Journal of Multivariate Analysis 159 (2017) 39–48

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Smooth copula-based estimation of the conditional density function with a single covariate

Paul Janssen [a], Jan Swanepoel [b], Noël Veraverbeke [a,b,*]

[a] Hasselt University, Center for Statistics, Agoralaan, Gebouw D, 3590 Diepenbeek, Belgium
[b] North-West University, Potchefstroom Campus, Potchefstroom, South Africa

A R T I C L E I N F O

Article history:
Received 7 July 2016
Available online 26 April 2017

AMS 2000 subject classifications:
62G05
62G07
62G20

Keywords:
Asymptotic distribution
Bernstein estimation
Copula
Conditional density

A B S T R A C T

Some recent papers deal with smooth nonparametric estimators for copula functions and copula derivatives. These papers contain results on copula-based Bernstein estimators for conditional distribution functions and related functionals such as regression and quantile functions. The focus in the present paper is on new copula-based smooth Bernstein estimators for the conditional density. Our approach avoids going through separate density estimation of numerator and denominator. Our estimator is defined as a smoother of the copula-based Bernstein estimator of the conditional distribution function. We establish asymptotic properties of bias and variance and discuss the asymptotic mean squared error in terms of the smoothing parameters. We also obtain the asymptotic normality of the new estimator. In a simulation study we show the good performance of the new estimator in comparison with other estimators proposed in the literature.

1. Introduction

Let (X, Y) be a random pair with joint absolutely continuous distribution function H, marginal distribution functions F and G, and corresponding densities f and g. Following [4,11], let C denote the unique copula corresponding to H, defined for all $x, y \in \mathbb{R}$, by

$$H(x, y) = C\{F(x), G(y)\},$$

or, since F and G are continuous, $C(u, v) = H\{F^{-1}(u), G^{-1}(v)\}$, with F^{-1} and G^{-1} the quantile functions. Let $C^{(1)}(u, v) = \partial C(u, v)/\partial u$ denote the first-order partial derivative of C. Then it is easily seen that the conditional distribution function $F_x(y) = \Pr(Y \le y \mid X = x)$ is given by

$$F_x(y) = C^{(1)}\{F(x), G(y)\}. \quad (1)$$

In [9] the following Bernstein estimator for $F_x(y)$ has been studied, viz.

$$\widehat{F}_{x,m}(y) = C^{(1)}_{m,n}\{F_n(x), G_n(y)\},$$

with $C^{(1)}_{m,n}(u, v)$ the Bernstein estimator for $C^{(1)}(u, v)$ defined as $C^{(1)}_{m,n}(u, v) = \partial C_{m,n}(u, v)/\partial u$, where $C_{m,n}(u, v)$ is the Bernstein estimator of C. It is defined as

$$C_{m,n}(u, v) = \sum_{k=0}^{m} \sum_{\ell=0}^{m} C_n\left(\frac{k}{m}, \frac{\ell}{m}\right) P_{m,k}(u) P_{m,\ell}(v),$$

* Corresponding author at: Hasselt University, Center for Statistics, Agoralaan, Gebouw D, 3590 Diepenbeek, Belgium.
E-mail addresses: paul.janssen@uhasselt.be (P. Janssen), jan.swanepoel@nwu.ac.za (J. Swanepoel), noel.veraverbeke@uhasselt.be (N. Veraverbeke).

http://dx.doi.org/10.1016/j.jmva.2017.04.008
0047-259X/

with, for all $k \in \{0, \ldots, m\}$ and $u \in [0, 1]$,

$$P_{m,k}(u) = \binom{m}{k} u^k (1-u)^{m-k},$$

the binomial probabilities and C_n the empirical copula estimator given for a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from (X, Y) by

$$C_n(u, v) = H_n\{F_n^{-1}(u), G_n^{-1}(v)\},$$

where

$$H_n(x, y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le x, Y_i \le y), \qquad F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le x), \qquad G_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(Y_i \le y),$$

with $\mathbf{1}$ the indicator function.

From (1) it follows that the conditional density function $f_x(y)$ of Y given $X = x$ leads to the following expression

$$f_x(y) = c\{F(x), G(y)\}g(y), \tag{2}$$

where c is the copula density function $c(u, v) = \partial^2 C(u, v)/\partial u \partial v$. Based on the representation in (2), Faugeras [3] proposed the following rank-based kernel-type estimator for $f_x(y)$

$$\widehat{f}^{0}_{x,h,b}(y) = \left[\frac{1}{nh_n^2}\sum_{i=1}^{n} K_1\left\{\frac{F_n(x) - F_n(X_i)}{h_n}\right\} K_2\left\{\frac{G_n(y) - G_n(Y_i)}{h_n}\right\}\right]\widehat{g}(y), \tag{3}$$

where $\widehat{g}$ is the Parzen–Rosenblatt kernel estimator of g given by

$$\widehat{g}(y) = \frac{1}{nb_n}\sum_{i=1}^{n} K_0\left(\frac{y - Y_i}{b_n}\right),$$

for given kernels K_0, K_1, K_2 and bandwidths h_n and b_n. Faugeras [3] studied the large-sample properties of this estimator and also provided some simulation results which showed that it compares favorably to its competitors.

A natural Bernstein-type estimator for $f_x(y)$ based on the copula-based expression in (2) could be proposed as

$$\widehat{f}^{*}_{x,m,b}(y) = c_{m,n}\{F_n(x), G_n(y)\}\widehat{g}(y),$$

where $c_{m,n}(u, v) = \partial^2 C_{m,n}(u, v)/\partial u \partial v$ is the Bernstein copula-density estimator studied, among others, in [8].

Both $\widehat{f}^{0}_{x,h,b}(y)$ and $\widehat{f}^{*}_{x,m,b}(y)$ are, however, not smooth estimators (as a function of y), because of the jumps caused by $G_n(y)$. In order to obtain continuous estimators for $f_x(y)$, further smoothing is required, leading to three smoothing parameters: h_n and b_n for $\widehat{f}^{0}_{x,h,b}$ (m and b_n for $\widehat{f}^{*}_{x,m,b}$) and a bandwidth for the final smoothing in each case.

Our new copula-based Bernstein estimator for the conditional density only needs two smoothing parameters m and b_n. It is defined as a kernel smoother of the Bernstein estimator $\widehat{F}_{x,m}$, viz.

$$\begin{aligned}
\widehat{f}_{x,m,b}(y) &= \frac{1}{b_n}\int K_0\left(\frac{y-s}{b_n}\right)\widehat{F}_{x,m}(ds) = \frac{1}{b_n}\int K_0\left(\frac{y-s}{b_n}\right) c_{m,n}\{F_n(x), G_n(s)\}dG_n(s) \\
&= \frac{1}{nb_n}\sum_{i=1}^{n} K_0\left(\frac{y - Y_{(i)}}{b_n}\right) c_{m,n}\left\{F_n(x), \frac{i}{n}\right\},
\end{aligned}$$

where $Y_{(1)} \le \cdots \le Y_{(n)}$ are the ordered $Y_1, \ldots, Y_n$.

Formal expressions for $\widehat{F}_{x,m}$ are given by

$$\begin{aligned}
\widehat{F}_{x,m}(y) &= C^{(1)}_{m,n}\{F_n(x), G_n(y)\} = \sum_{k=0}^{m}\sum_{\ell=0}^{m} C_n\left(\frac{k}{m}, \frac{\ell}{m}\right) P'_{m,k}\{F_n(x)\}P_{m,\ell}\{G_n(y)\} \\
&= m\sum_{k=0}^{m-1}\sum_{\ell=0}^{m}\left\{C_n\left(\frac{k+1}{m}, \frac{\ell}{m}\right) - C_n\left(\frac{k}{m}, \frac{\ell}{m}\right)\right\} P_{m-1,k}\{F_n(x)\}P_{m,\ell}\{G_n(y)\}.
\end{aligned}$$

To advocate Bernstein estimators, it is good to recall the advantage compared to standard kernel estimators of having a smaller order of the asymptotic variance. A further advantage is that the order of the bias is uniform and hence the Bernstein estimates are free of boundary effects; see [7,9,10]. In Section 2 we study the asymptotic properties (bias, variance, asymptotic distribution) of the new conditional density estimator $\widehat{f}_{x,m,b}$ and compare the new proposal with the Faugeras estimator (3) and other competitors.

In Section 3 we present, for a variety of copulas, a convincing finite-sample comparison with competitors of $\widehat{f}_{x,m,b}$. The comparison clearly shows that our copula-based estimator of the conditional density outperforms its competitors.

2. Asymptotic properties of the conditional density estimator

The following theorem describes the asymptotic behavior of the new estimator $\widehat{f}_{x,m,b}$.

Theorem. *Assume that the following conditions hold.*

(a) *The copula C has second-order partial derivatives that are Lipschitz in $(0, 1)^2$.*
(b) *$g(y)$ exists.*
(c) *K_0 is a continuous probability density function of bounded variation with bounded support $[-L, L]$, $K_0(-L) = K_0(L) = 0$ and*

$$\mu_1(K_0) = \int_{-L}^{L} tK_0(t)dt = 0.$$

(d) *The order m and the bandwidth b_n satisfy the relations, for $n \to \infty$,*

$$n^{1/2}m^{-5/4}b_n^{-1/2} \to 0, \qquad n^{-1/2}m^{5/6}b_n^{-1/2}(\ln n)^{1/2}(\ln\ln n)^{1/2} \to 0.$$

Then, for all x and y such that $0 < F(x) < 1$ and $0 < G(y) < 1$, one has

$$(nm^{-1/2}b_n)^{1/2}\{\widehat{f}_{x,m,b}(y) - E\widehat{f}_{x,m,b}(y)\} \rightsquigarrow \mathcal{N}\left[0, \frac{\|K_0\|^2}{2\sqrt{\pi F(x)\{1-F(x)\}}}f_x(y)\right],$$

where $\|K_0\|^2 = \int K_0^2(t)dt$.

Proof. Write

$$(A) = \widehat{f}_{x,m,b}(y) - \mathrm{E}\widehat{f}_{x,m,b}(y) = \frac{1}{b_n}\int \{\widehat{F}_{x,m}(y - b_nt) - \mathrm{E}\widehat{F}_{x,m}(y - b_nt)\}dK_0(t).$$

Note that $(A) = (A1) + (A2) + (A3)$ with

$$(A1) = \frac{1}{b_n}\int \{F_x(y - b_nt) - E\widehat{F}_{x,m}(y - b_nt)\}dK_0(t),$$

$$(A2) = \frac{1}{b_n}\int [\beta_m^0\{F_n(x), G_n(y)\} - F_x(y - b_nt)]dK_0(t),$$

with $\beta_m^0(u, v)$ as in Theorem 1 of [9] and

$$(A3) = \frac{1}{b_n}\int [\widehat{F}_{x,m}(y - b_nt) - \beta_m^0\{F_n(x), G_n(y)\}]dK_0(t).$$

From (the proof of) Theorem 3 and Remark 3 in [9], it follows that

$$(A1) = O(m^{-1}b_n^{-1}), \qquad (A2) = O\{m^{-1}b_n^{-1} + b_n^{-1}n^{-1/2}(\ln\ln n)^{1/2}\},$$

and

$$(A3) = \sum_{i=1}^{n}\frac{1}{nb_n}\int T_{in}\{F(x), G(y - b_nt)\}dK_0(t) + \frac{1}{b_n}\int \widetilde{R}_n(x, y - b_nt)dK_0(t),$$

with $T_{in}(u, v)$ and $\widetilde{R}_n(x, y)$ as in Section 2, respectively Theorem 3 of [9]. Combining these findings we obtain that, for all x and y,

$$(A) = \sum_{i=1}^{n}\frac{1}{nb_n}\int T_{in}\{F(x), G(y - b_nt)\}dK_0(t) + r_{x,m,b}(y),$$

where

$$\begin{aligned} T_{in}(u, v) = {} & m\sum_{k=0}^{m-1}\sum_{\ell=0}^{m}\left\{\mathbf{1}\left(\frac{k}{m} < U_i \le \frac{k+1}{m}, V_i \le \frac{\ell}{m}\right) - \Pr\left(\frac{k}{m} < U_i \le \frac{k+1}{m}, V_i \le \frac{\ell}{m}\right)\right\}P_{m-1,k}(u)P_{m,\ell}(v) \\ & - m\,C^{(1)}(u, v)\sum_{k=0}^{m-1}\left\{\mathbf{1}\left(\frac{k}{m} < U_i \le \frac{k+1}{m}\right) - \frac{1}{m}\right\}P_{m-1,k}(u), \end{aligned}$$

with $U_i = F(X_i)$, $V_i = G(Y_i)$, and

$$r_{x,m,b}(y) = O\left\{\frac{m^{-1}}{b_n} + \frac{1}{b_n} m^{1/2} n^{-3/4} (\ln n)^{1/2} (\ln\ln n)^{1/4} + \frac{1}{b_n} n^{-1/2} (\ln n)^{1/2} + \frac{1}{b_n} m^{13/12} n^{-1} (\ln n)^{1/2} (\ln\ln n)^{1/2}\right\}. \quad (4)$$

We denote the main term in this asymptotic representation by $\sum_{i=1}^n W_{in}$. For the variance we have

$$\operatorname{var}\left(\sum_{i=1}^n W_{in}\right) = \frac{1}{nb_n^2} \iint \mathrm{E}[T_{in}\{F(x), G(y - b_n s)\} T_{in}\{F(x), G(y - b_n t)\}] dK_0(s) dK_0(t).$$

From the calculations in Section 4 of [9] we have that

$$\mathrm{E}[T_{in}\{F(x), G(y - b_n s)\} T_{in}\{F(x), G(y - b_n t)\}] \sim m^{1/2} \frac{1}{2\sqrt{\pi F(x)\{1 - F(x)\}}} [F_x\{\min(y - b_n s, y - b_n t)\} - F_x(y - b_n s) F_x(y - b_n t)],$$

and hence

$$\operatorname{var}\left(\sum_{i=1}^n W_{in}\right) \sim \frac{m^{1/2}}{nb_n^2} \frac{J}{2\sqrt{\pi F(x)\{1 - F(x)\}}},$$

where

$$J = \int \left[\int [F_x\{\min(y - b_n s, y - b_n t)\} - F_x(y - b_n s) F_x(y - b_n t)] dK_0(s)\right] dK_0(t).$$

The inner integral is equal to

$$-\int K_0(s) \frac{\partial}{\partial s} [F_x\{\min(y - b_n s, y - b_n t)\} - F_x(y - b_n s) F_x(y - b_n t)] ds$$

and can be re-expressed successively as

$$-\int_v^L K_0(s) \frac{\partial}{\partial s} [F_x(y - b_n s)\{1 - F_x(y - b_n t)\}] \, ds - \int_{-L}^v K_0(s) \frac{\partial}{\partial s} \{F_x(y - b_n t)(1 - F_x(y - b_n s))\} \, ds$$

and

$$b_n \int_v^L K_0(s) f_x(y - b_n s)\{1 - F_x(y - b_n t)\} ds - b_n \int_{-L}^v K_0(s) f_x(y - b_n s) F_x(y - b_n t) ds.$$

Hence, by partial integration,

$$\begin{aligned} J &= -b_n \int K_0(t) \left[-K_0(t) f_x(y - b_n t)\{1 - F_x(y - b_n t)\} + b_n \left\{\int_t^L K_0(s) f_x(y - b_n s) ds\right\} f_x(y - b_n t)\right] dt \\ &\quad + b_n \int K_0(t) \left[-K_0(t) f_x(y - b_n t) F_x(y - b_n t) - b_n \left\{\int_{-L}^t K_0(s) f_x(y - b_n s) ds\right\} f_x(y - b_n t)\right] dt \\ &\sim b_n \int K_0^2(s) f_x(y - b_n s) ds \sim b_n \left\{\int K_0^2(s) ds\right\} f_x(y) = b_n \|K_0\|^2 f_x(y). \end{aligned}$$

Therefore,

$$\operatorname{var}\left(\sum_{i=1}^n W_{in}\right) \sim \frac{m^{1/2}}{nb_n} \frac{\|K_0\|^2}{2\sqrt{\pi F(x)\{1 - F(x)\}}} f_x(y).$$

Note that the convergence to zero of the factor $m^{1/2}/(nb_n)$ is implied by the first relation in condition (d). To show the asymptotic normality of $\sum_{i=1}^n W_{in}$ we check the Liapunov condition, viz.

$$\lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}(W_{in}^4) \Big/ \left\{\operatorname{var}\Big(\sum_{i=1}^n W_{in}\Big)\right\}^2 = 0.$$

Now,

$$\mathrm{E}(W_{in}^4) = \frac{1}{n^4 b_n^4} \iint \iint \mathrm{E}[T_{in}\{F(x), G(y-b_n s)\}T_{in}\{F(x), G(y-b_n t)\}$$
$$\times T_{in}\{F(x), G(y-b_n s')\}T_{in}\{F(x), G(y-b_n t')\}]dK_0(s)dK_0(t)dK_0(s')dK_0(t') = \frac{1}{n^4 b_n^4} O(m^{3/2}).$$

Hence, the Liapunov ratio becomes

$$O\left(\frac{1}{n^3}\frac{1}{b_n^4}m^{3/2}\right) \Big/ \left\{O\left(\frac{m^{1/2}}{nb_n}\right)\right\}^2 = O\left(\frac{m^{1/2}}{nb_n^2}\right) \to 0$$

because of the relations in condition (d).

It remains to check that the four O-terms in the remainder of (4) vanish after multiplication with the norming factor $(nm^{-1/2}b_n)^{1/2}$. This requires the following four conditions:

(i) $n^{1/2}m^{-5/4}b_n^{-1/2} \to 0$,
(ii) $n^{-1/4}m^{1/4}b_n^{-1/2}(\ln n)^{1/2}(\ln\ln n)^{1/4} \to 0$,
(iii) $m^{-1/4}b_n^{-1/2}(\ln n)^{1/2} \to 0$,
(iv) $n^{-1/2}m^{5/6}b_n^{-1/2}(\ln n)^{1/2}(\ln\ln n)^{1/2} \to 0$.

Some careful calculations show that (i) and (iv) imply (ii) and (iii) so that only restrictions (i) and (iv) remain in condition (d). □

Remark 1. It is easily verified that the choices $m = O(n^{2/5+\varepsilon})$ ($\varepsilon > 0$, small) and $b_n = O(n^{-\varepsilon'})$($\varepsilon' > 0$, small) minimize the factor $m^{1/2}/(nb_n)$ in the asymptotic variance. For these choices we have

$$\widehat{f}_{x,m,b}(y) - \mathrm{E}\widehat{f}_{x,m,b}(y) = O_P(n^{-2/5+\varepsilon/4+\varepsilon'/2}).$$

Taking $\varepsilon > 2\varepsilon'/5$ and $\varepsilon < 1/5 - 3\varepsilon'/5$ (needed to satisfy the conditions) we get an order very close to $O_P(n^{-2/5})$ which is better than the usual $O_P(n^{-1/3})$; see below.

Remark 2. The centering term $\mathrm{E}\widehat{f}_{x,m,b}(y)$ in the theorem can be further expanded if conditions (a) and (b) are replaced by the stronger conditions

(a′) The copula C has 4th-order partial derivatives.
(b′) $g(y)$ is twice continuously differentiable in an open neighborhood I of y and $\partial^4 C(u, v)/\partial u \partial v^3$ exists and is continuous on $(u, v) \in (0, 1) \times G(I)$.

By Remark 3 in [9] we have

$$\begin{aligned}\mathrm{E}\widehat{f}_{x,m,b}(y) &= \frac{1}{b_n}\int \mathrm{E}\widehat{F}_{x,m}(y-b_n t)dK_0(t)\\ &= \frac{1}{b_n}\int \left[F_x(y-b_n t) + \frac{1}{2m}b\{F(x), G(y-b_n t)\} + o(m^{-1})\right]dK_0(t),\end{aligned} \tag{5}$$

where

$$b(u, v) = (1-2u)C^{(1,1)}(u, v) + u(1-u)C^{(1,1,1)}(u, v) + v(1-v)C^{(1,2,2)}(u, v),$$

$C^{(1,2,2)}(u, v) = \partial^3 C(u, v)/\partial u \partial v^2$ and similarly for $C^{(1,1)}, C^{(1,1,1)}$.

Applying partial integration, a Taylor expansion, condition (b′) and the fact that $K_0(-L) = K_0(L) = 0$, the first term in (5) becomes

$$\int K_0(t)f_x(y-b_n t)dt = f_x(y) + \frac{1}{2}b_n^2 f_x''(y)\mu_2(K_0) + o(b_n^2),$$

where $\mu_2(K_0) = \int t^2 K_0(t)dt$.

Similarly, the second term in (5) is equal to

$$\begin{aligned}\frac{1}{2m}\int K_0(t)b^{(2)}\{F(x), G(y-b_n t)\}g(y-b_n t)dt &= \frac{1}{2m}\int K_0(t)[b^{(2)}\{F(x), G(y)\}g(y) + o(1)]dt\\ &= \frac{1}{2m}b^{(2)}\{F(x), G(y)\}g(y) + o(m^{-1}),\end{aligned}$$

where $b^{(2)}(u, v) = \partial b(u, v)/\partial v$. Hence we conclude that

$$\mathrm{E}\widehat{f}_{x,m,b}(y) = f_x(y) + \frac{1}{2}b_n^2 f_x''(y)\mu_2(K_0) + \frac{1}{2m}b^{(2)}\{F(x), G(y)\}g(y) + o(b_n^2) + o(m^{-1}). \tag{6}$$

Remark 3. The expression (6) for the bias of the estimator together with the asymptotic variance give that the asymptotic mean squared error is of the order

$$O\left(\frac{m^{1/2}}{nb_n}\right) + O\{(b_n^2 + m^{-1})^2\}.$$

Minimizing with respect to m and b_n gives that the optimal choices are of the form $b_n = c_1 n^{-1/6}$ and $m = c_2 n^{1/3}$ for some $c_1, c_2 > 0$ and that $\widehat{f}_{x,m,b}(y) - f_x(y) = O_P(n^{-1/3})$. These 'optimal' choices do not satisfy the restrictions in condition (d) of the theorem and hence the 'optimal' convergence rate $O_P(n^{-1/3})$ cannot be obtained. It can be easily verified however that the 'suboptimal' choices $m = O(n^{6/13+\varepsilon})$, $b_n = O(n^{-2/13})$, with $\varepsilon > 0$ arbitrary, do satisfy condition (d) and lead to a convergence rate $O_P(n^{-4/13+\varepsilon/4})$ which is close the optimal rate $O_P(n^{-1/3})$.

Corollary. *Assume conditions* (a′), (b′), (c), (d) *and the extra condition* $nm^{-1/2}b_n^5 \to C_1$ *for some finite constant* $C_1 \geq 0$. *Then, for all* x *and* y *such that* $0 < F(x) < 1$ *and* $0 < G(y) < 1$, *one has*

$$(nm^{-1/2}b_n)^{1/2}\{\widehat{f}_{x,m,b}(y) - f_x(y)\} \rightsquigarrow \mathcal{N}\left[b_x(y), \frac{\|K_0\|^2}{2\sqrt{\pi F(x)\{1-F(x)\}}} f_x(y)\right],$$

where $b_x(y) = C_1^2 f_x''(y)\mu_2(K_0)/2$. *Indeed, it can be verified that the expression* (6) *for the bias has the limit* $b_x(y)$ *after multiplication with the factor* $(nm^{-1/2}b_n)^{1/2}$. *This is due to the extra condition and the fact that* $(nm^{-1/2}b_n)^{1/2}/m \leq n^{1/2}m^{-5/4}b_n^{-1/2} \to 0$ *by the first relation in (d).* □

We now discuss the results derived in the theorem in relation to similar results proved in the literature. The Bernstein estimator for the conditional density function provides an alternative for the classical nonparametric kernel estimator of Rosenblatt [13]; see also [6]. It is given by

$$\widehat{f}_x^R(y) = \frac{\frac{1}{nh_nb_n}\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) K_0\left(\frac{y-Y_i}{b_n}\right)}{\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}, \tag{7}$$

and under typical conditions (e.g., $nh_nb_n \to \infty$), one has

$$(nh_nb_n)^{1/2}\{\widehat{f}_x^R(y) - f_x(y) + O(h_n^2 + b_n^2)\} \rightsquigarrow \mathcal{N}\left[0, \frac{f_x(y)}{f(x)}\|K\|^2\|K_0\|^2\right],$$

for all x such that $f(x) > 0$. The $O(h_n^2 + b_n^2)$ term above is a bias term (not made explicit here) which can be deleted if $nh_n^5b_n \to 0$ and $nh_nb_n^5 \to 0$. The bandwidths minimizing the asymptotic MSE are $h_n = O(n^{-1/6})$ and $b_n = O(n^{-1/6})$ and for these optimal bandwidths we have that $\widehat{f}_x^R(y) - f_x(y) = O_P(n^{-1/3})$.

A nonparametric estimator using local polynomial approximation in the y-direction has been studied in Fan et al. [1]. The local linear estimator is defined as (see also [2]):

$$\widehat{f}_x^{LL}(y) = \frac{1}{nh_nb_n}\sum_{i=1}^n K_n\left(\frac{x-X_i}{h_n}; x\right) K_0\left(\frac{y-Y_i}{b_n}\right), \tag{8}$$

with

$$K_n(z; x) = K(z)\frac{s_{n,2}(x) - zh_ns_{n,1}(x)}{s_{n,0}(x)s_{n,2}(x) - s_{n,1}^2(x)},$$

where, for $j \in \{0, 1, 2\}$,

$$s_{n,j}(x) = \frac{1}{nh_n}\sum_{i=1}^n (x - X_i)^j K\left(\frac{x-X_i}{h_n}\right).$$

The asymptotic order properties of bias and variance are the same as for the Rosenblatt estimator.

As far as the rank-based estimator $\widehat{f}_{x,h,b}^0$ is concerned, Faugeras [3] derived the asymptotic normality result

$$(nh_n^2)^{1/2}\{\widehat{f}_{x,h,b}^0(y) - f_x(y) + O(h_n^2)\} \rightsquigarrow \mathcal{N}[0, g(y)f_x(y)\|K_1\|\,\|K_2\|].$$

Again, by applying the asymptotic bias and variance expressions of [10], we have that the optimal rate $\widehat{f}_{x,h,b}^0(y) - f_x(y) = O_P(n^{-1/3})$ is obtained if $h_n = O(n^{-1/6})$.

Performing calculations similar to those in the proof of the theorem and applying results derived in [8] we can prove, under certain regularity conditions placed on C, F, G, the kernels and bandwidths (e.g., $m/n \to 0$ and $mb_n \to \infty$), that

$$(n/m)^{1/2}\{\widehat{f}^*_{x,m,b}(y) - f_x(y) + O(b_n^2 + m^{-1})\} \rightsquigarrow \mathcal{N}\left[0, \frac{g(y)f_x(y)}{4\pi\sqrt{F(x)\{1-F(x)\}G(y)\{1-G(y)\}}}\right].$$

Choosing any b_n such that $b_n = O(m^{-\alpha})$, $1/2 \leqq \alpha < 1$, we conclude that the optimal rate $\widehat{f}^*_{x,m,b}(y) - f_x(y) = O_P(n^{-1/3})$ holds if $m = O(n^{1/3})$.

It is interesting to discuss the comparison of the asymptotic variances of the various estimators. The order of magnitude of the Bernstein-based estimator $\widehat{f}_{x,m,b}(y)$ is $O\{(nm^{-1/2}b_n)^{-1}\}$. This is smaller than that of the competitors where we typically have $O\{(nh_nb_n)^{-1}\}$ (making the usual identification $h_n = m^{-1}$, as proposed in [10] and [14]). This improved order is due to the Bernstein method.

All estimators have $f_x(y)$ in the numerator of the asymptotic variance. Our Bernstein estimator $\widehat{f}_{x,m,b}(y)$ and also $\widehat{f}^*_{x,m,b}(y)$ have $\sqrt{F(x)\{1-F(x)\}}$ in the denominator instead of the typical $f(x)$ for the Rosenblatt $\widehat{f}^R_x(y)$ and local linear $\widehat{f}^{LL}_x(y)$. A comparison between these two factors can be made using a result of Parzen [12]; see also Remark 5 in [9]. It follows that $\sqrt{F(x)\{1-F(x)\}}$ is asymptotically (as $x \to \infty$) larger than $f(x)$ for all F with medium tails (e.g., exponential, logistic, Weibull, normal) and long tails (e.g., Cauchy, Pareto). A similar statement holds for the left tails. An exception is the estimator $\widehat{f}^0_{x,h,b}(y)$ of Faugeras [3] and also its Beta kernel version (on p. 2091 of [3]). In these asymptotic variances there is no $f(x)$ nor $\sqrt{F(x)\{1-F(x)\}}$ in the denominator but only $g(y)$ in the numerator. This 'advantage', mentioned in [10], p. 2091, did not show up in our simulations.

3. Monte Carlo simulations

In order to get a closer look at the finite-sample performance of an estimator $\widehat{f}_x(y)$ of $f_x(y)$, we compare the integrated squared error (ISE) at a fixed x on a set of y-values in an interval $[a, b]$, viz.

$$\mathrm{ISE}(x) = \int_a^b \{\widehat{f}_x(y) - f_x(y)\}^2 dy,$$

for the different estimators. The value of $\mathrm{ISE}(x)$ is computed by approximation on a grid of $N = 50$ values (y_j), equally spaced of Δ-length on the interval $[a, b]$,

$$I(x) = \Delta \sum_{j=1}^{N} \{\widehat{f}_x(y_j) - f_x(y_j)\}^2,$$

where $y_j = a+(b-a)(j-1)/(N-1)$ for all $j \in \{1, \ldots, N\}$, and $\Delta = (b-a)/(N-1)$. In the tables we report $AI(x)$, the average of $I(x)$ over 10,000 independent Monte Carlo replications of sample size $n = 100$, which is an accurate approximation of the mean integrated squared error, $\mathrm{MISE}(x)$. Throughout we choose $[a, b] = [-2.5, 2.5]$. The standard errors of the Monte Carlo estimates are found to be negligibly small and are not reported in the tables.

Random samples $(U_1, V_1), \ldots, (U_{100}, V_{100})$ are generated from known copula families C by means of the R package `copula`; see [5]. Data $(X_1, Y_1), \ldots, (X_{100}, Y_{100})$ are then obtained by setting $X_i = F^{-1}(U_i)$ and $Y_i = G^{-1}(V_i)$ for each $i \in \{1, \ldots, 100\}$, for prescribed marginals F and G.

Let ρ_S and ρ_P denote Spearman's rho and Pearson's correlation coefficient, respectively. To cover a variety of settings we consider the following models. (see [11]):

Model 1: $X \sim \mathcal{N}(0, 1)$, $Y \sim \mathcal{N}(0, 1)$, C_θ is the Frank copula with $\rho_S = 0.2\,(\theta = 1.224), 0.4\,(\theta = 2.610), 0.6\,(\theta = 4.466), 0.8\,(\theta = 7.902)$.

Model 2: $X \sim \mathcal{E}(0.5)$, $Y \sim \mathcal{N}(0, 1)$, C_θ is the Gumbel copula with $\rho_S = 0.2\,(\theta = 1.156), 0.4\,(\theta = 1.381), 0.6\,(\theta = 1.757), 0.8\,(\theta = 2.582)$. Throughout, $\mathcal{E}(0.5)$ refers to the exponential distribution with mean 2.

Model 3: $X \sim \mathcal{N}(0, 1)$, $Y \sim \mathcal{N}(0, 1)$, C_θ is the Gaussian copula with $\rho_P = 0.2, 0.4, 0.6, 0.8$.

Model 4: $X \sim \mathcal{E}(0.5)$, $Y \sim \mathcal{N}(0, 1)$, C_θ is the Clayton copula with $\rho_S = 0.2\,(\theta = 0.310), 0.4\,(\theta = 0.758), 0.6\,(\theta = 1.507), 0.8\,(\theta = 3.188)$.

Model 5: $X \sim \mathcal{N}(0, 1)$, $Y \sim \mathcal{N}(0, 1)$, C_θ is the Cuadras–Augé copula with $\rho_S = 0.2\,(\theta = 0.250), 0.4\,(\theta = 0.471), 0.6\,(\theta = 0.667), 0.8\,(\theta = 0.842)$. See Remark 4.

Model 6: $X \sim \mathcal{E}(0.5)$, $Y \sim \mathcal{N}(0, 1)$, $C_\theta(u, v)$ is the Plackett copula with $\rho_S = 0.2\,(\theta = 1.836), 0.4\,(\theta = 3.538), 0.6\,(\theta = 7.761), 0.8\,(\theta = 24.259)$.

Remark 4. Note that the Cuadras–Augé copula (Model 5) has an absolutely continuous and a singular component. Indeed $\Pr(U = V) = \theta/(2-\theta)$ and $\Pr(U < V) + \Pr(U > V) = (2-2\theta)/(2-\theta)$; see [11], p. 54. The corresponding copula 'density' therefore only captures a total probability weight of $(2-2\theta)/(2-\theta)$. In this case we need to rescale $f_x(y)$ in (2) to obtain a proper conditional density. Since $\int_0^1 c(u, v)dv = 1 - \theta u^{1-\theta}$ the rescaling factor is $k(x) = 1/[1 - \theta\{F(x)\}^{1-\theta}]$ and

Table 1
$MAI(x)$ and $BAI(x)$ for Model 1 (Frank, left column) and Model 2 (Gumbel, right column).

ρ	x	R	LL	QC_1	QC_2	BC_1	BC_2	x	R	LL	QC_1	QC_2	BC_1	BC_2
	0	**.01**	**.01**	.02	**.01**	**.01**	**.01**	1	**.01**	**.01**	.02	**.01**	**.01**	**.01**
0.2	1	**.01**	**.01**	.02	**.01**	**.01**	**.01**	2	**.01**	**.01**	.02	**.01**	**.01**	**.01**
	2	**.02**	.03	.06	**.02**	.06	.03	3	.02	.02	.02	**.01**	.02	**.01**
	0	**.01**	**.01**	.02	**.01**	**.01**	**.01**	1	**.01**	**.01**	.02	**.01**	**.01**	**.01**
0.4	1	.02	**.01**	.02	.02	.02	**.01**	2	**.01**	**.01**	.02	**.01**	.02	**.01**
	2	**.03**	.04	.06	**.03**	.06	**.03**	3	**.02**	**.02**	.03	**.02**	**.02**	**.02**
	0	.02	.02	.03	.02	.02	**.01**	1	**.02**	**.02**	**.02**	**.02**	**.02**	**.02**
0.6	1	**.02**	**.02**	.03	**.02**	.03	**.02**	2	**.02**	**.02**	.03	**.02**	**.02**	**.02**
	2	**.03**	.04	.08	**.03**	.06	.04	3	**.03**	**.03**	**.03**	**.03**	**.03**	**.03**
	0	**.04**	**.04**	**.04**	**.04**	**.04**	**.04**	1	.05	.05	**.03**	**.03**	**.03**	**.03**
0.8	1	.04	**.03**	**.03**	**.03**	**.03**	**.03**	2	.06	.05	**.03**	.04	.04	.05
	2	**.05**	**.05**	.09	**.05**	.07	**.05**	3	.07	.06	**.05**	**.05**	**.05**	.06

$f_x^{CA}(y) = k(x)c\{F(x), G(y)\}g(y)$. Further note that in this case $n = 100$ means that our estimator of $f_x^{CA}(y)$ is based on 100 sampled off-diagonal points, i.e., points related to the absolute continuous part of the copula.

We compare the performance of our estimators $\widehat{f}^*_{x,m,b}(y)$ and $\widehat{f}_{x,m,b}(y)$, referred to as BC_1 and BC_2 respectively in the tables, to that of the two estimators proposed by Faugeras [3], viz. $\widehat{f}^0_{x,h,b}(y)$ and a Beta kernel version of this estimator (see p. 2091 of [3]), indicated as QC_1 and QC_2 respectively in the tables and to that of the Rosenblatt estimator $\widehat{f}_x^R(y)$ given in (7) and the local linear estimator $\widehat{f}_x^{LL}(y)$ given in (8), referred to as R and LL respectively in the tables.

Following [3], we take $K_0 = K_1 = K_2 = \phi$, the standard normal density, in the definitions of QC_1, BC_1, BC_2, R and LL. Almost identical results are obtained for other choices, e.g., the Epanechnikov kernel. Furthermore, simulations are also conducted for the six copulas of Models 1–6 with different combinations ((Normal, Normal) and (Normal, Exponential)) of (X, Y). Similar behavior of the estimators is found as that reported in Tables 1–3.

In the practical implementation of the estimators some difficulties may arise, especially with the numerical calculation of a copula density estimator. Many copula densities become infinitely large in some corners of the unit square. Therefore, to avoid difficulties we replace $F_n(X_i)$ and $G_n(Y_i)$, for each $i \in \{1, \ldots, n\}$, by

$$\tilde{F}_n(X_i) = \frac{n}{n+1} F_n(X_i), \qquad \tilde{G}_n(Y_i) = \frac{n}{n+1} G_n(Y_i),$$

and $F_n(x)$, $G_n(y)$, for $x, y \in \mathbb{R}$, by

$$\widehat{F}_n(x) = \frac{1}{n+1} + \frac{n-1}{n+1} F_n(x), \qquad \widehat{G}_n(y) = \frac{1}{n+1} + \frac{n-1}{n+1} G_n(y).$$

Nonparametric estimators depend critically on the choice of the bandwidths. For conditional density estimation, bandwidth selection is a more delicate matter than for density estimation, due to the multidimensional nature of the problem. We leave a detailed study of a practical data-dependent method for bandwidth selection of the copula-based conditional density estimators for future research. Hence, we estimate b_n by the simple normal reference plug-in method (see, e.g., [15]) defined by $\hat{b}_n = 1.059\hat{\sigma}_Y n^{-1/5}$, where $\hat{\sigma}_Y$ is the sample standard deviation of $Y_1, \ldots, Y_{100}$. Furthermore, we choose a grid of m-values

$$A_n = \{m_k : m_k = \lfloor kn/8 \rfloor,\ k = 1, \ldots, 8\},$$

where $\lfloor a \rfloor$ denotes the integral part of a and compute $MAI(x) = \min_{m \in A_n} AI(x)$, for estimators BC_1 and BC_2. Likewise, we consider a grid of h_n-values

$$B_n = \{h_k : h_k = 1/m_k,\ k = 1, \ldots, 8\} \cup \{h_\ell : h_\ell = \lfloor 8/n \rfloor + \ell\Delta,\ \ell = 1, \ldots, L\}$$

with $\Delta = 0.04$ and $L = 14$ and calculate $BAI(x) = \min_{h \in B_n} AI(x)$, for estimators QC_1, QC_2, R and LL. Compared to A_n, the grid B_n contains more points to make sure that h_n can be chosen so that $\min_{h \in B_n} AI(x)$ is close to the bandwidth that makes $MISE(x)$ minimal. For comparison purposes between Bernstein and kernel estimators, it is natural to take $h = 1/m$ as the 'bandwidth'; see [10] p. 922, and [14] p. 549. The values of $MAI(x)$ and $BAI(x)$ are reported in the tables.

The results in Tables 1–3 illustrate the good performance of the new estimator $\widehat{f}_{x,m,b}(y)$ (BC_2 in the tables) when compared to five competitors. To have a fair comparison we give in Table 4 how often (for the 72 cases considered) each of the six estimators used in the comparison is better or at least as good as (in terms of $MAI(x)$ or $BAI(x)$) its competitors. From Table 4 it is clear that copula based estimators QC_2 and BC_2 have the best overall performance (82%, respectively 78%).

We therefore recommend the copula-based estimators BC_2 and QC_2 as effective nonparametric estimators for the conditional density function.

As an illustration we show in Fig. 1 the averaged performance of the six nonparametric estimators, considered in the simulation study, for the conditional density $f_x(y)$ corresponding to Model 4 (Clayton) and Model 5 (Cuadras–Augé). Note

Table 2
$MAI(x)$ and $BAI(x)$ for Model 3 (Gaussian, left column) and Model 4 (Clayton, right column).

ρ	x	R	LL	QC_1	QC_2	BC_1	BC_2	x	R	LL	QC_1	QC_2	BC_1	BC_2
	0	**.01**	**.01**	.02	**.01**	**.01**	**.01**	1	**.01**	**.01**	.02	**.01**	**.01**	**.01**
0.2	1	**.01**	**.01**	.03	.02	. 02	**.01**	2	**.01**	**.01**	.02	**.01**	**.01**	**.01**
	2	**.02**	.03	.06	**.02**	.06	.03	3	.02	.02	.02	**.01**	.02	**.01**
	0	**.01**	**.01**	.02	**.01**	**.01**	**.01**	1	.02	**.01**	.02	**.01**	**.01**	**.01**
0.4	1	.02	**.01**	.05	.02	.02	.02	2	**.01**	**.01**	.02	**.01**	.02	**.01**
	2	**.03**	.04	.07	**.03**	.07	.04	3	.02	.02	.02	.02	.02	**.01**
	0	.02	**.01**	.02	**.01**	**.01**	**.01**	1	**.02**	**.02**	.03	**.02**	**.02**	**.02**
0.6	1	**.02**	**.02**	.03	**.02**	**.02**	**.02**	2	.02	**.01**	.02	.02	.02	.02
	2	**.05**	**.05**	.09	.06	.09	.06	3	**.02**	**.02**	**.02**	**.02**	**.02**	**.02**
	0	**.03**	**.03**	**.03**	**.03**	**.03**	**.03**	1	.06	.06	**.04**	.05	**.04**	.05
0.8	1	.04	**.03**	**.03**	**.03**	.04	**.03**	2	.03	.03	**.02**	.03	.03	.03
	2	.07	**.06**	.12	.10	.15	.09	3	.03	**.02**	**.02**	**.02**	.03	**.02**

Table 3
$MAI(x)$ and $BAI(x)$ for Model 5 (Cuadras–Augé, left column) and Model 6 (Plackett, right column).

ρ	x	R	LL	QC_1	QC_2	BC_1	BC_2	x	R	LL	QC_1	QC_2	BC_1	BC_2
	0	**.01**	**.01**	.02	**.01**	**.01**	**.01**	1	**.01**	**.01**	.02	**.01**	**.01**	**.01**
0.2	1	**.01**	**.01**	.03	**.01**	**.01**	**.01**	2	**.01**	**.01**	.02	**.01**	**.01**	**.01**
	2	**.02**	**.02**	.06	**.02**	.03	**.02**	3	.02	.02	.02	**.01**	.02	**.01**
	0	**.01**	**.01**	.02	**.01**	**.01**	**.01**	1	**.01**	**.01**	.02	**.01**	.02	**.01**
0.4	1	**.01**	**.01**	.02	**.01**	**.01**	**.01**	2	.02	**.01**	.02	**.01**	.02	**.01**
	2	.03	.03	.05	**.02**	.03	**.02**	3	**.02**	**.02**	**.02**	**.02**	**.02**	**.02**
	0	**.01**	**.01**	.02	**.01**	**.01**	**.01**	1	.03	.03	.03	**.02**	**.02**	**.02**
0.6	1	**.01**	**.01**	.03	**.01**	**.01**	**.01**	2	.03	**.02**	.03	**.02**	.03	**.02**
	2	.03	.03	.05	**.02**	.03	**.02**	3	.03	.03	**.02**	**.02**	.03	.03
	0	**.01**	**.01**	.02	**.01**	**.01**	**.01**	1	.08	.08	**.05**	.06	**.05**	.08
0.8	1	**.01**	**.01**	.03	.02	**.01**	**.01**	2	.08	.08	**.06**	**.06**	**.06**	.08
	2	.03	.03	.05	**.02**	.03	**.02**	3	.07	.07	**.05**	**.05**	**.05**	.07

Table 4
Frequency of best performance of the estimators.

ρ	R	LL	QC_1	QC_2	BC_1	BC_2
Model 1	9	8	2	10	5	10
Model 2	8	8	5	11	9	10
Model 3	8	10	2	8	5	7
Model 4	5	8	4	8	6	9
Model 5	9	9	0	11	8	12
Model 6	4	6	5	11	7	8
Total	43	49	18	59	40	56
Percentage	60%	68%	25%	82%	56%	78%

that Remark 4 is instrumental when we work with the Cuadras–Augé copula. The dotted curve in the plot is the true conditional density. The two other curves in the plots are averages of 50 estimated curves. Note that this averaging masks the erratic behavior of non-smooth estimators like QC_1 and QC_2.

Remark 5. Computational aspects. To estimate $\widehat{f}^*_{x,m,b}(y)$ and $\widehat{f}_{x,m,b}(y)$ we use the following formulas:

$$\widehat{f}^*_{x,m,b}(y) = \frac{m^2}{n^2 b_n} \left\{ \sum_{i=1}^{n} K_0\left(\frac{y - Y_i}{b_n}\right) \right\} \times \left[\sum_{j=1}^{n} P_{m-1,\lceil m\widetilde{F}_n(X_j)-1\rceil}\{\widetilde{F}_n(x)\} P_{m-1,\lceil m\widetilde{G}_n(Y_j)-1\rceil}\{\widetilde{G}_n(y)\} \right],$$

$$\widehat{f}_{x,m,b}(y) = \frac{m^2}{n(n+1)b_n} \times \sum_{i=1}^{n} \left[K_0\left(\frac{y - Y_{(i)}}{b_n}\right) \sum_{j=1}^{n} P_{m-1,\lceil m\widetilde{F}_n(X_j)-1\rceil}\{\widetilde{F}_n(x)\} P_{m-1,\lceil m\widetilde{G}_n(Y_j)-1\rceil}\left(\frac{i}{n+1}\right) \right],$$

with $\lceil a \rceil$ the integral part of a plus 1.

These formulas are computationally efficient, e.g., the number of summations in $\widehat{f}^*_{x,m,b}(y)$, resp. $\widehat{f}_{x,m,b}(y)$, is $O(n)$, respectively $O(n^2)$. The use of these formulas keeps the computing time under control, e.g., for sample sizes up to 500 the computing time stays below one minute for $\widehat{f}_{x,m,b}(y)$, which is the most demanding one.

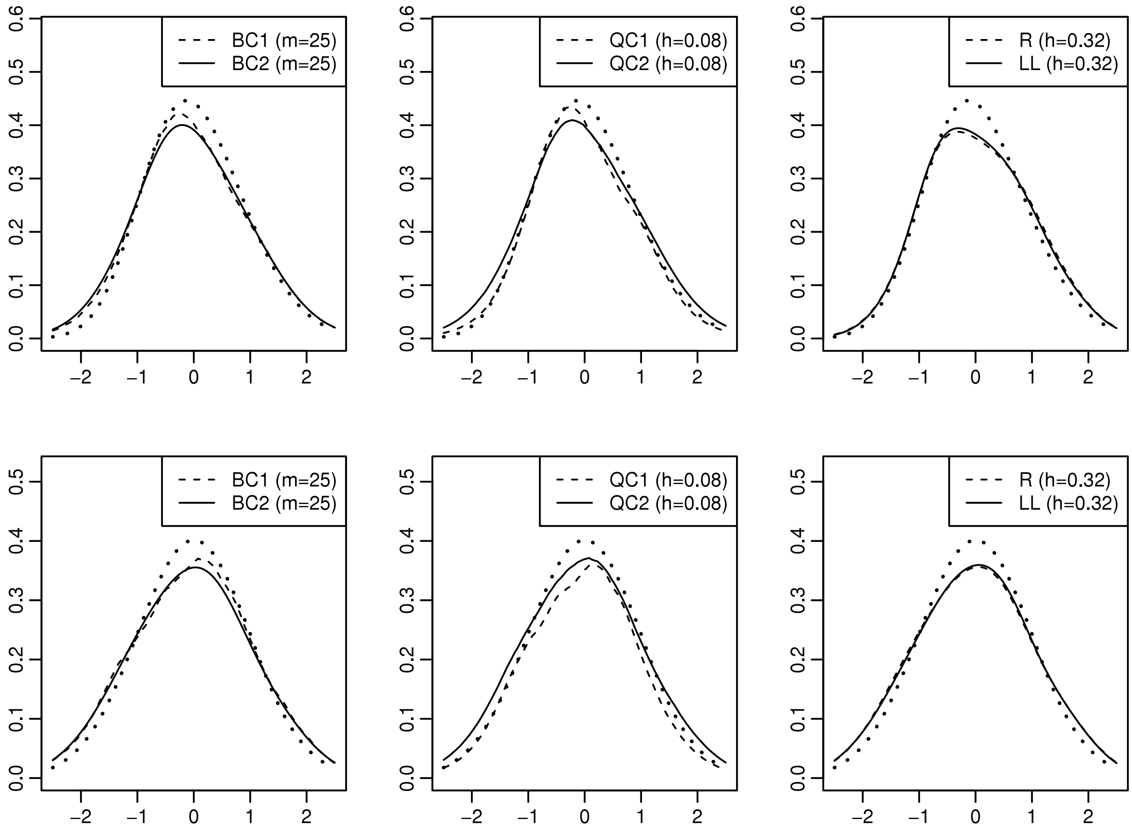

Fig. 1. Nonparametric estimators (averaged) for the conditional density corresponding (upper part) to Model 4 (Clayton copula) and (lower part) Model 5 (Cuadras–Augé copula). The dotted line is the true conditional density.

Acknowledgments

The authors thank Dr. Charl Pretorius for his important help with the simulations. They also thank the Editor, Associate Editor and a referee for their valuable comments and suggestions. The work was supported by the IAP Research Network P7/13 of the Belgian State (Belgian Science Policy). The second author thanks the National Science Foundation of South Africa for financial support (grant number 81038). The third author is also extraordinary professor at the North-West University, Potchefstroom, South Africa.

References

[1] J. Fan, Q. Yao, H. Tong, Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems, Biometrika 83 (1996) 189–206.
[2] J. Fan, T.H. Yim, A cross-validation method for estimating conditional densities, Biometrika 91 (2004) 819–834.
[3] O.P. Faugeras, A quantile-copula approach to conditional density estimation, J. Multivariate Anal. 100 (2009) 2083–2099.
[4] C. Genest, A.-C. Favre, Everything you always wanted to know about copula modeling but were afraid to ask, J. Hydrol. Eng. 12 (2007) 347–368.
[5] M. Hofert, I. Kojadinovic, M. Mächler, J. Yan, Copula: Multivariate dependence with copulas, R package version 0.999–16 URL https://CRAN.R-project.org/package=copula, (2017).
[6] R.J. Hyndman, D.M. Bashtannyk, G.K. Grunwald, Estimating and visualizing conditional densities, J. Comp. Graph. Stat. 5 (1996) 315–336.
[7] P. Janssen, J. Swanepoel, N. Veraverbeke, Large sample behavior of the Bernstein copula estimator, J. Statist. Plann. Inference 142 (2012) 1189–1197.
[8] P. Janssen, J. Swanepoel, N. Veraverbeke, A note on the asymptotic behavior of the Bernstein estimator of the copula density, J. Multivariate Anal. 124 (2014) 480–487.
[9] P. Janssen, J. Swanepoel, N. Veraverbeke, Bernstein estimation for a copula derivative with application to conditional distribution and regression functionals, TEST 25 (2016) 351–374.
[10] A. Leblanc, On estimating distribution functions using Bernstein polynomials, Ann. Inst. Statist. Math. 64 (2012) 919–943.
[11] R.B. Nelsen, An Introduction to Copulas, second ed., Springer, New York, 2006.
[12] E. Parzen, Nonparametric statistical data modeling, J. Amer. Statist. Assoc. 74 (1979) 105–121.
[13] M. Rosenblatt, Conditional probability and regression estimators, in: Multivariate Analysis II, Proc. Second International Symposium, Dayton, OH, 1968, Academic Press, New York, 1969, pp. 25–31.
[14] A. Sancetta, S. Satchell, The Bernstein copula and its applications to modeling and approximations of multivariate distributions, Econom. Theory 20 (2004) 535–562.
[15] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, London, 1986.