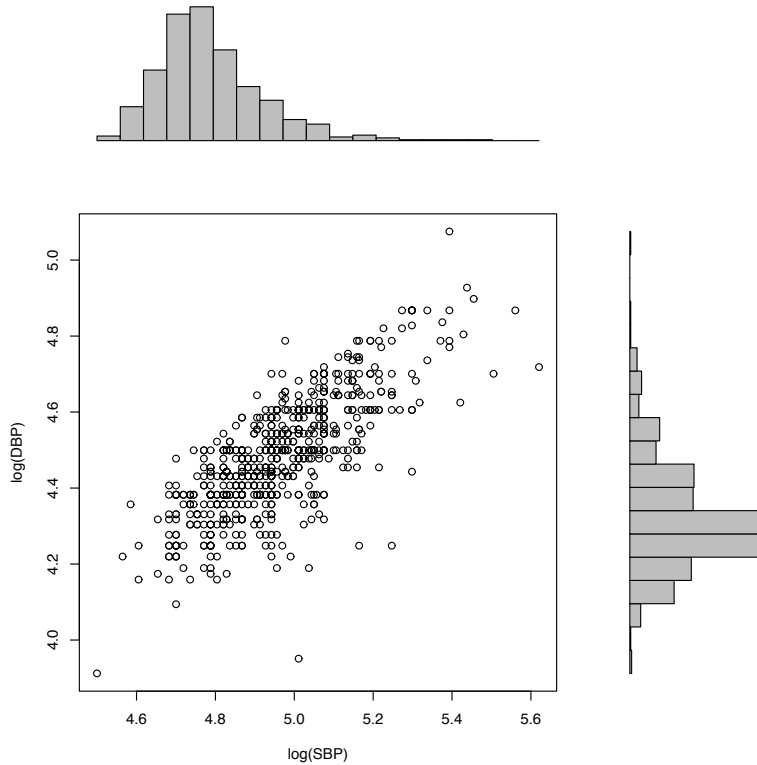


Fig. 4. Histograms and scatterplot of the log-diastolic and log-systolic blood pressures.

6 Application to real data

We propose to illustrate the method using a subset of the Framingham Heart study data (<http://www.framingham.com/heart/>). We shall focus our attention on the dependence structure underlying the diastolic (DBP) and the systolic (SBP) blood pressures (in *mmHg*) measured on 663 male subjects at their first visit (see Lambert and Vandenhende (2002) for another illustration of the use of copula to analyse hemodynamic parameters measured in a longitudinal study). The histograms of the log-blood pressures and the associated scatterplot can be found in Figure 4. It suggests asymmetric marginal distributions for the log-blood pressures and a strong positive dependence between the two responses.

The marginal distributions of $\log(\text{DBP})$ and $\log(\text{SBP})$ will be modelled using the 4-parameter skewed-Student distribution (Fernández and Steel, 1998). The location parameter, μ , which turns to be the mode, will be related to a covariate using the regression model

$$\mu = \beta_0 + \beta_1 \widetilde{\text{CHOL}}$$

where $\widetilde{\text{CHOL}}$ is the cholesterol level (mean 226 and standard error 42) of the considered subject minus the mean cholesterol level of the subjects in the

study. The dispersion parameter (which is the standard deviation when the distribution is symmetric), the skewness parameter (with a distribution which is left skewed when negative, symmetric when 0, right skewed when positive) and the tail parameter (which is the degrees of freedom) will not be related to that covariate.

We assume that the dependence structure in the blood pressures can be adequately described using an Archimedean copula with unknown (strict) generator. A B-splines specification, see Section 2, will be used for it.

Two estimation strategies were considered. The first one is the traditional two-steps strategy consisting of fitting marginal models first, and then estimating the copula parameters conditionally on the fitted marginals. The second one proposes to estimate the marginal and the dependence parameters in a single run. It enables to reflect the uncertainty attached to the marginal parameters estimation when it comes to the estimation of the parameters involved in the description of the dependence structure, and vice-versa.

We do not expect much difference in the results between the two strategies as the dependence structure is fully characterized by the copula in the continuous case. Our expectations were confirmed in our example. Note that this would not be the case anymore if we were dealing with discrete data, see Denuit and Lambert (2005), Vandenhende and Lambert (2002a) and the references therein for motivating arguments.

Summary measures of the posterior distributions can be found in Table 1 for the marginal skewed-Student regression models for $\log(\text{SBP})$ and $\log(\text{DBP})$, and in Table 2 for the B-splines parameters describing the fitted Archimedean copula generator.

Table 1 reveals a positive marginal association between the cholesterol level and blood pressures (see β_1), the positive skewness of the distributions of $\log(\text{DBP})$ and $\log(\text{SBP})$, as well as very moderate kurtosis. Note that the quality of the fits provided by the marginal parametric models was assessed and found to be excellent.

The results in Table 2 are summarized graphically in Figure 5. Like for the simulated data, we see that $\tilde{\lambda}(u|\bar{\alpha}_{\text{post}})$ is close to the non-parametric estimate proposed in Genest and Rivest (1993). It is smooth as required.

The posterior probability $p(Q_q|\mathbf{y})$ associated to the q th prior for τ_λ with $\mathcal{Q} = \{a_q = b_q = 10^{-q} : q = 1, \dots, 6\}$ was estimated and found to be

q	1	2	3	4	5	6
$p(\widetilde{Q}_q \mathbf{y})$	0.00	0.00	0.18	0.67	0.14	0.01

	log(DBP)		log(SBP)	
	Mean	MC error	Mean	MC error
	95% c.i.		95% c.i.	
β_0	4.44 (4.42, 4.47)	0.0003	4.87 (4.85, 4.91)	0.0003
β_1	4.1×10^{-4} $(1.5, 6.7) \times 10^{-4}$	3.0×10^{-6}	2.9×10^{-4} $(0.29, 5.5) \times 10^{-4}$	2.8×10^{-6}
$\log(\sigma)$	-1.92 (-1.98, -1.85)	0.0008	-1.89 (-1.96, -1.82)	0.0008
skewness	0.152 (0.040, 0.261)	0.0013	0.319 (0.195, 0.441)	0.0013
1/df	0.084 (0.022, 0.154)	0.0007	0.057 (0.007, 0.122)	0.0006

Table 1

Posterior means and 95% credible intervals for the parameters involved in the marginal regression models for the diastolic and the systolic blood pressures ; Monte-Carlo errors with a chain of length 10,000 using the univariate Metropolis algorithm after a rotation and a translation, see Section 3.

showing the relative contribution of each prior to the mixture prior.

As an illustration, the lambda function corresponding to three well-known (parametric) Archimedean copulas evaluated at their MLEs for the dependence parameter are also plotted on Figure 5. One can see that the Frank's and Clayton's copula are not adequate to summarize the dependence structure in the (log-) blood pressures, while the Gumbel copula is obviously more appropriate without being fully satisfactory.

The fitted joint distribution for the log-diastolic and log-systolic blood pressures for an average cholesterol level is given as a contour plot (solid curves) in Figure 6 together with the contours (dashed curves) for the Gumbel generator. The fitted distributions are very close with some differences in the shape of the contours around the modal values for the responses. This feature was already revealed in Figure 5 with the Gumbel lambda function lying slightly outside the 90% credible region around 0.45.

	Mean	90% c.i.	MC error
θ_1	-0.030	(-0.044,-0.019)	0.0003
θ_2	-0.068	(-0.081,-0.056)	0.0003
θ_3	-0.106	(-0.120,-0.093)	0.0003
θ_4	-0.132	(-0.148,-0.116)	0.0004
θ_5	-0.156	(-0.174,-0.138)	0.0005
θ_6	-0.169	(-0.187,-0.149)	0.0005
θ_7	-0.170	(-0.185,-0.153)	0.0004
θ_8	-0.159	(-0.174,-0.144)	0.0003
θ_9	-0.148	(-0.163,-0.133)	0.0004
θ_{10}	-0.141	(-0.158,-0.124)	0.0004
θ_{11}	-0.146	(-0.161,-0.129)	0.0004
θ_{12}	-0.144	(-0.158,-0.130)	0.0003
θ_{13}	-0.138	(-0.151,-0.124)	0.0003
θ_{14}	-0.122	(-0.136,-0.109)	0.0003
θ_{15}	-0.102	(-0.115,-0.090)	0.0003
θ_{16}	-0.086	(-0.097,-0.075)	0.0003
θ_{17}	-0.074	(-0.085,-0.063)	0.0003
θ_{18}	-0.054	(-0.063,-0.045)	0.0002
θ_{19}	-0.026	(-0.033,-0.020)	0.0001
θ_{20}	-0.015	(-0.021,-0.010)	0.0001

Table 2

B-splines estimation of the Archimedean copula generator: posterior means and 90% credible intervals for the B-splines parameters associated to $K = 20$ equidistant knots on $(0, 1)$; Monte-Carlo errors in the two-steps estimation approach with a chain of length 10,000 using the multivariate Metropolis algorithm after a rotation and a translation, see Section 3.

7 Discussion

We have shown how to construct a smooth approximation to the generator of an Archimedean copula using cubic B-splines. That approximation is constrained to be a valid generator. A Bayesian setting was found convenient to express some of these constraints and to guarantee smoothness.

Smoothness was obtained using a Bayesian translation (Berry et al., 2002;

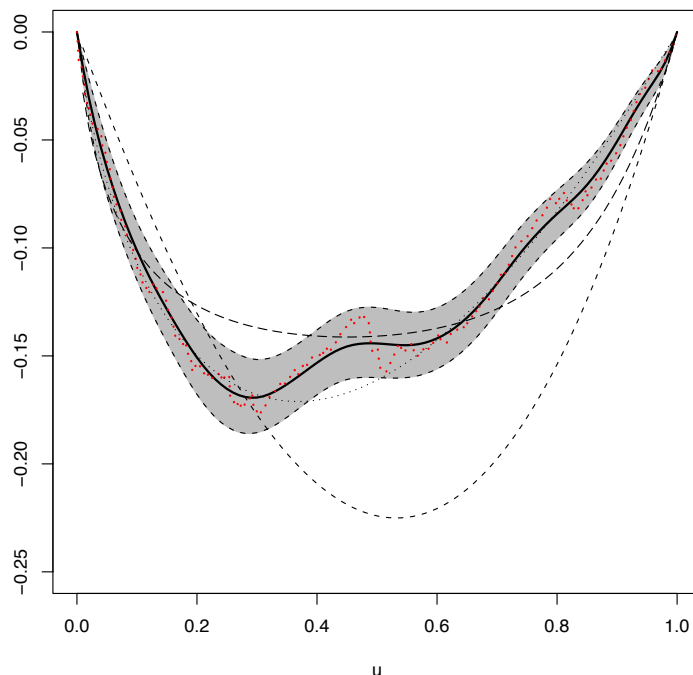


Fig. 5. $\lambda(u) = \varphi(u)/\varphi'(u)$: our B-splines estimate $\tilde{\lambda}(u|\bar{\theta}_{\text{post}})$ (solid line) and the associated 90% credible envelope (grey area) ; the Genest & Rivest (1993) non-parametric estimate (discontinuous dotted line) ; best fitting Clayton (dashed), Gumbel (dotted line) and Frank (long-dashed) copulas.

Lang and Brezger, 2004; Lambert and Eilers, 2005) of the roughness penalty proposed by Eilers and Marx (1996) through the prior distribution of the B-splines parameters. We extended the corresponding procedure by specifying a mixture prior for the penalty parameter as the usual gamma prior was found to affect the smooth of the fitted curve in specific circumstances. The marginal posterior distributions of the spline parameters was also derived.

The generated chains can be used to obtain an approximation to the posterior distribution of any function of the generator such as Kendall's tau or Spearman's rho. It was used here to derive credible intervals for $\lambda(\cdot) = \varphi(\cdot)/\varphi'(\cdot)$. The utility of the method was illustrated using simulated datasets and data from an epidemiological study.

Acknowledgements

I would like to thank Prof. Paul Eilers (Univ. of Leiden, The Netherlands) for useful discussions on smoothing techniques with B-splines. This paper greatly benefited from the excellent course on 'Advanced use of MCMC methods' de-

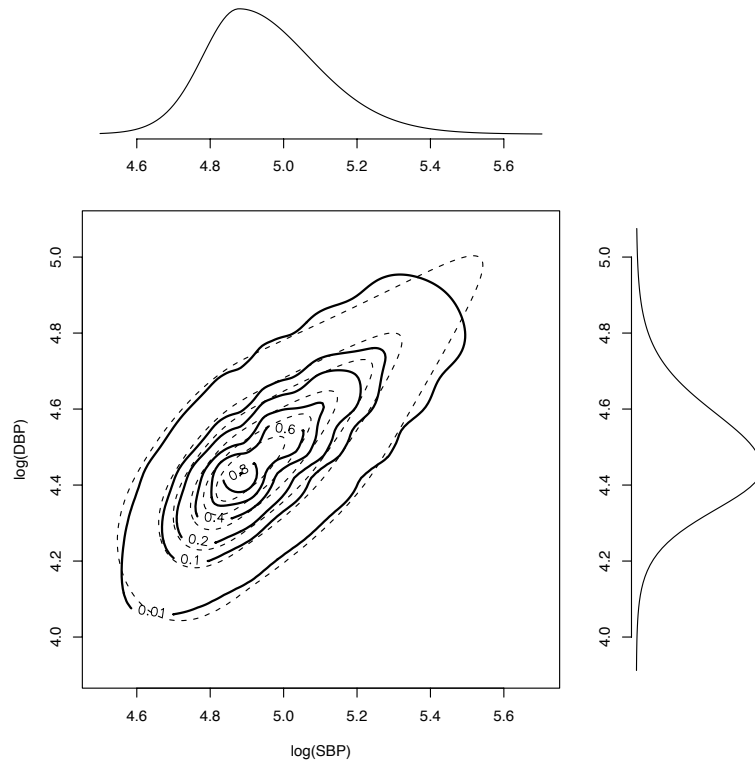


Fig. 6. Fitted joint distribution for the log-diastolic and log-systolic blood pressures for an average cholesterol level using a B-splines generator (solid contours) or the Gumbel generator (dashed contours).

livered by Prof. Petros Dellaportas (Athens Univ.) during his visit at KUL (Leuven, Belgium) in September 2003. Of course, I am fully responsible for the potential misuses of MCMC techniques in the present paper.

Financial support from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

References

- Berry, S. M., Carroll, R. J., Ruppert, D., 2002. Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association* 97, 160–169.
- Bolstad, W., 2004. *Introduction to Bayesian Statistics*. Wiley.
- Boor, C. D., 1978. *A Practical Guide to Splines*. Springer, Berlin.
- Brezger, A., Lang, S., 2006. Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis* 50, 967–991.
- Carlin, B. P., Louis, T. A., 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall / CRC.

- Denuit, M., Lambert, P., 2005. Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis* 93, 40–57.
- Eilers, P. H. C., Marx, B. D., 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* 11, 89–121.
- Fernández, C., Steel, M. F. J., 1998. On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association* 93, 359–371.
- Gelman, A., Roberts, G. O., Gilks, W. R., 1996. Efficient Metropolis jumping rules. In: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M. (Eds.), *Bayesian Statistics 5*. Oxford: Oxford University Press, pp. 599–607.
- Genest, C., Mackay, R. J., 1986. Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *La revue canadienne de statistique* 14, 145–159.
- Genest, C., Quessy, J.-F., Rémillard, B., 2006. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics* 32, 337–366.
- Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association* 88, 1034–1043.
- Gilks, W. R., Richardson, S., Spiegelhalter, D. J., 1996. *Markov chain Monte Carlo in Practice*. Chapman and Hall/CRC.
- Hobert, J. P., Casella, G., 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association* 91, 1461–1473.
- Jullion, A., Lambert, P., 2007. Specification of the roughness penalty prior distribution in Bayesian P-splines models. *Computational Statistics and Data Analysis* 51, 2542–2558.
- Lambert, P., Eilers, P. H., 2005. Bayesian proportional hazards model with time varying regression coefficients: a penalized Poisson regression approach. *Statistics in Medicine* 24, 3977–3989.
- Lambert, P., Vandenhende, F., 2002. A copula based model for multivariate non normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. *Statistics in Medicine* 21, 3197–3217.
- Lang, S., Brezger, A., 2004. Bayesian P-splines. *Journal of Computational and Graphical Statistics* 13, 183–212.
- Nelsen, R. B., 1999. *An Introduction to Copulas*. Springer-Verlag.
- O’Sullivan, F., 1986. A statistical perspective on ill-posed inverse problems. *Statistical Science* 1, 502–518.
- Schumaker, L., 1981. *Spline Functions: Basic Theory*. Wiley.
- Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8, 229–231.
- Vandenhende, F., Lambert, P., 2002a. Improved rank-based dependence measures for categorical data. *Statistics and Probability Letters* 63, 157–163.
- Vandenhende, F., Lambert, P., 2002b. On the joint analysis of longitudinal responses and early discontinuation in randomized trials. *Journal of Biopharmaceutical Statistics* 12, 425–440.

Vandenhende, F., Lambert, P., 2005. Local dependence estimation using semi-parametric Archimedean copulas. *Canadian Journal of Statistics* 33, 377–388.