Baker's distribution, Bernstein copula and B-spline copulas

Xiaoling DOU (Waseda University)*

Abstract

A method that uses order statistics to construct multivariate distributions with fixed marginals is proposed by Baker (2008). We investigate the properties of Baker's bivariate distributions. The properties include the weak convergence to the Fréchet–Hoeffding upper bound, the product-moment convergence and the totally positivity of order 2. As Baker's distribution utilizes a representation of the Bernstein copula in terms of a finite mixture distribution, we propose expectation-maximization (EM) algorithms to estimate the Bernstein copula and give illustrative examples using real data sets and a 3-dimensional simulated data set. These studies show that the Bernstein copula is able to represent various distributions flexibly and that the proposed EM algorithms work well for such data. Using B-spline functions, we construct a new class of copulas, B-spline copulas, that includes the Bernstein copulas as a special case. The range of correlation of the B-spline copulas is examined, and the Fréchet–Hoeffding upper bound is proved to be attained when the number of B-spline functions goes to infinity. As the B-spline functions are well-known to be an order-complete weak Tchebycheff system, we show that the property of total positivity of any order (TP_∞) follows for the maximum correlation case. These results extend the classical results for the Bernstein copulas. In addition, we derive in terms of the Stirling numbers of the second kind an explicit formula for the moments of the related B-spline functions on $[0, \infty)$.

1. Introduction

A novel method that applied the theory of order statistics to construct multivariate distributions with given marginal distributions was developed by Baker [1]. We refer to Lin, et al. [15] for a recent survey of this topic.

Baker's idea, applied to univariate cumulative distribution functions F and G, can be described as follows: Let $\{X_1, \dots, X_m\}$ and $\{Y_1, \dots, Y_n\}$ be independent random samples from the distributions F and G, respectively. Let $X_{k,m}$ be the kth smallest order statistic of $\{X_1, \dots, X_n\}$, and denote by $F_{k,m}$ the distribution of $X_{k,m}$; we write this as $X_{k,m} \sim F_{k,m}$. Similarly, we denote by $Y_{l,n}$ the lth smallest order statistic of $\{Y_1, \dots, Y_n\}$ and we let $G_{l,n}$ be its corresponding distribution, written $Y_{l,n} \sim G_{l,n}$. (Note that F and G can be discrete distributions.)

Let $R = (r_{kl})_{1 \le k,l \le n}$ be a parameter matrix whose matrix entries r_{kl} satisfy the conditions

$$\sum_{k=1}^{m} r_{kl} = \frac{1}{n}, \quad \sum_{l=1}^{n} r_{kl} = \frac{1}{m}, \quad r_{kl} \ge 0, \quad k = 1, 2, \dots m, \quad l = 1, 2, \dots, n. \tag{1}$$

This work was supported by KAKENHI (26730024, 16K00060).
Keywords: Bernstein function, B-spline function, Fréchet–Hoeffding upper bound, Maximum correlation, Order-complete weak Tchebycheff system, Order statistics, Total positivity of order r.

*e-mail: xiaoling@aoni.waseda.jp

Now choose the pair $(X_{k,m}, Y_{l,n})$ with probability r_{kl}, $k = 1, 2, \ldots m$, $\quad l = 1, 2, \ldots, n$. Then $(X_{k,m}, Y_{l,n})$ follows Baker's bivariate distribution: For $x, y \in \mathbb{R}$, the joint cumulative distribution function $H(x, y) := \Pr(X_{k,m} \le x,\ Y_{l,n} \le y)$ satisfies

$$\begin{aligned} H(x,y) &= \sum_{k=1}^{m}\sum_{l=1}^{n} r_{kl} F_{k,m}(x) G_{l,n}(y) \\ &= (F_{1,m}(x), \ldots, F_{m,m}(x))\, R\, (G_{1,n}(y), \ldots, G_{n,n}(y))^{\mathsf{T}}, \end{aligned} \tag{2}$$

where "T" denotes transpose.

For maximal correlation, Baker [1] proposes the bivariate distribution of (X, Y):

$$\begin{aligned} H_n(x,y) &= \Pr(X \le x, Y \le y) \\ &= \frac{1}{n}\sum_{k=1}^{n} F_{k,n}(x) G_{k,n}(y) \qquad \text{for all } x, y \in \mathbb{R} \equiv (-\infty, \infty), \text{ when } m = n, \end{aligned} \tag{3}$$

where

$$\begin{aligned} F_{k,n}(x) = \Pr(X_{k,n} \le x) &= n\binom{n-1}{k-1}\int_0^{F(x)} t^{k-1}(1-t)^{n-k} dt \\ &= k\binom{n}{k}\int_0^{F(x)} t^{k-1}(1-t)^{n-k} dt, \end{aligned}$$

and $G_{k,n}$ has a similar form (see, e.g., (1) of Hwang and Lin[9]). Indeed, the distribution H_n has marginals F and G because $H_n(x, \infty) = \frac{1}{n}\sum_{k=1}^{n} F_{k,n}(x) = F(x)$ and $H_n(\infty, y) = \frac{1}{n}\sum_{k=1}^{n} G_{k,n}(y) = G(y)$.

This article is organized as follows. In Section 2, we consider the asymptotic properties of Baker's distribution. In Section 3, we confirm Baker's distribution is the Bernstein copula and provide an EM algorithm for estimating joint density function. In Section 4, we define the B-spline copula as a generalization of the Bernstein copula and examine its dependence properties for its maximum correlation case. Finally, we make a summary in Section 5.

2. Asymptotic properties of Baker's distributions

In this section we provide the limiting distribution of Baker's bivariate distribution H_n as well as the limits of product moments $E[(X_n)^p (Y_n)^q]$.

Theorem 1 *Let (X_n, Y_n) be a sample generated from Baker's distribution H_n with absolutely continuous marginals F and G. Let $U_n = \sqrt{n}(G(Y_n) - F(X_n))$. Then, as $n \to \infty$, (X_n, U_n) converges in distribution to $(\widetilde{X}, \widetilde{U})$ whose joint density is*

$$k(x,u) = \frac{1}{2\sqrt{\pi F(x)\bar{F}(x)}} \exp\left\{ -\frac{u^2}{4F(x)\bar{F}(x)} \right\} f(x),$$

where f is the density of F. The conditional distribution of $\widetilde{U}$ given $\widetilde{X} = x$ is the normal distribution with mean 0 and variance $2F(x)\bar{F}(x)$:

$$\widetilde{U}|_{\widetilde{X}=x} \sim N(0, 2F(x)\bar{F}(x)).$$

The proof can be done by applying a generalization of the so-called local limit theorem for binomial probability.

From Theorem 1, we can see that $G(Y_n) - F(X_n) \xrightarrow{p} 0$ as $n \to \infty$. This implies the weak convergence

$$(X_n, Y_n) \xrightarrow{d} (\widetilde{X}, \widetilde{Y}) = (\widetilde{X}, G^{-1}(F(\widetilde{X}))), \quad \widetilde{X} \sim F \quad (n \to \infty), \tag{4}$$

where G^{-1} is the quantile function of G, namely, $G^{-1}(t) = \inf\{y : G(y) \geq t\}$, $t \in (0,1)$. The continuity of F and G are assumed in Theorem 1. Indeed, (4) holds for general F and G in the following form.

Theorem 2 *Let* (X_n, Y_n) *be a sample generated from Baker's distribution* H_n *with general marginals* F *and* G*. Then,* (X_n, Y_n) *converges in distribution to* $(\widetilde{X}, \widetilde{Y}) = (F^{-1}(Z), G^{-1}(Z))$ *as* $n \to \infty$*, where* Z *is a uniform random variable on* $(0,1)$.

To illustrate Theorem 2, we now give three simulation results. Baker's distribution has the advantage that a random number $(X_n, Y_n) \sim H_n$ can easily be generated as follows:

(i) Generate $X_1^*, \ldots, X_n^* \sim F$ i.i.d., and $Y_1^*, \ldots, Y_n^* \sim G$ i.i.d.

(ii) Generate $K \sim \mathrm{Unif}\{1, \ldots, n\}$.

(iii) Then, $(X_n, Y_n) \stackrel{d}{=} (X_{(K)}^*, Y_{(K)}^*)$ (the pair of the Kth smallest order statistics).

Figure 1 shows the scatter plots of two random samples of size 50 from Baker's distribution H_n with $n = 3$ and 100. In this simulation, F is a normal distribution $N(0,1)$, and G is a logistic distribution with mean 0 and variance 1, denoted by Logistic$(0,1)$. We can see that the larger n is, the more tightly random points accumulate around the curve $\{(x,y) : F(x) = G(y)\} = \{(F^{-1}(t), G^{-1}(t)) : t \in (0,1)\}$.

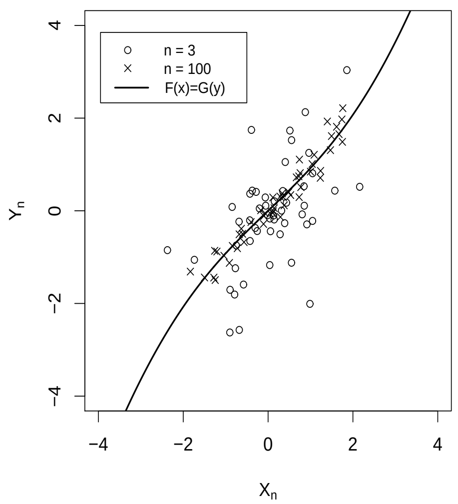

Figure 1: Two random samples (circles and crosses) from H_n with marginals $N(0,1)$ and Logistic$(0,1)$.

When F is a continuous distribution with a density function f, whereas G is a discrete distribution defined by $\Pr(Y = y_i) = q_i$, $i \in \mathbb{Z}$, where $\cdots < y_{-1} < y_0 < y_1 < \cdots$ is an increasing sequence, the limiting distribution of H_n is described as follows:

$$\frac{\Pr(\widetilde{X} \in (x, x + dx), \widetilde{Y} = y_j)}{dx} = \begin{cases} f(x), & \text{if } F(x) \in (Q_{j-1}, Q_j), \\ 0, & \text{otherwise.} \end{cases}$$

The examples depicted in Figure 2 are generated from Baker's distribution H_n with F the standard normal distribution $N(0, 1)$ and G a binomial distribution $\mathrm{Bin}(6; 0.3)$ for $n = 3$ and 100, respectively. The limiting support is displayed by the solid line. Sample sizes in both panels are 50. Most random points gather on the limiting support when $n = 100$.

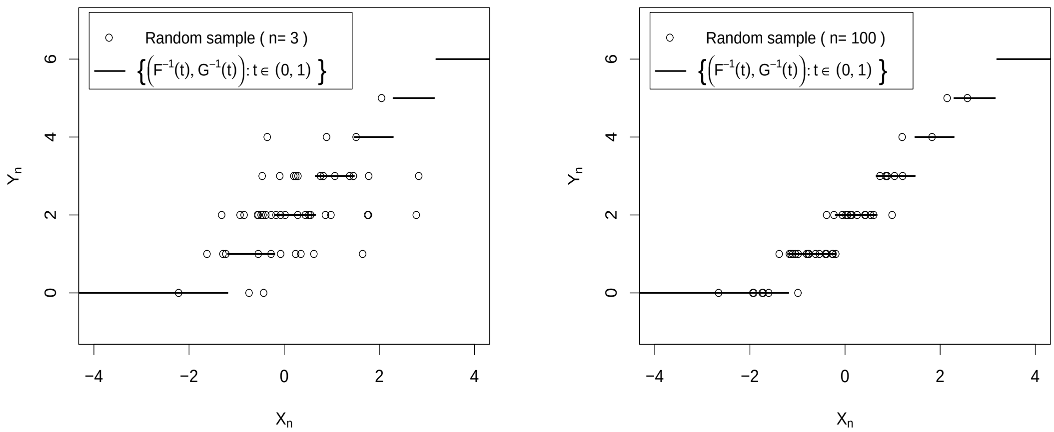

Figure 2: Two random samples (circles) from H_n with marginals $N(0, 1)$ and $\mathrm{Bin}(6; 0.3)$.

Suppose that the marginal distributions F and G are discrete such that

$$\Pr(X = x_i) = p_i, \quad \Pr(Y = y_i) = q_i, \quad i \in \mathbb{Z},$$

where $\cdots < x_{-1} < x_0 < x_1 < \cdots$ and $\cdots < y_{-1} < y_0 < y_1 < \cdots$ are increasing sequences. Then, the limiting distribution of H_n is a discrete distribution with probability

$$\begin{aligned} \Pr((\widetilde{X}, \widetilde{Y}) = (x_i, y_j)) &= \Pr(Z \in (P_{i-1}, P_i), Z \in (Q_{j-1}, Q_j)) \\ &= \max\{\min\{P_i, Q_j\} - \max\{P_{i-1}, Q_{j-1}\}, 0\}, \end{aligned}$$

where $P_i = \sum_{j \le i} p_j$ and $Q_i = \sum_{j \le i} q_j$. The two random point sets plotted in Figure 3 are sampled from Baker's distribution H_n with binomial distributions $F = \mathrm{Bin}(5; 0.5)$ and $G = \mathrm{Bin}(6; 0.3)$ for $n = 3$ and $n = 100$, respectively. The limiting support $\{(F^{-1}(t), G^{-1}(t)) : t \in (0, 1)\}$ is displayed by the small pluses. Sample sizes of the two samples are 50. Because of the overprinting, only few random points are shown. We can see that in case of $n = 100$, more random points overlap on the limiting support than in case of $n = 3$.

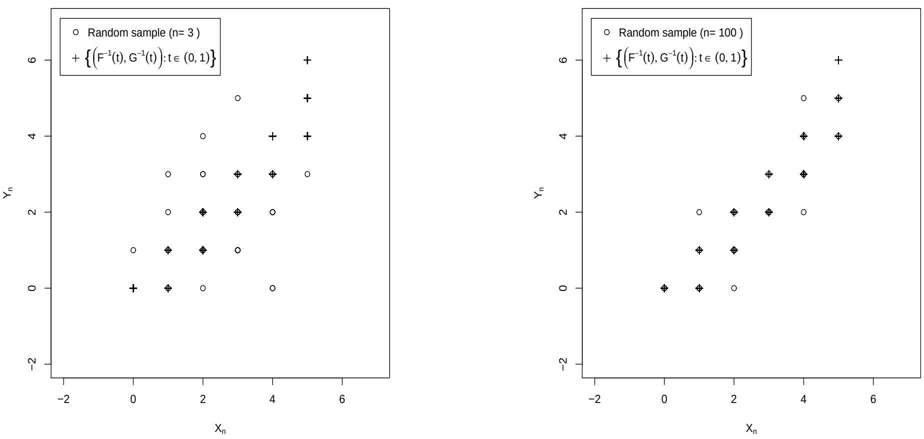

Figure 3: Two random samples (circles) from H_n with marginals $F = \mathrm{Bin}(5; 0.5)$ and $G = \mathrm{Bin}(6; 0.3)$.

The next theorem states that the convergence of $\{H_n\}_{n=1}^{\infty}$ to the Fréchet–Hoeffding upper bound is not only in the sense of weak convergence but in the sense of convergence of moments. This is a particular feature of the distributions with their marginals fixed. The proof is given in [5].

Theorem 3 *Let (X_n, Y_n) be a sample generated from Baker's distribution H_n with general marginals F and G. Let $(\widetilde{X}, \widetilde{Y}) = (F^{-1}(Z), G^{-1}(Z)), Z \sim Unif(0,1)$, having the limiting distribution. Let p and q be positive integers. If $E[|X_n|^{p+q}]$, $E[|Y_n|^{p+q}] < \infty$, then*

$$\lim_{n\to\infty} E[X_n^p Y_n^q] = E[(\widetilde{X})^p (\widetilde{Y})^q] = \int_0^1 \left(F^{-1}(t)\right)^p \left(G^{-1}(t)\right)^q dt.$$

3. Bernstein copula

The Bernstein polynomial and its integral are known as the following forms:

$$b_{k,n}(u) = \binom{n}{k} u^k (1-u)^{n-k}, \quad B_{k,n}(u) = \int_0^u b_{k,n}(t)dt, \quad u \in [0,1]. \tag{5}$$

We also know that the distribution function of the kth smallest order statistic $X_{k,m}$ of $X_1, \ldots, X_m$ is

$$F_{k,m}(x) = \sum_{j=k}^{m} \binom{m}{j} F(x)^j (1 - F(x))^{m-j} = m \int_0^{F(x)} \binom{m-1}{k-1} t^{k-1} (1-t)^{m-k} dt.$$

Using the notation above, we can rewrite the cdf as

$$F_{k,m}(x) = m B_{k-1,m-1}\left(F(x)\right).$$

Similar for the lth smallest order statistic $y_{l,n}$ of $Y_1, \ldots, Y_n$, the cdf can be written into

$$G_{l,n}(y) = n B_{l-1,n-1}\left(G(y)\right).$$

Then (2) becomes

$$\begin{aligned} H(x,y;\boldsymbol{R}) &= mn\sum_{k=1}^{m}\sum_{l=1}^{n} r_{k,l}B_{k-1,m-1}\left(F(x)\right)B_{l-1,n-1}\left(G(y)\right) \\ &= C^{\mathrm{Bern}}(F(x),G(y);\boldsymbol{R}), \end{aligned} \tag{6}$$

which is called the Bernstein copula [20]. If the random variables are continuous, taking derivatives, we can find their densities of the order statistics as

$$\begin{aligned} f_{k,m}(x) &= m\binom{m-1}{k-1}F(x)^{k-1}(1-F(x))^{m-k}f(x), \\ g_{l,n}(y) &= n\binom{n-1}{l-1}G(y)^{l-1}(1-G(y))^{n-l}g(y). \end{aligned}$$

Rewriting them with the Bernstein polynomial, we have

$$f_{k,m}(x) = mb_{k-1,m-1}\left(F(x)\right)f(x), \quad g_{l,n}(y) = nb_{l-1,n-1}\left(G(y)\right)g(y).$$

Then the density function of the bivariate Baker's distribution can be rewritten into a form with the Bernstein copula density

$$\begin{aligned} h(x,y;\boldsymbol{R}) &= mn\sum_{k=1}^{m}\sum_{l=1}^{n} r_{k,l}b_{k-1,m-1}(F(x))b_{l-1,n-1}(G(y))f(x)g(y) \\ &= c^{\mathrm{Bern}}(F(x),G(y);\boldsymbol{R})f(x)g(y). \end{aligned} \tag{7}$$

Hence, for random variable $U,V \sim i.i.d.\mathrm{Unif}(0,1)$, the Bernstein copula and its density are

$$\begin{aligned} C^{\mathrm{Bern}}(u,v;\boldsymbol{R}) &= mn\sum_{k=1}^{m}\sum_{l=1}^{n} r_{k,l}B_{k-1,m-1}\left(u\right)B_{l-1,n-1}\left(v\right) \quad \text{and} \\ c^{\mathrm{Bern}}(u,v;R) &= mn\sum_{k=1}^{m}\sum_{l=1}^{n} r_{kl}b_{k-1,m-1}(u)b_{l-1,n-1}(v), \end{aligned} \tag{8}$$

respectively.

3.1. EM algorithms based on the pseudo-likelihood function

Copulas are used to model dependence structures for multivariate data sets([18], [22]). Among the class of copulas, the Bernstein copula has two remarkable features. First, because of the Weierstrass approximation theorem, any 2-dimensional copula can be approximated uniformly on $[0,1]^2$ by the Bernstein copula density (8), when m and n are sufficiently large. Therefore, any continuous bivariate density function can be approximated by the density arising from the Bernstein copula. A second remarkable feature of the Bernstein copula is that it is a finite mixture distribution. This allows us to apply the expectation-maximization (EM) algorithm [17] to estimate parameters. In this subsection, we suppose that X and Y are continuous random variables, and that F and G are absolutely continuous with densities f and g, respectively. We also assume that the marginal distributions F and G have been estimated in advance, and we shall treat them in the subsequent analysis as known functions.

On the basis of a random sample of size N on (X, Y), let F_N and G_N denote the marginal empirical distributions of X and Y. We take F and G to be estimated by $NF_N/(N+1)$ and $NG_N/(N+1)$, respectively. If f and g, the corresponding density functions of F and G, exist then we estimate them with kernel estimators. The likelihood function with F, G, f and g replaced by their corresponding estimators is called the *pseudo-likelihood function*.

Suppose that an independent, identically distributed (i.i.d.) sample (x_i, y_i), $i = 1, \ldots, N$, is obtained from Baker's distribution (7). According to the standard method for estimating a finite mixture distribution, we introduce a pair of unobserved variables (K_i, L_i) for observation i, with probability $\Pr(K_i = k, L_i = l) = r_{k,l}$, $k \in \{1, \ldots, m\}$, $l \in \{1, \ldots, n\}$, $i = 1, \ldots, N$. We also define an $m \times n$ matrix $\tau_i = (\tau_{i,k,l})$ as a dummy variable with elements

$$\tau_{i,k,l} = \begin{cases} 1, & \text{if } (K_i, L_i) = (k, l) \\ 0, & \text{if } (K_i, L_i) \neq (k, l) \end{cases}$$

$i = 1, \ldots, N$. Note that τ_i and (K_i, L_i) are one-to-one. The likelihood for the full data set (x_i, y_i, τ_i), $i = 1, \ldots, N$, is given by

$$\prod_{i=1}^{N} \prod_{k=1}^{m} \prod_{l=1}^{n} \big\{ r_{k,l} f_{k:m}(x_i) g_{l:n}(y_i) \big\}^{\tau_{i,k,l}}. \tag{9}$$

The E-step in the EM algorithm calculates the conditional expectation of $\tau_{i,k,l}$ given (x_i, y_i), $i = 1, \ldots, N$; that is,

$$\begin{aligned} \widehat{\tau}_{i,k,l} &= E\big[\tau_{i,k,l} \,|\, (x_i, y_i)_{1 \le i \le N}; R\big] \\ &= \frac{r_{k,l} f_{k:m}(x_i) g_{l:n}(y_i)}{h(x_i, y_i; R)} \\ &= \frac{r_{k,l} b_{k-1,m-1}(F(x_i)) b_{l-1,n-1}(G(y_i))}{c(F(x_i), G(y_i); R)}. \end{aligned} \tag{10}$$

The M-step maximizes the logarithm of the likelihood (9) with respect to $r_{k,l}$ by assuming $\tau_{i,k,l} = \widehat{\tau}_{i,k,l}$. The logarithm of the expectation of (9) divided by N is

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{m} \sum_{l=1}^{n} \widehat{\tau}_{i,k,l} \log(r_{k,l} f_{k:m}(x_i) g_{l:n}(y_i)) = \sum_{k=1}^{m} \sum_{l=1}^{n} \bar{\tau}_{k,l} \log r_{k,l} + \text{const.}, \tag{11}$$

where $\bar{\tau}_{k,l} = \sum_{i=1}^{N} \widehat{\tau}_{i,k,l}/N$.

Maximizing the function (11) is a convex problem which has a unique maximizer $R^* = (r^*_{k,l})$ because (11) is a proper concave function in $r_{k,l}$ and the region for $R = (r_{k,l})$ defined by (1) is convex. Moreover, if $\bar{\tau}_{k,l} > 0$ for all k, l then the maximizer R^* is a (relative) interior point of the region (1); in that case, the maximizer R^* is obtained by the Lagrange multiplier method under the conditions $\sum_{l=1}^{n} r_{k,l} = 1/m$, $\sum_{k=1}^{m} r_{k,l} = 1/n$ for all k and l.

We introduce Lagrange multipliers μ_k and λ_l, and proceed to maximize

$$L = \sum_{k=1}^{m} \sum_{l=1}^{n} \bar{\tau}_{k,l} \log r_{k,l} - \sum_{k} \mu_k \Big(\sum_{l} r_{k,l} - \frac{1}{m} \Big) - \sum_{l} \lambda_l \Big(\sum_{k} r_{k,l} - \frac{1}{n} \Big)$$

with respect to $r_{k,l}$, μ_k and λ_l. Then, the maximizers $r^*_{k,l}$, μ^*_k and λ^*_l are obtained as the solution of

$$\frac{\partial L}{\partial r_{k,l}} = \frac{\bar{\tau}_{k,l}}{r_{k,l}} - \mu_k - \lambda_l = 0$$

subject to the restrictions in (1).

To find μ^*_k and λ^*_l satisfying

$$r_{k,l} = \frac{\bar{\tau}_{k,l}}{\mu_k + \lambda_l} > 0 \tag{12}$$

as well as the restriction (1), we propose the following procedure:

Algorithm 3.1

Step M0: Set $\mu_k^{(0)} = 1/2$ *and* $t = 0$.

Step M1: For fixed $\boldsymbol{\mu}^{(t)} = \left(\mu_1^{(t)}, \ldots, \mu_m^{(t)}\right)'$, *and for* $1 \le l \le n$, *find* $\lambda_l^{(t)}$ *numerically as a unique solution* λ_l *of*

$$\sum_{k=1}^{m} \frac{\bar{\tau}_{k,l}}{\mu_k^{(t)} + \lambda_l} = \frac{1}{n} \quad \text{such that} \quad \lambda_l > -\min_k\left(\mu_k^{(t)}\right).$$

Step M2: For fixed $\boldsymbol{\lambda}^{(t)} = \left(\lambda_1^{(t)}, \ldots, \lambda_n^{(t)}\right)'$, *and for* $1 \le k \le m$, *find* $\widetilde{\mu}_k^{(t)}$ *numerically as a unique solution* $\widetilde{\mu}_k$ *of*

$$\sum_{l=1}^{n} \frac{\bar{\tau}_{k,l}}{\widetilde{\mu}_k + \lambda_l^{(t)}} = \frac{1}{m} \quad \text{such that} \quad \widetilde{\mu}_k > -\min_l\left(\lambda_l^{(t)}\right).$$

Step M3: Let

$$\mu_k^{(t)} = \widetilde{\mu}_k^{(t)} - \frac{1}{m}\left(\sum_{k=1}^{m} \widetilde{\mu}_k^{(t)} - \sum_{k=1}^{m} \mu_k^{(0)}\right), \quad 1 \le k \le m.$$

Increase the counter t *by* 1, *and repeat Steps M1–M3 until (12) converges.*

To apply the EM algorithm, we will use

$$\widetilde{r}_{k,l} = \#\left\{ i \;\Big|\; \frac{k-1}{m} < \frac{N}{N+1} F_N(x_i) \le \frac{k}{m},\; \frac{l-1}{n} < \frac{N}{N+1} G_N(y_i) \le \frac{l}{n} \right\} \Big/ N. \tag{13}$$

as an initial value of $r_{k,l}$ the estimator given by [20] and [10]. Then the EM algorithm is summarized as follows.

Algorithm 3.2

Step 0: Set $r_{k,l}$ *equal to* $\widetilde{r}_{k,l}$ *in (13).*

Step 1: Find $\widehat{\tau}_{i,k,l}$ *by (10) (E-step).*

Step 2: Update $r_{k,l}$ *by Algorithm 3.1, Steps M0–M3 (M-step).*

Repeat Steps 1 and 2 until $\widehat{\tau}_{i,k,l}$ *converges.*

Note that this algorithm can be extended to Baker's distributions with three or more variables.

3.2. Illustrative examples

In this section, we demonstrate how our algorithms perform in practical data analysis. The results show that the algorithms work well in all the illustrative examples.

3.2.1. Consomic mouse data

The first data set consists of measurements of blood concentrations of biochemical substances in mice, and we apply Algorithm 3.2 for fitting Baker's distribution (7) with continuous variables.

The data set ([24], [25])consists of measurements of triglycerides (TG) and plasma high-density lipoprotein cholesterol (HDL) as plotted in Figure 5.

Using the Gaussian kernel estimator, we first estimate the marginal density functions. The bandwidths are selected according to Silverman's "rule of thumb" [23]. We use the empirical distribution functions to approximate the (cumulative) distribution functions. The estimated marginal densities and distribution functions are shown in the left and right panels, respectively, of Figure 4. Subsequently, we estimate the Bern-

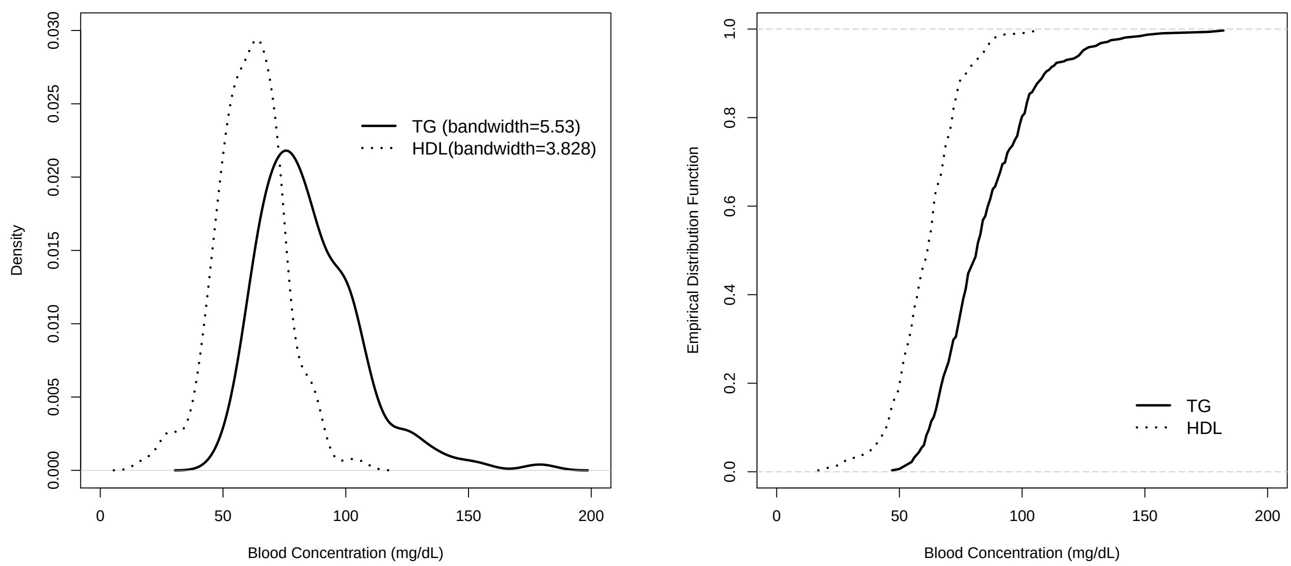

Figure 4: Estimated marginals of TG and HDL.
(Left: density functions. Right: cumulative distribution functions.)

stein copula density (8) with the EM algorithm (Algorithm 3.2) for fixed m and n. In the estimation, we determine the matrix size of R by the Akaike information criterion (AIC) [13]. From Table 1, we find that the AIC attains its minimum value, 5210.52, when $(m, n) = (2, 3)$. Table 1 also shows that the cases in which $(m, n) = (2, 2)$ and $(m, n) = (2, 3)$ have very close AIC values; indeed, the estimated contours based on these two cases are very similar.

For the case in which $(m, n) = (2, 3)$, the initial value $\widetilde{R}$ in (13) and the MLE $\widehat{R}$ obtained as the limit of sequence starting from $\widetilde{R}$ are

$$\widetilde{R} = \begin{pmatrix} 0.232 & 0.137 & 0.118 \\ 0.099 & 0.188 & 0.226 \end{pmatrix} \quad \text{and} \quad \widehat{R} = \begin{pmatrix} 0.333 & 0.106 & 0.061 \\ 0.000 & 0.227 & 0.273 \end{pmatrix},$$

respectively. The consistent estimate of the covariance of $(\widehat{r}_{11}, \widehat{r}_{12}, \widehat{r}_{13}, \widehat{r}_{21}, \widehat{r}_{22}, \widehat{r}_{23})'$ is

Table 1: AIC for female consomic mouse data.
(The minimum AIC is indicated with a box.)

$m \setminus n$	1	2	3	4	5	6	8	10
1	5242.00	5242.00	5242.00	5242.00	5242.00	5242.00	5242.00	5242.00
2	5242.00	5210.57	$\boxed{5210.52}$	5212.15	5211.15	5210.80	5212.67	5216.23
3	5242.00	5212.55	5211.94	5214.22	5215.47	5217.64	5223.91	5230.53
4	5242.00	5214.56	5215.69	5219.16	5220.33	5224.48	5234.19	5244.29
5	5242.00	5215.37	5218.51	5223.65	5226.87	5232.10	5246.20	5259.89
6	5242.00	5216.59	5220.58	5225.99	5231.44	5238.67	5256.04	5273.77
8	5242.00	5218.77	5225.45	5233.77	5242.13	5253.45	5277.90	5302.69
10	5242.00	5221.55	5229.78	5241.92	5253.58	5268.72	5300.85	5332.22

calculated as

$$\begin{pmatrix} 0.003 & -0.001 & -0.003 & -0.001 & 0.001 & -0.002 \\ -0.001 & 0.003 & 0.003 & -0.002 & 0.000 & 0.002 \\ -0.003 & 0.003 & 0.010 & -0.003 & -0.004 & 0.003 \\ -0.001 & -0.002 & -0.003 & 0.006 & 0.002 & -0.003 \\ 0.001 & 0.000 & -0.004 & 0.002 & 0.003 & -0.002 \\ -0.002 & 0.002 & 0.003 & -0.003 & -0.002 & 0.005 \end{pmatrix}.$$

A contour plot of the estimated joint density $h(x, y; \widehat{R})$ is shown in Figure 5.

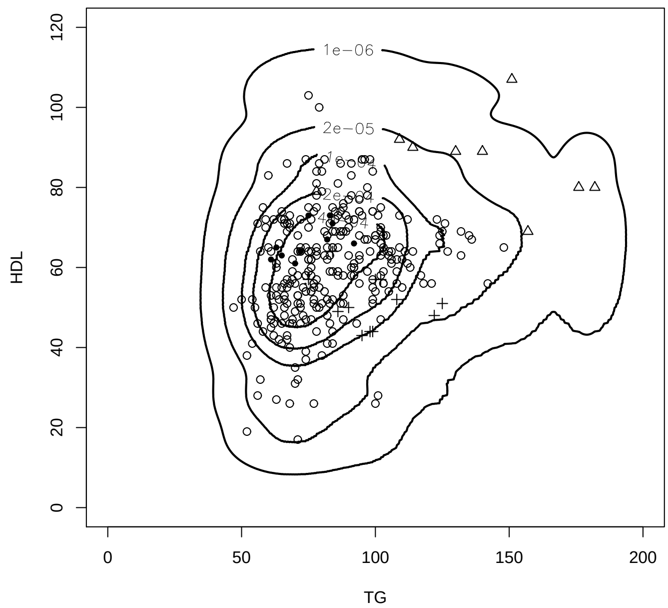

Figure 5: TG and HDL data (female consomic mice) and estimated contour.
(Dots: B6, Pluses: B6-Chr4MSM, Triangles: MSM, Circles: others.)

3.2.2. Simulated trivariate data with interaction

The second example is an artificial trivariate continuous data. The data are generated by the following two steps. First, we generate data $(u_{1,i}, u_{2,i}, u_{3,i})$, $i = 1, \ldots, N$, from a trivariate Baker's distribution with the copula density

$$c(u_1, u_2, u_3) = n_1 n_2 n_3 \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \sum_{k_3=1}^{n_3} r_{k_1,k_2,k_3} \prod_{j=1}^{3} b_{k_j-1,n_j-1}(u_j). \tag{14}$$

Here, the parameter $R = (r_{k_1,k_2,k_3})$ is defined as

$$r_{k_1,k_2,1} = \frac{1}{2n_1n_2} \quad (\text{for all } k_1, k_2), \qquad r_{k_1,k_2,2} = \begin{cases} \frac{1}{2n_1} & (\text{if } k_1 = k_2), \\ 0 & (\text{if } k_1 \neq k_2), \end{cases}$$

with $n_1 = n_2 = 20$ and $n_3 = 2$. The sample size is chosen to be $N = 2000$. Also, we convert the uniform marginals to normal marginals by $x_i = \Phi^{-1}(u_{1,i})$, $y_i = \Phi^{-1}(u_{2,i})$, $z_i = \Phi^{-1}(u_{3,i})$, where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. We then obtain random data (x_i, y_i, z_i) whose marginals are the standard normal distribution.

The first row of Figure 6 depicts scatter plots for the first and second variates (X, Y) stratified with the third variable Z. The correlation between X and Y is designed to be increasing in Z, and the marginals of (X, Z) and (Y, Z) are independent. From the three panels, we can see that X and Y are almost independent when Z is small and they are highly correlated when Z is large.

We fit the Bernstein copula density (14) with an extended version of Algorithm 3.2 as well as Algorithm 3.1 for this 3-dimensional data set. The contours of the estimated density function are shown in the second row of Figure 6, and we see that the Bernstein copula represents well the characteristic of the changing correlation. For comparison, we also plot the contours estimated with the Gaussian copula in the last row of the figure even though a Gaussian copula obviously cannot adapt to the change of correlation (i.e., the 3-way interaction). Nevertheless, this example demonstrates the flexibility of the Bernstein copula and the usefulness of the EM algorithm for 3-dimensional data.

EM algorithms are also available when some of the random variables are discrete and when the joint distribution is considered as a mixture of an independent case and the maximum correlated case of Baker's distribution. All the EM algorithms, more examples and detailed analysis can be found in [6].

4. B-spline copulas

We consider first a general setting based on *order-complete weak Tchebycheff systems* (OCWT-systems) [11], and then we define a class of B-spline copulas that includes the Bernstein copulas as special cases. After that, we investigate the dependence properties of the most correlated case of the B-spline copula. We also involve the moments of the B-spline functions.

4.1. Definition of the B-spline copula

Let $q_k \geq 0$, $k = 1, \ldots, n$, $\sum_{k=1}^{n} q_k = 1$, and let $\phi_1, \ldots, \phi_n$ be probability densities on $[0, 1]$ such that

$$\sum_{k=1}^{n} q_k \, \phi_k(t) = 1, \tag{15}$$

$t \in [0, 1]$. We assume further that $\{\phi_1, \ldots, \phi_n\}$ is an OCWT-system, i.e.,

(i) $\phi_1, \ldots, \phi_n$ are linearly independent, and

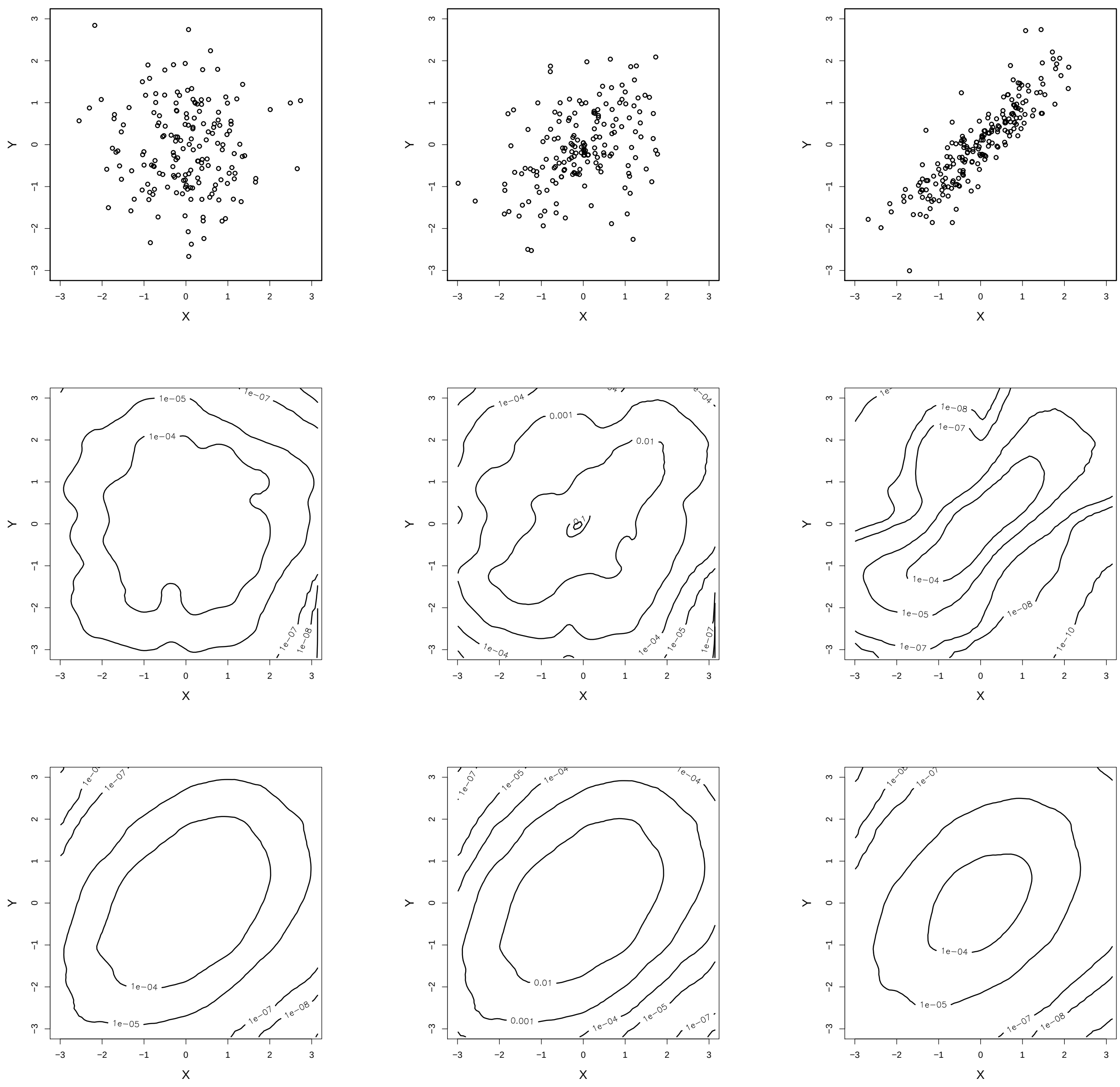

Figure 6: Simulated trivariate data and estimated contours.
Scatter plots for stratified data (first row), the estimated density with a Bernstein copula (second row), and the estimated density with a Gaussian copula (third row) for three cases of $(X, Y|Z)$.
Left: Z is small ($Z < 0.1$); Center: Z is moderately-sized ($0.45 < Z < 0.55$); Right: Z is large ($Z > 0.9$).

(ii) $\phi_k(t)$ is *totally positive of order* n *(TP_n) in* (k,t), i.e., for each $r = 1,\ldots,n$,

$$\det\left(\phi_{k_i}(t_j)\right)_{r\times r} \ge 0 \tag{16}$$

for all $k_1 > \cdots > k_r$ and $t_1 > \cdots > t_r$.

See Karlin and Studden [12, Chapter 1] or Schumaker [21, Chapter 2] for examples of OCWT systems.

Let $q_{1k} \ge 0$, $k = 1,\ldots,n_1$, such that $\sum_{k=1}^{n_1} q_{1k} = 1$. Also, let $q_{2l} \ge 0$, $l = 1,\ldots,n_2$, such that $\sum_{l=1}^{n_2} q_{2l} = 1$. Letting

$$\Phi_k(u) = \int_0^u \phi_k(t)\,\mathrm{d}t,$$

$u \in [0,1]$, we define the *B-spline copula*, a generalization of the Bernstein copula (6), by

$$C(u,v;R) = \sum_{k=1}^{n_1}\sum_{l=1}^{n_2} r_{k,l}\,\Phi_k(u)\,\Phi_l(v), \tag{17}$$

$u, v \in [0,1]$, with parameter matrix

$$\begin{aligned} &R = (r_{k,l})_{1\le k\le n_1;\,1\le l\le n_2}, \quad r_{k,l} \ge 0,\\ &\sum_{k=1}^{n_1} r_{k,l} = q_{2l},\ \sum_{l=1}^{n_2} r_{k,l} = q_{1k}, \quad k = 1,2,\ldots,n_1,\ l = 1,2,\ldots,n_2. \end{aligned} \tag{18}$$

The copula (17) is a *bona fide* copula since, for any $u \in [0,1]$,

$$\begin{aligned} C(u,1;R) &= \sum_{k=1}^{n_1}\sum_{l=1}^{n_2} r_{k,l}\Phi_k(u) = \sum_{k=1}^{n_1} q_{1k}\Phi_k(u)\\ &= \int_0^u \sum_{k=1}^{n_1} q_{1k}\phi_k(t)\,\mathrm{d}t = \int_0^u 1\,\mathrm{d}t = u; \end{aligned}$$

and similarly, $C(1,v;R) = v$, $v \in [0,1]$.

4.2. The maximum correlation copula

From now on, we restrict our attention to the case in which $n_1 = n_2 = n$ and $q_{1k} = q_{2k} = q_k$; further, we use the notation $Q = \mathrm{diag}(q_k)_{1\le k\le n}$ for the diagonal matrix with diagonal entries $q_1,\ldots,q_n$.

Theorem 4 *For the copula (17) with the parameter space (18), the maximum correlation is attained when* $r_{kl} = q_k\delta_{kl}$*, equivalently,* $R = Q$*.*

In the maximum correlation case, $C(u,v;R)$ becomes

$$C^*(u,v) := C(u,v;Q) = \sum_{k=1}^{n} q_k\Phi_k(u)\Phi_k(v), \tag{19}$$

$u, v \in [0,1]$.

To prove Theorem 4, we need the following crucial lemma.

Lemma 1 *Let $a_1 \geq \cdots \geq a_n \geq 0$ and $b_1 \geq \cdots \geq b_n \geq 0$ be given. Let $q_1, \ldots, q_n \geq 0$ satisfy $\sum_{k=1}^n q_k = 1$. Then,*

$$\max_{\substack{\sum_k r_{kl}=q_l \\ \sum_l r_{kl}=q_k \\ r_{kl}\geq 0}} \sum_{k=1}^n \sum_{l=1}^n r_{kl} a_k b_l \;=\; \sum_{k=1}^n q_k a_k b_k.$$

Proof of Theorem 4. Since $\{\phi_1, \ldots, \phi_n\}$ is an OCWT-system then, for all $i < j$ and $s < t$,

$$\phi_i(s)\phi_j(t) - \phi_j(s)\phi_i(t) = \det \begin{pmatrix} \phi_i(s) & \phi_i(t) \\ \phi_j(s) & \phi_j(t) \end{pmatrix} \geq 0.$$

Integrating this inequality with respect to (s, t) over $s \in (0, u)$ and $t \in (u, 1)$, we obtain

$$\Phi_i(u)(1 - \Phi_j(u)) - \Phi_j(u)(1 - \Phi_i(u)) = \Phi_i(u) - \Phi_j(u) \geq 0,$$

$u \in [0, 1]$. Therefore, we obtain the stochastic order,

$$\Phi_1(u) \geq \Phi_2(u) \geq \cdots \geq \Phi_k(u)$$

for all $u \in [0, 1]$. Combining this result with Lemma 1, we obtain the inequality

$$C^*(u, v) := C(u, v; Q) \geq C(u, v; R)$$

for all $u, v \in [0, 1]$ and R satisfying (18). The theorem now follows from Hoeffding's covariance formula,

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \big[\Pr(X \leq x, Y \leq y) - \Pr(X \leq x) \Pr(Y \leq y) \big] \, \mathrm{d}x \mathrm{d}y$$

(see, e.g., [15]). The proof is complete. □

Functions ϕ_k satisfying (15) and (16) can be constructed by B-spline functions as we now show. Let N_i^d be a B-spline function on $[0, 1]$ of degree d (≥ 0) defined as a non-zero B-spline basis with $m + 2d + 1$ knots:

$$\underbrace{t_{-d} = \cdots = t_{-1}}_{d} = t_0 = 0 < t_1 < \cdots < t_{m-1} < 1 = t_m = \underbrace{t_{m+1} = \cdots = t_{m+d}}_{d}. \tag{20}$$

Then, $N_i^d(t)$ is generated by the recursion formula,

$$\begin{aligned} N_i^d(t) &= \frac{t - t_i}{t_{i+d} - t_i} N_i^{d-1}(t) + \frac{t_{i+d+1} - t}{t_{i+d+1} - t_{i+1}} N_{i+1}^{d-1}(t) \\ &= \frac{t - t_i}{t_{i+d} - t_i} N_i^{d-1}(t) + \left(1 - \frac{t - t_{i+1}}{t_{i+d+1} - t_{i+1}}\right) N_{i+1}^{d-1}(t), \end{aligned}$$

$t \in [0, 1]$, for $i = -d, \ldots, -1, 0, 1, \ldots, m - 1$, with initial conditions

$$N_i^0(t) = \begin{cases} 1, & i < m \text{ and } t \in [t_i, t_{i+1}), \\ & \text{or } i = m - 1 \text{ and } t = t_m = 1, \\ 0, & \text{otherwise} \end{cases}$$

(see [2],[4] and [19]). The number of non-zero bases is

$$n = m + d.$$

The B-spline is known to satisfy

(i) $N_i^d(t) \geq 0,\ t \in [0,1]$,

(ii) The support is given by

$$\operatorname{supp} N_i^d = \overline{\{t \mid N_i^d(t) > 0\}} = [t_i, t_{i+d+1}],$$

$i = -d, \ldots, -1, 0, 1, \ldots, m-1$, and

(iii) The "partition of unity" property:

$$\sum_{i=-d}^{m-1} N_i^d(t) = 1 \quad \text{for all } t \in [0,1].$$

For given d and m, let

$$q_k = q_{k,d} = \int_0^1 N_{k-d-1}^d(t)\,\mathrm{d}t, \quad \phi_k(t) = \phi_{k,d}(t) = \frac{1}{q_k} N_{k-d-1}^d(t), \tag{21}$$

where $t \in [0,1]$ and $k = 1, 2, \ldots, n$ $(= m+d)$. Then, (15) holds, and we have the following result (see [3], or [21, Theorems 4.18 and 4.65]).

Theorem 5 *Under the hypotheses (20) and (21), the set $\{N_i^d\}_{i=-d}^{m-1}$ of B-spline functions, and hence also the B-spline system $\{\phi_1, \ldots, \phi_n\}$, forms an OCWT-system satisfying (16).*

Theorem 6 *Let $m = 1$ and the degree $d = n-1 (= n-m)$. Then the B-splines (21) reduce to the Bernstein system (5). Specifically, for $k = 1, \ldots, n$ and $t \in [0,1]$,*

$$q_k = q_{k,d} = \frac{1}{n}, \quad \phi_k(t) = \phi_{k,d}(t) = b_{k,n}(t).$$

This implies that the B-spline copula include the Bernstein copula as a special case.

From now on, for simplicity, we consider only the B-spline with equally-spaced knots, i.e., the B-spline functions on $[0,1]$ of order d having knots given in (20) with $t_i = i/m$, $i = 1, 2, \ldots, m-1$.

4.3. Range of correlation of the maximum correlation copula

For copula functions, the range of the correlation is of particular importance. In particular, great attention is paid to the maximum achievable correlation (see, e.g., Lin and Huang [16]). By Theorem 4, the maximum is attained when the copula density is

$$c^*(u,v) = \sum_{k=1}^{n} q_k \phi_k(u) \phi_k(v), \quad u, v \in [0,1]. \tag{22}$$

Suppose that (U,V) is from the copula density (22). Then,

$$E[UV] = \sum_{k=1}^{n} q_k \left(\int_0^1 u \phi_k(u)\,\mathrm{d}u \right)^2.$$

Noting that $E[U] = E[V] = 1/2$ and $\operatorname{Var}(U) = \operatorname{Var}(V) = 1/12$, it follows that

$$\operatorname{corr}(U,V) = 12\Big(E[UV] - \frac{1}{4}\Big).$$

In the Bernstein case ($m = 1$), it follows from Theorem 6 that $E[UV] = (2n+1)/6(n+1)$ and hence

$$\mathrm{corr}(U, V) = 1 - \frac{2}{n+1}.$$

In order to calculate the maximum correlation for general d, we present first a lemma in which it is understood that the vectors (q_k) and (r_k) reduce to the central parts when $d = 0$.

Lemma 2 *Suppose that $m \geq d \geq 0$, i.e., $n = m + d \geq 2d \geq 0$. Let N_i^d, $i = -d, -d+1, \ldots, m-1$, be the B-spline functions on $[0, 1]$ of order d having knots (20) with $t_i = i/m$, $i = 0, 1, \ldots, m$. In addition, denote the integral and the first moment of N_{k-d-1}^d by*

$$q_k = \int_0^1 N_{k-d-1}^d(t)\,\mathrm{d}t \quad \text{and} \quad r_k = \int_0^1 t N_{k-d-1}^d(t)\,\mathrm{d}t,$$

$k = 1, \ldots, n$. Then,

$$(q_k)_{1\leq k\leq n} = \frac{1}{m}\Bigg(\underbrace{\frac{1}{d+1}, \frac{2}{d+1}, \ldots, \frac{d}{d+1}}_{d}, \underbrace{1, \ldots, 1}_{m-d}, \underbrace{\frac{d}{d+1}, \frac{d-1}{d+1}, \ldots, \frac{1}{d+1}}_{d}\Bigg),$$

$$(r_k)_{1\leq k\leq n} = \frac{1}{m^2}\Bigg(\underbrace{\frac{1^2(1+1)}{2(d+1)(d+2)}, \frac{2^2(2+1)}{2(d+1)(d+2)}, \ldots, \frac{d^2(d+1)}{2(d+1)(d+2)}}_{d},$$
$$\underbrace{\frac{d+1}{2}, \frac{d+3}{2}, \ldots, \frac{2m-1-d}{2}}_{m-d}, \underbrace{m^2(q_d - r_d), \ldots, m^2(q_1 - r_1)}_{d}\Bigg).$$

Theorem 7 *Under the assumptions of Lemma 2, suppose that (U, V) have the copula density c^* in (22) with ϕ_k defined through the B-spline functions (21) having knots given in Lemma 2. Then the correlation of (U, V) is*

$$\mathrm{corr}(U, V) = 1 - \frac{d+1}{(n-d)^2} + \frac{d(d+3)(2d+3)}{5(d+2)(n-d)^3}.$$

Theorem 8 *Let C^* be the maximum correlation copula function (19) that is constructed by the B-spline*

$$\{N_{k-d-1}^d\}_{k=1}^n = \{N_i^d : i = -d, -d+1, \ldots, m-2, m-1\}$$

on $[0, 1]$ of degree $d \geq 0$, having equally-spaced knots (20) with $t_i = i/m$, $i = 0, 1, \ldots, m$, where $m \geq d$. As $m \to \infty$, $C^*(u, v) \to \min\{u, v\}$ for all u, v, the Fréchet–Hoeffding upper bound.

This property is also inherited from the Bernstein copula [8].

Table 2 shows the maximum correlations when the number of basis functions is n. In view of Table 2, the range of correlation for the B-spline copulas of small order d is wider than that of the Bernstein copula. Indeed,

$$\mathrm{corr}(U, V) \approx 1 - \frac{d+1}{n^2}.$$

Table 2: Maximum correlations

	Bernstein*	$d=0$	$d=1$	$d=2$	$d=3$
$n=2$	0.333	0.75	0.333	NA	NA
$n=3$	0.5	0.889	0.667	0.5*	NA
$n=4$	0.6	0.938	0.827	0.688	0.6*
$n=5$	0.667	0.96	0.896	0.796	0.72
$n=6$	0.714	0.972	0.931	0.867	0.796
$n=7$	0.75	0.980	0.951	0.908	0.851
$n=8$	0.778	0.984	0.963	0.933	0.892
$n=9$	0.8	0.988	0.971	0.949	0.919
$n=10$	0.818	0.99	0.977	0.960	0.937
n	$1-\dfrac{2}{n+1}$	$1-\dfrac{1}{n^2}$	$1-\dfrac{2(3n-5)}{3(n-1)^3}$	$1-\dfrac{6n-19}{2(n-2)^3}$	$1-\dfrac{2(50n-231)}{25(n-3)^3}$

*: Bernstein case ($m=n-d=1$).

4.4. Total positivity of the maximum correlation copula

The next two results improve significantly the previous ones about the Bernstein copulas.

Theorem 9 *The copula C^* in (19) is TP_∞, i.e., for any $r \geq 1$,*

$$\det\left(C^*(u_i, v_j)\right)_{r\times r} \geq 0$$

for all $u_1 > \cdots > u_r$ and $v_1 > \cdots > v_r$.

Theorem 10 *The copula density c^* in (22) is TP_∞ .*

This can be proved by using the fact that B-spline functions are the Chebyshev system. The detailed proofs and more TP_∞ properties can be found in [7] and [14].

4.5. Moments of the B-spline functions with initial boundary

In this section, we provide the moment formula for the B-spline functions with initial boundary at $t=0$ defined on $\mathbb{R}_+ = [0,\infty)$. Let N_i^d be a B-spline function of degree $d \geq 0$ on $\mathbb{R}_+$ with knots:

$$\underbrace{t_{-d} = \cdots = t_{-1}}_{d} = t_0 = 0 < t_1 = 1 < t_2 = 2 < \ \cdots \tag{23}$$

Here, we have $t_i = (i)_+ = \max\{i, 0\}$ and, $N_i^d(t)$ is generated by the following recursion formula:

$$N_i^d(t) = \frac{t-(i)_+}{(i+d)_+ - (i)_+} N_i^{d-1}(t) + \frac{(i+d+1)_+ - t}{(i+d+1)_+ - (i+1)_+} N_{i+1}^{d-1}(t), \tag{24}$$

$d \geq 1$, with initial conditions

$$N_i^0(t) = \begin{cases} 1, & i \geq 0 \text{ and } t \in [i, i+1), \\ 0, & \text{otherwise.} \end{cases}$$

For each $i \geq -d$, N_i^d is a non-zero function with support $[\max\{i,0\}, i+d+1]$. The recurrence (24) can be written more concretely as

$$N_i^d(t) = \begin{cases} \dfrac{t-i}{d} N_i^{d-1}(t) + \dfrac{i+d+1-t}{d} N_{i+1}^{d-1}(t), & i \geq 0, \\ \dfrac{t}{i+d} N_i^{d-1}(t) + \dfrac{i+d+1-t}{i+d+1} N_{i+1}^{d-1}(t), & -d < i \leq -1, \\ (1-t) N_{-d+1}^{d-1}(t), & i = -d, \\ 0, & i < -d. \end{cases}$$

For $h \geq 0$, denote the h-moment of N_i^d,

$$\gamma_i^d(h) := \int_{-\infty}^{\infty} t^h N_i^d(t)\,\mathrm{d}t = \int_{\max\{i,0\}}^{i+d+1} t^h N_i^d(t)\,\mathrm{d}t;$$

this quantity was used in the proof of Lemma 2 above. Then, we have the following recurrence relation for these moments.

$$\gamma_i^d(h) = \begin{cases} \dfrac{\gamma_i^{d-1}(h+1) - i\gamma_i^{d-1}(h)}{d} \\ \quad + \dfrac{(i+d+1)\gamma_{i+1}^{d-1}(h) - \gamma_{i+1}^{d-1}(h+1)}{d}, & i \geq 0, \\ \dfrac{\gamma_i^{d-1}(h+1)}{i+d} \\ \quad + \dfrac{(i+d+1)\gamma_{i+1}^{d-1}(h) - \gamma_{i+1}^{d-1}(h+1)}{i+d+1}, & -d < i < 0, \\ \gamma_{-d+1}^{d-1}(h) - \gamma_{-d+1}^{d-1}(h+1), & i = -d, \\ 0, & i < -d, \end{cases}$$

with boundary condition

$$\gamma_i^0(h) = \begin{cases} \dfrac{(i+1)^{h+1} - i^{h+1}}{h+1}, & i \geq 0, \\ 0, & i < 0. \end{cases}$$

The next result, which is interesting in its own right, presents the solution of the recurrence system in terms of the Stirling numbers of the second kind:

$$S(n,k) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^n.$$

Here, $S(n,0) = \delta_{n0}$, $S(n,k) = 0$ for $n < k$, and $0^0 \equiv 1$ whenever it arises. Note also that $S(n,1) = S(n,n) = 1$ and $S(n,n-1) = n(n-1)/2$.

Theorem 11 *For $d \geq 0$, the h-moment of the B-spline function N_i^d in (23) is of the* *form*

$$\gamma_i^d(h) = \begin{cases} \displaystyle\sum_{l=0}^{h} i^l \binom{h}{l} \frac{S(h+d+1-l, d+1)}{\binom{h+d+1-l}{d+1}}, & i \geq 0, \\ \displaystyle\frac{i+d+1}{d+1} \frac{S(h+i+d+1, i+d+1)}{\binom{h+d+1}{d+1}}, & -d \leq i \leq 0, \\ 0, & i < -d. \end{cases}$$

The proof is given in [7].

5. Conclusion Remark

For given marginals, Baker's distribution is constructed with order statistics and the parameter matrix (or array) which contains the dependence structure of the order statistics. For the maximum correlation case, we investigated the limiting distribution and the convergence of the product-moment as the size of the parameter (matrix or array) is big enough. Baker's distribution can be written into a copula form, and it is called the Bernstein copula. As the Bernstein copula can be considered as a finite mixture distribution, we proposed EM algorithms to estimate the parameters. With the given marginals, we then estimated the joint density functions of multivariate data sets. The Bernstein copula is expressed in terms of Bernstein functions. We know that the Bernstein function is a special case of the B-spline function in which there is no interior knot in the domain. We then could generalize the Bernstein copula, and define the B-spline copula. For the maximum correlated case of the B-spline copula, we examined the range of the correlation, proved its TP_∞ property and showed that it can reach the Fréchet–Hoeffding upper bound.

Acknowledgments

The author thanks the organizers and councilors for the opportunity to present this work at the MSJ Autumn Meeting 2019, at Kanazawa University.

References

[1] Baker, R. An order-statistics-based method for constructing multivariate distributions with fixed marginals. *Journal of Multivariate Analysis* **99**, 2312–2327 (2008)

[2] de Boor, C. On calculating with B-splines. *Journal of Approximation Theory* **6**, 50–62 (1972)

[3] de Boor, C. Total positivity of the spline collocation matrix. *Indiana University Mathematics Journal* **25**, 541–551 (1976)

[4] de Boor, C. *A Practical Guide to Splines*, Revised edn. Springer, New York (2001)

[5] Dou, X., Kuriki, S., and Lin, G. D. Dependence structures and asymptotic properties of Baker's distributions with fixed marginals. *Journal of Statistical Planning and Inference* **143**, 1343–1354 (2013)

[6] Dou, X., Kuriki, S., Lin, G. D., and Richards, D. EM algorithms for estimating the Bernstein copula. *Computational Statistics & Data Analysis* **93**, 228–245 (2014)

[7] Dou, X., Kuriki, S., Lin, G. D., and Richards, D. Dependence properties of B-spline copulas. *arXiv*:1902.04749 (2019)

[8] Huang, J. S., Dou, X., Kuriki, S., and Lin, G. D. Dependence structure of bivariate order statistics with applications to Bayramoglu's distributions. *Journal of Multivariate Analysis* **114**, 201–208 (2013)

[9] Hwang, J. S. and Lin, G. D. Characterizations of distributions by linear combinations of moments of order statistics. *Bulletin of the Institute of Mathematics, Academia Sinica* **12**, 179–202 (1984)

[10] Janssen, P., Swanepoel, J. and Veraverbeke, N. (2012). Large sample behavior of the Bernstein copula estimator, *Journal of Statistical Planning and Inference*, **142** (5), 1189–1197.

[11] Karlin, S. *Total Positivity*, Vol. 1. Stanford University Press, Palo Alto, CA (1968)

[12] Karlin, S. and Studden, W. J. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley, New York (1966)

[13] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*, Springer, New York.

[14] Lin, G. D., Dou, X., and Kuriki, S. The bivariate lack-of-memory distributions. *Sankhyā Ser. A.* 1–25 Available at `https://doi.org/10.1007/s13171-017-0119-1` (2017)

[15] Lin, G. D., Dou, X., Kuriki, S., and Huang, J. S. Recent developments on the construction of bivariate distributions with fixed marginals. *Journal of Statistical Distributions and Applications* **1**: 14 (2014)

[16] Lin, G. D. and Huang, J. S. A note on the maximum correlation for Baker's bivariate distributions with fixed marginals. *Journal of Multivariate Analysis* **101**, 2227–2233 (2010)

[17] McLachlan, G. and Krishnan, T. (2008). *The EM Algorithm and Extensions*, 2nd edn., Wiley, New York.

[18] Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd edn., Springer, New York.

[19] Nürnberger, G. *Approximation by Spline Functions*. Springer, New York (1989)

[20] Sancetta, A. and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions, *Econometric Theory*, **20** (3), 535–562.

[21] Schumaker, L. L. *Spline Functions: Basic Theory*, 3rd edn. Cambridge University Press, Cambridge (2007)

[22] Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* **8**, 229–231 (1959)

[23] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

[24] Takada, T., Mita, A., Maeno, A., Sakai, T., Shitara, H., Kikkawa, Y., Moriwaki, K., Yonekawa, H. and Shiroishi, T. (2012). Mouse inter-subspecific consomic strains for genetic dissection of quantitative complex traits, *Genome Research*, **18** (3), 500–508. The data set is available at
`http://molossinus.lab.nig.ac.jp/phenotype/index.html`

[25] Takada, T. and Shiroishi, T. (2012). Complex quantitative traits cracked by the mouse inter-subspecific consomic strains, *Experimental Animals*, **61** (4), 375–388.

[26] Tsukahara, H. (2005). Semiparametric estimation in copula models, *The Canadian Journal of Statistics*, **33** (3), 357–375. Erratum: ibid (2011), **39** (4), 734–735.