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

Noname manuscript No.
(will be inserted by the editor)

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 The estimation of a bi-dimensional density in the case of knowledge of the copula density is investigated. We propose a general estimator and study its performance under the mean integrated squared error (MISE). Then we derive an adaptive estimator based on wavelets. Under some smoothness assumptions, we show that it attains the standard unidimensional rate of convergence for a large class of unknown functions. We also report a short simulation study to support our theoretical findings.

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

2000 Mathematics Subject Classification 62G08, 62G20.

1 Motivations

We consider the bi-dimensional density estimation problem described as follows. Let $(X_i, Y_i)_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with continuous density $h : [a, b]^2 \to [0, \infty)$ with $a < b$. Following the Sklar Theorem (see Sklar (1959)), there exists an unique function $c : [0, 1]^2 \to [0, \infty)$, called copula density, such that

$$h(x, y) = f(x)g(y)c(F(x), G(y)), \tag{1}$$

[(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)] Laboratoire de Mathématiques Jean Leray, 2, Rue de la Houssinire, 44322 Nantes, France E-mail: Fabien.Navarro@univ-nantes.fr
[(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). We aim to estimate h from the n random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ under the assumption that c is known. As mentioned in Segers *et al.* (2008) and Einmahl and Van den Akker (2011), the knowledge of c is artificial but provides an important step toward the solution of more realistic problems, as the estimation of h when c belongs to a parametric family.

In the first part of this note, following the idea of Faugeras (2009), we develop a general "plug-in estimator" for the estimation of h. We evaluate its performance by considering the mean integrated squared error (MISE). A result providing sharp upper bound for the MISE is proved. A second part is devoted to an adaptive wavelet version of our estimator. It is based on the wavelet block thresholding elaborated by Hall *et al.* (1998, 1999). We prove that it attains a fast rate of convergence under the MISE over a wide class of unknown functions: the Besov balls. In particular, we show that, under some smoothness assumptions on f, g and c, our estimator attains the standard unidimensional rate of convergence. This result provides a contribution to the so called "curse of dimensionality", also studied in other estimation copula contexts by Segers *et al.* (2008) and Einmahl and Van den Akker (2011).

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 over Besov balls are presented in Section 3. Simulation results are presented 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

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

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 $c : [0,1]^2 \to [0,\infty)$ 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 assumptions **(H3)** characterizes the smoothness of c. It is satisfied for numerous copula densities. For instance

- If $c(x,y) = 1$ (i.e., the independent case), then **(H3)** is satisfied with $C_3 = 0$ (and any $\theta \in (0,1]$),
- If c is the Farlie-Gumbel-Morgenstern copula density, i.e., $c(x,y) = 1 + \upsilon(2x-1)(2y-1)$ with $\upsilon \in [-1,1]$ (see Nelsen (1999)), then **(H3)** is satisfied with $C_3 = 8\upsilon$ and $\theta = 1$.

2.2 MISE result

Theorem 1 presents an estimator for h and shows an upper bound for its MISE.

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

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

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

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

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

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

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

(where θ refers to **H3***).*

The general estimator (2) is derived from the estimator introduced by Faugeras (2009) for another statistical problem. The proof of Theorem 1 is based on a suitable decomposition of $\hat{h} - h$ and moment properties of $\hat{F}$ and $\hat{G}$.

Remark 1 If the estimators $\hat{f}$ and $\hat{g}$ are bounded from above, we can bound the MISE of $\hat{h}$ by the MISE's of $\hat{f}$ and $\hat{g}$; we have

$$\mathbb{E}\left(\int_a^b\int_a^b(\hat{h}(x,y)-h(x,y))^2dxdy\right)$$
$$\leq C\left(\max\left(\mathbb{E}(||\hat{g}-g||_2^2),\mathbb{E}(||\hat{f}-f||_2^2),\frac{1}{n^\theta}\right)\right),$$

In the next section, we will introduce basics on wavelets and the considered wavelet estimators. 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

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

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

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

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

$$\alpha_{j,k}=\int_0^1 v(x)\phi_{j,k}(x)dx,\qquad \beta_{j,k}=\int_0^1 v(x)\psi_{j,k}(x)dx. \tag{3}$$

3.2 Besov balls

We work with the sequential version of Besov balls defined as follows. Let $M > 0$, $s > 0$, $p \geq 1$ and $r \geq 1$. A function v belongs to $B^s_{p,r}(M)$ if and only if there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (3) 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, $B^s_{p,r}(M)$ 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 estimators

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

$$\hat{v}(x, W) = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k} \phi_{\tau,k}(x) + \sum_{j=j_0}^{j_1} \sum_{K \in A_j} \sum_{k \in U_{j,K}} \hat{\beta}_{j,k} \mathbf{1}_{\{|\hat{b}_{j,K}| \geq \kappa \lambda_j\}} \psi_{j,k}(x), \quad (4)$$

where

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

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

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

This estimator 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 (4). Suppose that v is bounded from above and $v \in B^s_{p,r}(M)$ with $M > 0$, $s \in (0, N)$, $p \geq 1$ and $r \geq 1$. Then there exists a constant $C > 0$ such that*

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

where

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

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)).

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

- Let $W_o = (X_1, \ldots, X_{[n/2]})$ and $\hat{v}$ be (4). We define $\hat{f}$ as

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

  with $a_n = [n/2]$.
- Let $W_\bullet = (Y_{[n/2]+1}, \ldots, Y_n)$ and $\hat{v}$ be (4). 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):

$$\hat{h}(x, y) = \hat{f}(x)\hat{g}(y)c(\hat{F}(x), \hat{G}(y)), \tag{5}$$

where

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

Remark 2 Since $\hat{f}$ is defined with W_o, $\hat{g}$ is defined with $W_\bullet$ and $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent, $\hat{f}$ and $\hat{g}$ are independent.

Remark 3 If X_i and Y_i are independent, i.e. $c(x, y) = 1$, we can consider $\hat{h}$ with $\hat{f}$ defined with $W_o = (X_1, \ldots, X_n)$ and $\hat{g}$ defined with $W_\bullet = (Y_1, \ldots, Y_n)$. In this case, $\hat{f}$ and $\hat{g}$ are always independent.

3.4 Rate of convergence

Theorem 3 investigates the rate of convergence attained by $\hat{h}$ 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* (5). *Suppose that*

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

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

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

where

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

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

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 is the near optimal one in the minimax sense for the unidimensional density estimation problem under the MISE over Besov balls $B^{s_*}_{p,r}(M)$ (see, e.g., Tsybakov (2004) and Härdle *et al.* (1998)). Theorem 3 proves, with some informations on c, our estimator escapes to the so-called "curse of dimension"; the bi-dimensionality of the considered estimation problem has no influence on the performance of our estimator. This completes the studies of Segers *et al.* (2008) and Einmahl and Van den Akker (2011) for other estimation problems using copula.

4 Simulation

We now illustrate these theoretical results with a short simulation study. 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. The densities were evaluated at $T = 2^J$ equispaced points $t_i = 2ib/T$, $i = -T/2, \ldots, T/2 - 1$ between $-b$ and b, where J is the index of the highest resolution level and T is the number of discretization points. The primary level $j_0 = 3$, $T = 256$ and the finest resolution level j_1 used in all our simulations was chosen to be the maximum resolution level allowed by the discretization. All simulations have been implemented using Matlab.

We first consider the case in which X_i and Y_i are independent, i.e. $c(x, y) = 1$. In order to satisfy the assumptions (**H1**) and (**H2**), we have choosen compactly supported densities. More precisely, we consider the two-sided truncated normal distribution on $[a, b]$ denoted by $\mathcal{N}(\mu, \sigma^2, a, b)$, with density

$$f(x; \mu, \sigma, a, b) = \begin{cases} \frac{\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise,} \end{cases}$$

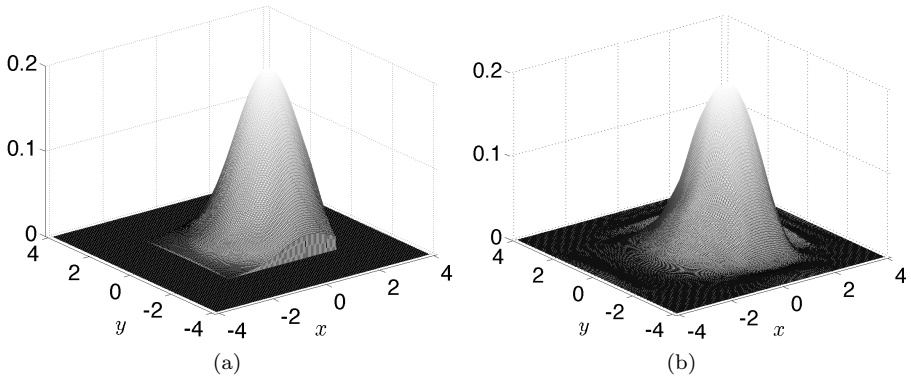

Fig. 1 (a) Theoretical density h. (b) Block thresholding estimator $\hat{h}$ from a single simulation.

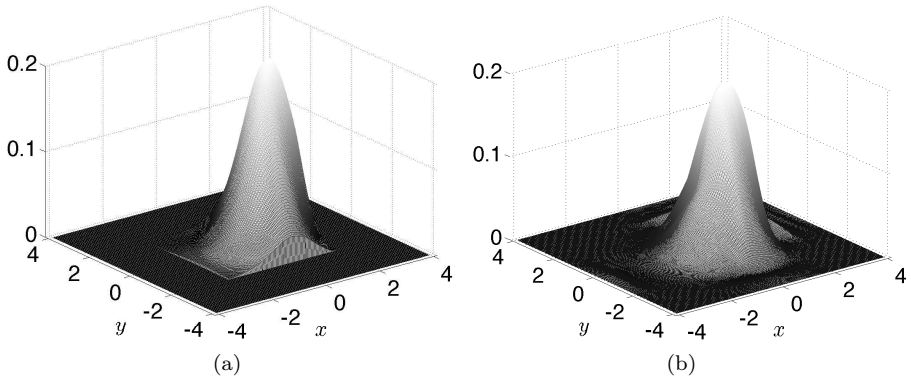

Fig. 2 (a) Theoretical density h. (c) Block thresholding estimator $\hat{h}$ from a single simulation.

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$ is the probability density function of the standard normal distribution, Φ is its cumulative distribution function and the parameters μ and σ are respectively the mean and the standard deviation of the distribution.

Let us first consider an example where $X_1 \sim \mathcal{N}(0, 1, -2, 2)$ and $Y_1 \sim \mathcal{N}(1/5, 1, -2, 2)$ with $n = 1000$. A typical example of estimation is given in Fig 1.

In our second example, we propose to illustrate our proposed estimator with the FGM copula density with associated dependence parameter $\upsilon = 1$ (i.e., $c(x, y) = 1 + \upsilon(2x - 1)(2y - 1)$). Typical reconstructions from a single simulation are depicted in Fig 2. In both examples, one can see that our adaptive block thresholding estimator $\hat{h}$ is effective to estimate the unknown density h.

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

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

We have

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

It follows from **(H3)** that

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

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

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

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

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

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

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

Similarly, we have

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

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

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

Theorem 1 is proved.

□

Proof of Theorem 3. It follows from Theorem 2 that

- If $f\in B^{s_1}_{p_1,r_1}(M_1)$ with $M_1>0$, $r_1\geq 1$, either $\{p_1\geq 2$ and $s_1\in(0,N)\}$ or $\{p_1\in[1,2)$ 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 $a_n=[n/2]$.
- If $g\in B^{s_2}_{p_2,r_2}(M_2)$ with $M_2>0$, $r_2\geq 1$, either $\{p_2\geq 2$ and $s_2\in(0,N)\}$ or $\{p_2\in[1,2)$ 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 $b_n=n-[n/2]$.

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

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

This ends the proof of Theorem 3.

□

References

Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion), *Journal of the Italian Statistical Society Series B*, 6, 97-144.

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, *A*nnals of Statistics, 24, 508-539.

Einmahl, J. and Van den Akker, R. (2011). Superefficient estimation of the marginals by exploiting knowledge on the copula. *Journal of Multivariate Analysis*, 102 (9), 1315-1319.

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*.

Meyer, Y. (1992). Wavelets and Operators. *Cambridge University Press*, Cambridge.

Nelsen, R. (1999). *An introduction to copulas (1st ed.)*. New York: Springer- Verlag.

Segers, Van den Akker, R. and Werker, B. (2008). Improving upon the Marginal Empirical Distribution Functions When the Copula is Known, *Discussion paper, Tilburg University, Center for Economic Research*.

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.