



HAL
open science

A note on the adaptive estimation of a bi-dimensional density in the case of knowledge of the copula density

Ingo Bulla, Christophe Chesneau, Fabien Navarro, Tanya Mark

► To cite this version:

Ingo Bulla, Christophe Chesneau, Fabien Navarro, Tanya Mark. A note on the adaptive estimation of a bi-dimensional density in the case of knowledge of the copula density. *Statistics and Probability Letters*, 2015, 105, 10.1016/j.spl.2015.02.024 . hal-00979729v2

HAL Id: hal-00979729

<https://hal.science/hal-00979729v2>

Submitted on 5 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A note on the adaptive estimation of a bi-dimensional density in the case of knowledge of the copula density

Ingo Bulla^(a), Christophe Chesneau^(b),
Fabien Navarro^(c), Tanya Mark^(d)

Received:

Abstract This paper attempts to better understand the influence of the smoothness of the copula density in the bi-dimensional estimation density problem. We provide an element of answer by studying the MISE properties of an adaptive estimator based on a plug-in approach and wavelet methods.

Keywords Bi-dimensional density estimation · Copula density · Mean integrated squared error · Wavelet methods.

2000 Mathematics Subject Classification 62G08, 62G20.

1 Motivations

We consider the common bi-dimensional density estimation problem where $(X_i, Y_i)_{i \in \mathbb{N}}$ is a sequence of independent and identically distributed random vectors defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with continuous density $h : [a, b]^2 \rightarrow [0, \infty)$ with $a < b$. The goal is to estimate h from the n random vectors $(X_1, Y_1), \dots, (X_n, Y_n)$. Numerous and various nonparametric methods exist (see, e.g., Simonof (1996) and Tsybakov (2004)).

Instead of estimating h directly, one can use the Sklar Theorem (see Sklar (1959)): there exists an unique function $c : [0, 1]^2 \rightarrow [0, \infty)$, called copula density, such that

$$h(x, y) = f(x)g(y)c(F(x), G(y)), \quad (1)$$

^(a) Theoretical Biology and Biophysics, Group T-6, Los Alamos National Laboratory, Los Alamo, NM 87545, USA E-mail: ingobulla@gmail.com

^(b) Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Basse-Normandie, Campus II, Science 3, 14032 Caen, France. E-mail: chesneau@math.unicaen.fr

^(c) Department of Mathematics and Statistics, Concordia University 1455, de Maisonneuve Blvd. W. Montreal, Quebec, Canada, H3G 1M8, E-mail: fabien.navarro@concordia.ca

^(d) University of Guelph, 50 Stone Road, Guelph, Ontario, Canada, N1G2W1 E-mail: markt@uoguelph.ca

where f is the density of X_1 , g the density of Y_1 , F is the distribution function of X_1 and G is the distribution function of Y_1 . In this expression, c measures the dependence between X_1 and Y_1 (examples of copula densities can be found in Nelsen (1999)). Then a natural estimator for $h(x, y)$ is

$$\hat{h}(x, y) = \hat{f}(x)\hat{g}(y)\hat{c}(\hat{F}(x), \hat{G}(y)), \quad (2)$$

where \hat{f} , \hat{g} , \hat{c} , \hat{F} and \hat{G} are estimators of f , g , c , F and G respectively. Such “plug-in estimator” has been investigated in, e.g., Liebscher (2005), Bouezmarni and Rombouts (2008) and Faugeras (2009) via non-adaptive kernel methods. Their mean integrated squared error (MISE) properties, uniform strong consistency and asymptotic normality was proved. Among the existing results, the following one deserves to be highlighted: if f , g are twice differentiable and c is Lipschitz continuous, the estimator developed in Liebscher (2005) is robust to the well known curse of dimensionality problem; it has similar performances to its analog for the standard univariate density estimation problem.

The following question arises: what is the precise influence of the nature of c in the estimation of h ? This paper provides an element of answer. First, assuming that c is known and satisfies a Hölder condition (including the Lipschitz one), we prove a general result on the MISE properties of (2). Then we develop an adaptive estimator based on (1), wavelets and a block thresholding technique elaborated by Hall *et al.* (1998, 1999). The great advantages of wavelet methods over the concurrences are to reach the goal of adaptivity and to be efficient for a wide class of unknown functions (see, e.g., Härdle *et al.* (1998)). We prove that our estimator attains a fast rate of convergence under the MISE over a wide class of functions for f and g (including twice differentiable functions). A consequence of our result is that whenever the dependence structure is “too complex” in terms of smoothness of c relatively to the smoothness of f and g , then the curse of dimensionality still occurs. Finally, a short simulation study illustrates the performances of our wavelet estimator. In particular, the numerical tests indicate that our block thresholding estimator compares favorably to standard kernel-based methods.

The paper is organized as follows. Section 2 is devoted to our general method and the main MISE result. Our wavelet estimator and its rate of convergence under the MISE are presented in Section 3. The simulation study can be found in Section 4. Technical proofs are collected in Section 5.

2 Upper bound for the MISE

2.1 Assumptions on the model

For any $p \geq 1$, we set

$$\mathbb{L}^p([a, b]) = \left\{ v : [0, 1] \rightarrow \mathbb{R}; \|v\|_p = \left(\int_a^b |v(x)|^p dx \right)^{1/p} < \infty \right\}.$$

We formulate the following assumptions.

(H1) There exists a constant $C_1 > 0$ such that

$$\sup_{x \in [a, b]} f(x) \leq C_1.$$

(H2) There exists a constant $C_2 > 0$ such that

$$\sup_{y \in [a, b]} g(y) \leq C_2.$$

(H3) The copula density $c : [0, 1]^2 \rightarrow [0, \infty)$ is known and there exist two constants $C_3 \geq 0$ and $\theta \in (0, 1]$ such that, for any $(x, y, u, v) \in [0, 1]^4$,

$$|c(x, y) - c(u, v)| \leq C_3(|x - u|^\theta + |y - v|^\theta).$$

The boundedness assumptions **(H1)** and **(H2)** are a bit restrictive but standard for the density estimation problem. See, e.g., Simonof (1996) and Tsybakov (2004).

The assumption **(H3)** is a Hölder condition where the parameter θ characterizes the smoothness of c . A simple example is the copula density $c(x, y) = 1$ (i.e., the independent case), where **(H3)** is satisfied with $C_3 = 0$ (and $\theta = 1$). Another less trivial example is the copula density investigated by Huang and Kotz (1999), which is an extension of the Farlie-Gumbel-Morgenstern copula, i.e.,

$$c(x, y) = 1 + v(1 - (a + 1)x^a)(1 - (a + 1)y^a), \quad (x, y) \in [0, 1]^2, \quad (3)$$

with $a \in (0, 1]$, $v \in (-1, (1/a))$, which satisfies **(H3)** with the parameter $\theta = a$. Other examples of polynomial-type copula densities satisfying **(H3)** can be found in (Balakrishnan and Lai 2009, Chapter 2).

Finally, note that it includes the Lipschitz condition on c considered in (Liebscher 2005, Assumption \mathcal{C} with $t_1 = t_2$) or (Bouezmarni and Rombouts 2008, Assumption $P1$).

2.2 MISE result

Theorem 1 shows an upper bound for the MISE of \hat{h} (2) under mild assumptions on f , g , c , \hat{f} and \hat{g} .

Theorem 1 *We consider the bi-dimensional density estimation problem described in Section 1 under **(H1)**, **(H2)** and **(H3)**, and the following estimator for h (1):*

$$\hat{h}(x, y) = \hat{f}(x)\hat{g}(y)c(\hat{F}(x), \hat{G}(y)), \quad (4)$$

where \hat{f} denotes an arbitrary estimator for f in $\mathbb{L}^2([a, b])$, \hat{g} denotes an arbitrary estimator for g in $\mathbb{L}^2([a, b])$,

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}, \quad \hat{G}(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i \leq y\}}$$

and $\mathbf{1}$ denotes the indicator function.

Then there exists a constant $C > 0$ such that

$$\begin{aligned} & \mathbb{E} \left(\int_a^b \int_a^b (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \\ & \leq C \left(\mathbb{E}(\|\hat{g} - g\|_2^2) + \mathbb{E}(\|\hat{f} - f\|_2^2) + \mathbb{E}(\|\hat{f} - f\|_2^2 \|\hat{g} - g\|_2^2) + \frac{C_3^2}{n^\theta} \right), \end{aligned}$$

(where θ refers to **H3**).

The proof of Theorem 1 is based on a suitable decomposition of $\hat{h} - h$ and moment properties of \hat{F} and \hat{G} .

Theorem 1 shows the non negligible influence of the smoothness parameter θ in the estimation of h by \hat{h} (4). In particular, if the estimators \hat{f} and \hat{g} are bounded from above (as most of the kernel density estimators), we have

$$\begin{aligned} & \mathbb{E} \left(\int_a^b \int_a^b (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \\ & \leq C \left(\max \left(\mathbb{E}(\|\hat{g} - g\|_2^2), \mathbb{E}(\|\hat{f} - f\|_2^2), \frac{1}{n^\theta} \right) \right). \end{aligned}$$

In the next section, we aim to exhibit fast rates of convergence for \hat{h} . To reach this goal, we construct wavelet adaptive estimators for \hat{f} and \hat{g} . Further details and discussions on wavelets in nonparametric statistics can be found in, e.g., Antoniadis (1997), Vidakovic (1999) and Härdle *et al.* (1998).

3 Adaptive wavelet estimation

Before introducing our wavelet estimators, let us present some basics on wavelets. For the sake of simplicity in the notations, we work on the interval $[0, 1]$ (so $a = 0$ and $b = 1$).

3.1 Wavelet basis on $[0, 1]$

Let us briefly recall the construction of wavelet basis on the interval $[0, 1]$ introduced by Cohen *et al.* (1993). Let N be a positive integer, ϕ and ψ be the initial wavelets of the Daubechies orthogonal wavelets $db2N$. We set

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k).$$

With appropriate treatments at the boundaries, there exists an integer τ satisfying $2^\tau \geq 2N$ such that the collection $\mathcal{S} = \{\phi_{\tau,k}(\cdot), k \in \{0, \dots, 2^\tau - 1\}; \psi_{j,k}(\cdot); j \in \mathbb{N} - \{0, \dots, \tau - 1\}, k \in \{0, \dots, 2^j - 1\}\}$, is an orthonormal basis of $\mathbb{L}^2([0, 1])$.

Any $v \in \mathbb{L}^2([0, 1])$ can be expanded on \mathcal{S} as

$$v(x) = \sum_{k=0}^{2^\tau-1} \alpha_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(x), \quad x \in [0, 1],$$

where $\alpha_{j,k}$ and $\beta_{j,k}$ are the wavelet coefficients of v defined by

$$\alpha_{j,k} = \int_0^1 v(x) \phi_{j,k}(x) dx, \quad \beta_{j,k} = \int_0^1 v(x) \psi_{j,k}(x) dx. \quad (5)$$

3.2 Besov balls

In our statistical study, we will suppose that the unknown functions f and g belong to a wide class of functions: the Besov balls. Let us now briefly present their sequential versions.

Let $M > 0$, $s > 0$, $p \geq 1$ and $r \geq 1$. A function v belongs to $B_{p,r}^s(M)$ if and only if there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (5) satisfy

$$2^{\tau(1/2-1/p)} \left(\sum_{k=0}^{2^\tau-1} |\alpha_{\tau,k}|^p \right)^{1/p} + \left(\sum_{j=\tau}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq M^*.$$

In this expression, s is a smoothness parameter and p and r are norm parameters. For a particular choice of s , p and r , $B_{p,r}^s(M)$ contains standard balls of function spaces, as the Hölder and Sobolev balls. See, e.g., Devore and Popov (1988), Meyer (1992) and Härdle *et al.* (1998).

3.3 Block thresholding estimator

Let us now present the considered family of wavelet estimators based on block thresholding technique.

Let n be a positive integer, (a_n) be a sequence of positive integers such that $\lim_{n \rightarrow \infty} a_n = \infty$ and $W = (W_1, \dots, W_{a_n})$ be a random vector of a_n i.i.d. random variables with common density v . We define the \hat{v} by

$$\hat{v}(x, W) = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k} \phi_{\tau,k}(x) + \sum_{j=j_0}^{j_1} \sum_{K \in A_j} \sum_{k \in U_{j,K}} \hat{\beta}_{j,k} \mathbf{1}_{\{|\hat{b}_{j,K}| \geq \kappa \lambda_j\}} \psi_{j,k}(x), \quad (6)$$

where

$$\hat{\alpha}_{j_0,k} = \frac{1}{a_n} \sum_{i=1}^{a_n} \phi_{j_0,k}(W_i), \quad \hat{\beta}_{j,k} = \frac{1}{a_n} \sum_{i=1}^{a_n} \psi_{j,k}(W_i),$$

$$\hat{b}_{j,K} = \left(\sum_{k \in U_{j,K}} \hat{\beta}_{j,k}^2 \right)^{1/2},$$

$2^{j_0} = L$, $2^{j_1} = \lceil n/L \rceil$, $A_j = \{1, \dots, 2^j/L\}$ et $U_{j,K} = \{k \in \{0, \dots, 2^j - 1\}; (K-1)L \leq k \leq KL-1\}$, κ is a positive constant, $\lambda_n = \sqrt{L/a_n}$ and $L = \lceil \ln a_n \rceil$.

The idea of \hat{v} is to select $(\hat{\beta}_{j,k})_{j,k}$ by overlapping blocks of common length L according to the magnitude of the ℓ_2 norm of $(\hat{\beta}_{j,k})_{j,k}$ in each block. It was introduced by Hall *et al.* (1998, 1999) and improved by Chicken and Cai (2005). It is of interest thanks to his adaptivity and its good asymptotic performances in terms of MISE. In particular, the following result determines the rate of convergence attained by this estimator under the MISE over Besov balls.

Theorem 2 *Let \hat{v} be (6). Suppose that v is bounded from above and $v \in B_{p,r}^s(M)$ with $M > 0$, $s \in (0, N)$, $p \geq 1$ and $r \geq 1$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}(\|\hat{v}(\cdot, W) - v\|_2^2) \leq C\varphi(a_n, s, p),$$

where

$$\varphi(a_n, s, p) = \begin{cases} a_n^{-2s/(2s+1)} & \text{if } p \geq 2 \text{ and } s \in (0, N), \\ \left(\frac{\ln a_n}{a_n} \right)^{2s/(2s+1)} & \text{if } p \in [1, 2) \text{ and } s \in (1/p, N). \end{cases}$$

The details of the proof can be found in Chesneau (2010) taking $w(x) = \mu = 1$.

Note that the rate of convergence φ_n is the near optimal one in the minimax sense. The only difference is a logarithmic term for the case $p \in [1, 2)$ (see, e.g., Härdle *et al.* (1998) and Tsybakov (2004)).

3.4 Main estimator

Let us now consider the bi-dimensional density estimation problem described in Section 1.

– Let $W_o = (X_1, \dots, X_{\lfloor n/2 \rfloor})$ and \hat{v} be (6). We define \hat{f} as

$$\hat{f}(x) = \hat{v}(x, W_o),$$

with $a_n = \lfloor n/2 \rfloor$.

– Let $W_\bullet = (Y_{[n/2]+1}, \dots, Y_n)$ and \hat{v} be (6). We define \hat{g} as

$$\hat{g}(y) = \hat{v}(y, W_\bullet),$$

with $a_n = b_n = n - [n/2]$.

Finally we consider the following estimator for h (1):

$$\hat{h}(x, y) = \hat{f}(x)\hat{g}(y)c(\hat{F}(x), \hat{G}(y)), \quad (7)$$

where

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}, \quad \hat{G}(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i \leq y\}}.$$

Remark 1 Since \hat{f} is defined with W_o , \hat{g} is defined with W_\bullet and $(X_1, Y_1), \dots, (X_n, Y_n)$ are independent, \hat{f} and \hat{g} are independent.

Remark 2 If X_i and Y_i are independent, i.e. $c(x, y) = 1$, we can consider \hat{h} with \hat{f} defined with $W_o = (X_1, \dots, X_n)$ and \hat{g} defined with $W_\bullet = (Y_1, \dots, Y_n)$. In this case, \hat{f} and \hat{g} are always independent.

3.5 Rate of convergence

Theorem 3 investigates the rate of convergence attained by \hat{h} (7) under the MISE over Besov balls.

Theorem 3 *We consider the bi-dimensional density estimation problem described in Section 1 under (H1), (H2) and (H3). Let \hat{h} be (7). Suppose that*

- $f \in B_{p_1, r_1}^{s_1}(M_1)$ with $M_1 > 0$, $r_1 \geq 1$, either $\{p_1 \geq 2 \text{ and } s_1 \in (0, N)\}$ or $\{p_1 \in [1, 2) \text{ and } s_1 \in (1/p_1, N)\}$,
- $g \in B_{p_2, r_2}^{s_2}(M_2)$ with $M_2 > 0$, $r_2 \geq 1$, either $\{p_2 \geq 2 \text{ and } s_2 \in (0, N)\}$ or $\{p_2 \in [1, 2) \text{ and } s_2 \in (1/p_2, N)\}$.

Then there exists a constant $C > 0$ such that

$$\mathbb{E} \left(\int_0^1 \int_0^1 (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \leq C \max(\vartheta_n, C_3^2 n^{-\theta}),$$

where

$$\vartheta_n = \begin{cases} n^{-2s_*/(2s_*+1)}, & \text{if } p \geq 2 \text{ and } s \in (0, N), \\ \left(\frac{\ln n}{n}\right)^{2s_*/(2s_*+1)}, & \text{if } p \in [1, 2) \text{ and } s \in (1/p, N) \end{cases}$$

and $s_* = \min(s_1, s_2)$.

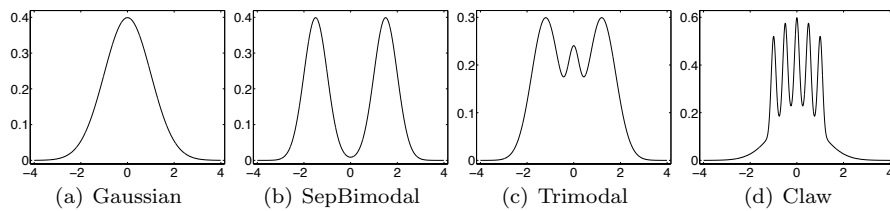


Fig. 1 Test densities.

The proof of Theorem 3 combines several elements of Theorems 1 and 2.

Let us remark that, when $s_1 = s_2$ and $\theta \in [2s_1/(2s_1 + 1), 1]$, the rate of convergence ϑ_n in Theorem 3 is the near optimal one in the minimax sense for the unidimensional density estimation problem (see, e.g., Tsybakov (2004) and Härdle *et al.* (1998)). In this case, our estimator escapes to the so-called “curse of dimensionality”; the bi-dimensionality of the considered estimation problem has no influence on the performance of our estimator. This result can be viewed as a generalization of the phenomenon exhibited by Liebscher (2005) for $\theta = 1$ with kernel methods. When $\theta \in (0, 2s_1/(2s_1 + 1))$, ϑ_n becomes $n^{-\theta}$, which is slower than the previous one. Note that, when $\theta \in [0, s_1/(s_1 + 1)]$, the optimal rate of convergence for the bi-dimensional density estimation problem is conserved; the curse of dimensionality still occurs.

4 A short simulation study

In this section, since it is new in the literature, we aim to illustrate the numerical performances of the adaptive wavelet estimator \hat{h} (7). We have compared its performance to the traditional kernel estimator based on the same plug-in approach on several densities. All experiments were conducted using a Gaussian kernel and we have been focused on a global bandwidth selector: the *rule of thumb* (ROT) bandwidth selector (see, e.g., Silverman (1986)). Thus, the optimal bandwidth is given by $b_{\text{ROT}} = 1.06 \min(\hat{\sigma}, Q/1.34)n^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation and Q is the interquartile range. For both estimation methods we used the first half and the last half of the data to estimate f and g respectively. We have considered the copula densities (3) with associated dependence parameter $v = 1$ and $a = 1/2$ and the *Gaussian copula* with correlation coefficient $\rho = 0.9$. Four true distributions are considered, three for f : “SeparatedBimodal”, “Trimodal” and “Claw”, initially introduced in Marron and Wand (1992) and the Gaussian distribution for g (see Figure.1).

Since our estimation method is adaptive, we have chosen a predetermined threshold κ (universal threshold, see, e.g., Donoho *et al.* (1996)) for all the tests and the Symmlet wavelet with 6 vanishing moments was used throughout all experiments. For numerical implementation, we consider an interval $[a_1, a_2]$ that cover the range of the data and the marginal density estimates were evaluated at $M = 2^r$ equally spaced points $t_i = a_1 + (a_2 - a_1)/M$,

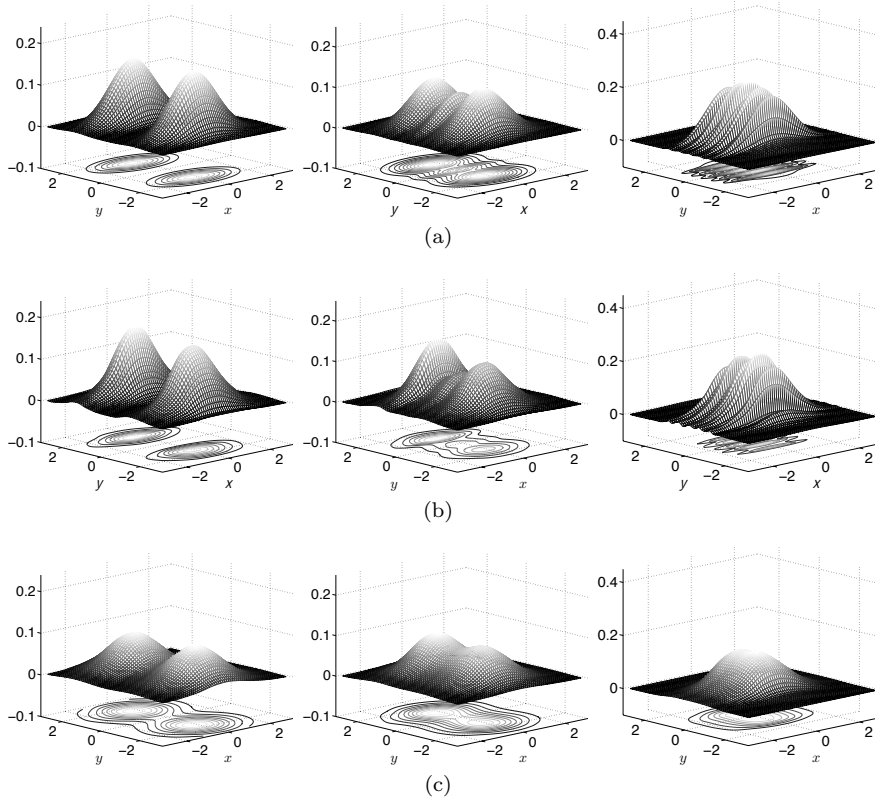


Fig. 2 Huang and Kotz copula density. (a) Theoretical densities h . (b) Block thresholding estimators. (c) Kernel estimators. From left to right SeparatedBimodal, Trimodal and Claw.

$i = 0, 1, \dots, M-1$, between a_1 and a_2 , with $r = 6$, $a_2 = -a_1 = 4$ and $M = 64$ is the number of discretization points. The primary level $j_0 = 3$ and the finest resolution level j_1 used in all our simulations was chosen to be the maximum resolution level allowed by the discretization.

Note that \hat{h} may sometimes be negative (especially on the tails and on the regions where the sample is sparse) and fail to integrate to 1. To solve this problem, we have truncated the estimate to its positive part and then renormalized the truncation.

Typical reconstructions from a single simulation with $n = 1000$ are depicted in Fig. 2 and Fig. 3 for Huang and Kotz (1999) and Gaussian copula respectively. One can see that our adaptive block thresholding estimator \hat{h} is effective to estimate the unknown density h and provides very competitive results in comparison to the kernel ones.

For each distribution, 100 data sets of sample size $n = 500, 1000$ and 2000 were simulated. The root mean squared error (RMSE) from 100 repetitions are tabulated in Table 1. It shows that our wavelet-based estimator clearly

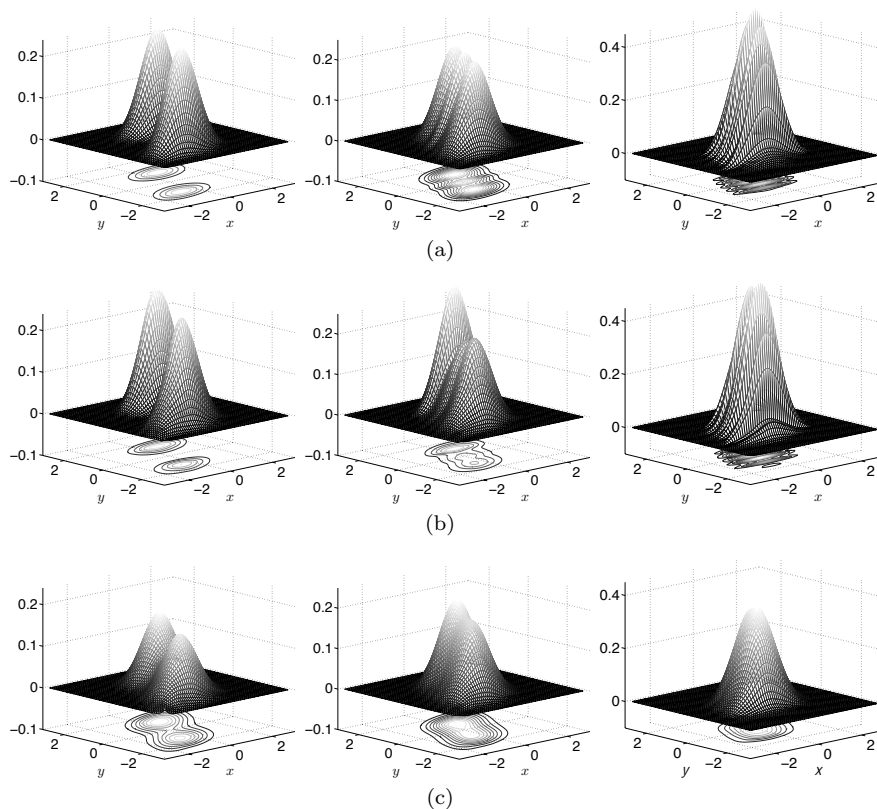


Fig. 3 Gaussian copula density. (a) Theoretical densities h . (b) Block thresholding estimators. (c) Kernel estimators. From left to right SeparatedBimodal, Trimodal and Claw.

Table 1 $1000 \times \text{RMSE}$ values from 100 replications.

n	<i>SeparatedBimodal</i>			<i>Trimodal</i>			<i>Claw</i>		
	500	1000	2000	500	1000	2000	500	1000	2000
<i>Huang and Kotz Copula, $v = 1, a = 1/2$</i>									
Wavelet	7.57	6.15	5.66	6.84	5.71	4.89	17.91	11.09	7.79
Kernel	20.62	17.78	14.72	8.77	7.68	7.05	18.93	18.44	18.12
<i>Gaussian Copula, $\rho = 0.9$</i>									
Wavelet	10.48	8.69	7.55	10.82	9.33	7.79	32.88	19.18	15.16
Kernel	29.95	25.41	19.99	13.11	11.20	10.42	35.96	34.74	34.70

outperforms the kernel estimator for all considered densities and all sample sizes of the simulation study. Furthermore, as expected, for both methods, and in all cases, the RMSE is decreasing as the sample size increases.

5 Proofs

In this section, for the sake of simplicity, C denotes a generic constant; its value may change from one term to another.

Proof of Theorem 1. For the sake of simplicity, we set

$$c_{x,y} = c(F(x), G(y)), \quad \hat{c}_{x,y} = c(\hat{F}(x), \hat{G}(y)).$$

We have

$$\begin{aligned} \hat{h}(x,y) - h(x,y) &= \hat{f}(x)\hat{g}(y)\hat{c}_{x,y} - f(x)g(y)c_{x,y} \\ &= f(x)(\hat{g}(y) - g(y))(\hat{c}_{x,y} - c_{x,y}) + g(y)(\hat{f}(x) - f(x))(\hat{c}_{x,y} - c_{x,y}) \\ &\quad + c_{x,y}(\hat{f}(x) - f(x))(\hat{g}(y) - g(y)) + f(x)g(y)(\hat{c}_{x,y} - c_{x,y}) + f(x)c_{x,y}(\hat{g}(y) - g(y)) \\ &\quad + g(y)c_{x,y}(\hat{f}(x) - f(x)) + (\hat{f}(x) - f(x))(\hat{g}(y) - g(y))(\hat{c}_{x,y} - c_{x,y}). \end{aligned}$$

It follows from **(H3)** that

$$|\hat{c}_{x,y} - c_{x,y}| \leq C_3(|\hat{F}(x) - F(x)|^\theta + |\hat{G}(y) - G(y)|^\theta).$$

Using **(H1)**, **(H2)**, $c_{x,y} \leq C$ due to **(H3)**, and the previous inequalities, we obtain

$$\begin{aligned} &|\hat{h}(x,y) - h(x,y)| \\ &\leq C(|\hat{g}(y) - g(y)| + |\hat{f}(x) - f(x)| + |\hat{f}(x) - f(x)||\hat{g}(y) - g(y)| \\ &\quad + C_3|\hat{F}(x) - F(x)|^\theta + C_3|\hat{G}(y) - G(y)|^\theta). \end{aligned}$$

The elementary inequality $(\sum_{i=1}^5 a_i)^2 \leq 5 \sum_{i=1}^5 a_i^2$ and an integration of $[0, 1]^2$ yield

$$\begin{aligned} &\mathbb{E} \left(\int_a^b \int_a^b (\hat{h}(x,y) - h(x,y))^2 dx dy \right) \\ &\leq C(\mathbb{E}(\|\hat{g} - g\|_2^2) + \mathbb{E}(\|\hat{f} - f\|_2^2) + \mathbb{E}(\|\hat{f} - f\|_2^2 \|\hat{g} - g\|_2^2)) \\ &\quad + C_3^2 \sup_{x \in [a,b]} \mathbb{E}((\hat{F}(x) - F(x))^{2\theta}) + C_3^2 \sup_{y \in [a,b]} \mathbb{E}((\hat{G}(y) - G(y))^{2\theta}). \end{aligned}$$

Since $\theta \in (0, 1]$, $\mathbb{E}(\hat{F}(x)) = F(x)$ and $\mathbb{V}(\hat{F}(x)) = (1/n)F(x)(1 - F(x)) \leq 1/(4n)$, the Hölder inequality yields

$$\sup_{x \in [a,b]} \mathbb{E}((\hat{F}(x) - F(x))^{2\theta}) \leq \sup_{x \in [a,b]} \left(\mathbb{V}(\hat{F}(x)) \right)^\theta \leq \frac{1}{(4n)^\theta}.$$

Similarly, we have

$$\sup_{y \in [a,b]} \mathbb{E}((\hat{G}(y) - G(y))^{2\theta}) \leq \frac{1}{(4n)^\theta}.$$

Therefore the MISE of \hat{h} can be bounded as

$$\begin{aligned} & \mathbb{E} \left(\int_a^b \int_a^b (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \\ & \leq C \left(\mathbb{E}(\|\hat{g} - g\|_2^2) + \mathbb{E}(\|\hat{f} - f\|_2^2) + \mathbb{E}(\|\hat{f} - f\|_2^2 \|\hat{g} - g\|_2^2) + \frac{C_3^2}{n^\theta} \right). \end{aligned}$$

Theorem 1 is proved. □

Proof of Theorem 3. It follows from Theorem 2 that

- If $f \in B_{p_1, r_1}^{s_1}(M_1)$ with $M_1 > 0$, $r_1 \geq 1$, either $\{p_1 \geq 2 \text{ and } s_1 \in (0, N)\}$ or $\{p_1 \in [1, 2) \text{ and } s_1 \in (1/p_1, N)\}$, there exist two constants $C > 0$ and $C_* > 0$ such that

$$\mathbb{E}(\|\hat{f} - f\|_2^2) \leq C\varphi(a_n, s_1, p_1) \leq C_*\vartheta_n,$$

with $a_n = \lfloor n/2 \rfloor$.

- If $g \in B_{p_2, r_2}^{s_2}(M_2)$ with $M_2 > 0$, $r_2 \geq 1$, either $\{p_2 \geq 2 \text{ and } s_2 \in (0, N)\}$ or $\{p_2 \in [1, 2) \text{ and } s_2 \in (1/p_2, N)\}$, there exist two constants $C > 0$ and $C_* > 0$ such that

$$\mathbb{E}(\|\hat{g} - g\|_2^2) \leq C\varphi(b_n, s_2, p_2) \leq C_*\vartheta_n,$$

with $b_n = n - \lfloor n/2 \rfloor$.

Moreover, using the independence between \hat{f} and \hat{g} , we get

$$\mathbb{E}(\|\hat{f} - f\|_2^2 \|\hat{g} - g\|_2^2) = \mathbb{E}(\|\hat{f} - f\|_2^2) \mathbb{E}(\|\hat{g} - g\|_2^2) \leq C\vartheta_n.$$

Theorems 1 yields

$$\begin{aligned} & \mathbb{E} \left(\int_0^1 \int_0^1 (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \\ & \leq C \left(\mathbb{E}(\|\hat{g} - g\|_2^2) + \mathbb{E}(\|\hat{f} - f\|_2^2) + \mathbb{E}(\|\hat{f} - f\|_2^2 \|\hat{g} - g\|_2^2) + \frac{C_3^2}{n^\theta} \right) \\ & \leq C \left(\vartheta_n + \frac{C_3^2}{n^\theta} \right) \leq C \max(\vartheta_n, C_3^2 n^{-\theta}). \end{aligned}$$

This ends the proof of Theorem 3. □

Acknowledgements

The authors are very grateful to the Associate Editor and the anonymous referees for their careful reading and valuable comments that have led to an improvement of the presentation.

References

- Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion), *Journal of the Italian Statistical Society Series B*, 6, 97-144.
- Balakrishnan, N. and Lai, C.D., (2009). *Continuous Bivariate Distributions*, Springer, New York, NY, USA, 2nd edition.
- Bouezmarni, T. and Rombouts, J.V.K., (2008). Semiparametric Multivariate Density Estimation for Positive Data Using Copulas, *Computational Statistics and Data Analysis*, 53, 2040-2054.
- Chesneau, C. (2010). Wavelet block thresholding for density estimation in the presence of bias, *Journal of the Korean Statistical Society*, 39(1), 43-53.
- Chicken, E. and Cai, T. (2005). Block thresholding for density estimation: Local and global adaptivity. *J. Multivariate Analysis*, 95, 76-106.
- Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24, 1, 54-81.
- DeVore, R. and Popov, V. (1988). Interpolation of Besov spaces, *Trans. Amer. Math. Soc.*, 305, 397-414.
- Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *Annals of Statistics*, 24, 508-539.
- Faugeras, O.P. (2009). A quantile-copula approach to conditional density estimation, *J. Multivariate Anal.*, 100, 9, 2083-2099.
- Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods, *Ann. Statist.*, 26,922-942.
- Hall, P., Kerkyacharian, G. and Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators, *Statist. Sinica*, 9, 33-50.
- Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics New York 129, Springer Verlag.
- Huang, J.S. and Kotz, S. (1999). Modifications of the Farlie Gumbel Morgenstern distributions. A tough hill to climb, *Metrika*, 49, 135 -145.
- Liebscher, E. (2005). Semiparametric density estimators using copulas, *Comm. Statist. - Theory Methods*, 34, 59-71.
- Marron, J.S. and Wand, M.P. (1992). Exact Mean Integrated Squared Error, *Ann. Statist.*, 20, 712-736.
- Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press, Cambridge.
- Nelsen, R. (1999). *An introduction to copulas (1st ed.)*. New York: Springer- Verlag.
- Silverman, B. W. (1986). *Density estimation: for statistics and data analysis*, Chapman and Hall, London.
- Simonoff, J. S. (1996). *Smoothing Methods in Statistics*, Springer-Verlag, New York.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* 8, 229-231.
- Tsybakov, A.B. (2004). *Introduction à l'estimation non-paramétrique*, Springer.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.