Int. J. Nonlinear Anal. Appl. In Press, 1–7
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2025.34631.5178

The role of the copula density function in the estimation of the conditional density function

Esmaeil Shirazi

Faculty of Science, Gonbad Kavous University, Gonbad Kavous, Iran

(Communicated by Zakieh Avazzadeh)

Abstract

In this article, a new wavelet-based method for estimating the conditional density function using the wavelet method is investigated. Based on this method, we will explain how to obtain an estimate with the optimal convergence rate for the conditional density function based on the information in the quantiles and using the copula density function. We also discuss the convergence rate of the new estimator.

Keywords: BlockShrink estimator, Copula function, Conditional density, Quantiles, Wavelets
2020 MSC: Primary 62G07; Secondary 62G08, 62G20

1 Introduction

One of the important statistical subjects with wide application in many practical problems, especially problems related to forecasting, is the estimation of the conditional probability function. For example, this function plays a central role in financial econometrics (see [2] and [9]). Now, suppose that $(X_i, Y_i)_{i\geq 1}$ be mutually independent random vectors of a pair (X, Y). We can show the structure of function $\upsilon(y \mid x)$ as below

$$\upsilon(y \mid x) = \frac{f(x, y)}{\upsilon(x)} \tag{1.1}$$

where $f(x, y)$ and $\upsilon(x)$ denote the joint density of (X, Y) and X, respectively. The kernel estimators of these functions are given by

$$\hat{f}_n(x, y) = \frac{1}{n}\sum_{i=1}^{n} W'_{h'}(X_i - x)W_h(Y_i - y)$$

$$\hat{\upsilon}_n(x) = \frac{1}{n}\sum_{i=1}^{n} W'_{h'}(X_i - x)$$

respectively. Here, $W_h(.) = 1/hW(./h)$ and $W'_{h'}(.) = 1/h'W(./h')$ are (rescaled) kernels whose associated sequences $h = h_n$ and $h' = h'_n$ of bandwidth vanish as $n \to \infty$. Accordingly, an estimator of $\upsilon(y \mid x)$ is given by

$$\hat{\upsilon}(y \mid x) = \frac{\hat{f}_n(x, y)}{\hat{\upsilon}_n(x)}$$

Email address: esmaeil.shirazi@gonbad.ac.ir (Esmaeil Shirazi)

Received: July 2024 *Accepted:* January 2025

Kernel estimators have been studied in various fields of application due to their ease of working with them. But these estimators create limitations, especially in the estimation of bounds in discontinuous functions. For this reason, it is more preferred in this situation to suggest alternative estimation methods such as wavelets, which have higher efficiency. The complete background on wavelets can be found in [7].

In this study, following the idea of [3], we propose two new wavelet estimators for $\upsilon(y \mid x)$ by using the observation detail function. The first estimator is linear and non-adaptive because it depends on the smoothing parameter s. But the second estimator, which is based on the block threshold idea, is non-linear and entirely adaptive. Also, we evaluate their performance by applying the Mean Integrated Error (MISE) criterion on a wide class of Besov space functions. In the end, we show that the introduced estimators obtain the optimal convergence rate $(n/\log n)^{-\frac{2s}{1+2s}}$, where, in section 2, we will discuss how to choose the smoothing parameter s.

The structure of this article is compiled as follows. In section 2, we have an overview of wavelets and Besov space, and the general structure of the desired threshold estimator is introduced. In section 3, the convergence behavior of the estimator is investigated, and in the last section, the proof of the main results of the article is given.

2 Preliminaries

In this section, some basic mathematical concepts such as multi-resolution analysis and functional spaces as well as the considered estimator are introduced.

2.1 Wavelets

First, we define an orthogonal bases of the wavelet used in this article are based on the "father" compact supported wavelet $\phi(.)$ and the "mother" compact supported wavelet $\psi(.)$. For this purpose, we consider the periodic wavelet bases on the unit interval. For any $x \in [0,1]$, any integer i and any $j \in \{0, ..., 2^i - 1\}$, let

$$\phi_{i_0 j}(x) = 2^{i_0/2}\phi(2^{i_0}x - j),\ \psi_{ij}(x) = 2^{i/2}\psi(2^i x - j)$$

be the elements of the wavelet basis and

$$\phi^{per}_{i,j}(x) = \sum_{k\in Z} \phi_{i,j}(x-k),\ \psi^{per}_{i,j}(x) = \sum_{k\in Z} \psi_{i,j}(x-k)$$

there periodized version. There exists an integer τ such that the collection $\Omega =$ defined by $\Omega = \{\phi^{per}_{\tau,j},\ j = 0, ..., 2^{\tau} - 1;\ \psi^{per}_{i,j},\ i = \tau, ..., \infty,\ j = 0, ..., 2^i - 1\}$, constitutes an orthonormal basis of $L^2([0,1])$. In what follows, the superscript "per" will be suppressed from the notations for convenience. For any integer $k \geq \tau$, a function $h \in L^2([0,1])$ can be expanded into a wavelet series as

$$h(x) = \sum_{j\in\{0,...,2^{i_0}-1\}} \alpha_{i_0,j}\phi_{i_0,j}(x) + \sum_{i\geq i_0} \sum_{j\in\{0,...,2^{i}-1\}} \beta_{i,j}\psi_{i,j}(x),$$

where the wavelet coefficients $\alpha_{i_0,j}$ and $\beta_{i,j}$ are given by

$$\alpha_{i_0,j} = \int_{[0,1]} h(x)\phi_{i_0,j}(x)dx, \quad and \quad \beta_{i,j} = \int_{[0,1]} h(x)\psi_{i,j}(x)dx.$$

All the details about these wavelet bases, including the expansion into wavelet series as described above, can be found in, for example, [7] and [6].

Let $M > 0,\ s > 0,\ p \geq 1$, and $q \geq 1$. We say that a function $h(.) \in L_2(R)$ pertains to the Besov balls $B^s_{p,q}(M)$ if and only if the associated wavelet coefficients (2.1) satisfy

$$\left(\sum_{k=0}^{2^{\tau}-1} \mid \alpha_{\tau,k} \mid^p\right)^{1/p} + \left(\sum_{j=\tau}^{\infty}\left(2^{j\sigma}\left(\sum_{k=0}^{2^j-1} \mid \beta_{jk} \mid^p\right)^{1/p}\right)^q\right)^{1/q} \leq M \tag{2.1}$$

where $\sigma = s + 1/2 - 1/p$. and with the usual modifications for $p = \infty$ or $q = \infty$. These sets contain function classes of significant spatial inhomogeneity, including Sobolev balls and Holder balls. Details about Besov balls can be found in, for example [7].

2.2 A product-shaped estimator

The kernel estimator of the conditional density function introduced in the previous section has several basic drawbacks.According to a scientific explanation, these estimators may be unstable when their denominator approaches zero. The large-sample behavior of the estimators is also difficult to track down, due to the quotient form. On the other hand, it is very difficult to investigate the behavior of a large sample for them due to the quotient form. Of course, this part of the problem can be solved directly by linearizing the inverse after centering the numerator and denominator.; see, e.g., [12] or [1] for details. From a theoretical point of view, it can be argued that applying regression estimation techniques in this matter is somewhat artificial: density estimation, even conditional, should only use density estimation techniques.

Here, to solve these problems, we have proposed an estimator based on the idea of pseudo-observations, that is, a transformation of the original data. To be specific, A quantile transformation of the data is seen through a copula representation into a product estimator.

$$\hat{v}_n(y \mid x) = \hat{g}_Y(y)\hat{c}_n\{\hat{F}_n(x), \hat{G}_n(y)\} \tag{2.2}$$

where $\hat{g}_Y, \hat{c}_n, \hat{F}_n(x), \hat{G}_n(y)$ are estimators of the density g_Y of Y, the copula density c, the c.d.f. F of X and G of Y, respectively. As will be shown, the properties of $\hat{v}_n(y \mid x)$ are easily deduced from existing results in nonparametric density estimation.

2.3 Estimator presentation

We are now in a position where we can express the estimator of the conditional density function based on the information obtained from the copula density function described in the explanation of the previous section using quantile observations.

2.3.1 Quantile transformation

Data transformation strategy is one of the common methods to improve the application range and performance of classical estimation techniques, which is often used in dealing with skewed data, heavy tails, or restrictions on the support, among others; see, e.g., Chapter 14 of [5]. In order to make inference on Y from X , a quantile transformation seems appropriate, the details of which are explained in the following. Based on this, the selection of transformation $U = F(X)$ which turns a continuous random variable X with c.d.f. F into a uniform random variable U on the interval $[0, 1]$.

2.3.2 The copula representation

The copula function is introduced as an interface between the multivariate cumulative distribution function to it's one-dimensional marginal distribution functions. The word "copula" was first proposed in 1959 by [11] in the following theorems. following fundamental result:

Theorem 2.1. For any bivariate c.d.f. $F_{X,Y}$ on $\mathbb{R}^2$, with marginal c.d.f. F of X and G of Y , there exists some function $C : [0, 1]^2 \longrightarrow [0, 1]$, called the dependence or copula function, such as

$$F_{X,Y}(x, y) = C\{F(x), G(y)\}, \quad -\infty < x, y < \infty \tag{2.3}$$

If F and G are continuous, this representation is unique. The copula C is itself a c.d.f. on $[0, 1]^2$ with uniform margins.

In this theorem, the relationship between the bivariate distribution function and its marginal distribution functions is shown. In other words, the copula function shows the dependence structure in the pair (X, Y), regardless of the marginal distribution functions F and G. In fact, this way allows one to deal with the randomness of the dependence structure and the randomness of the margins separately.

Copulas are basically related to quantile transformations according to formula (2.3) as follows

$$C(u, v) = F_{X,Y}\{F^{-1}(u), G^{-1}(v)\}$$

To see more details about copulas and their properties, you can refer, for example, to Joe's book [4]. Furthermore, we assume that the copula function $C(u,v)$ has density

$$c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v).$$

Considering the Lebesgue measure on $[0,1]^2$ and accepting that F and G are strictly increasing and differentiable with density υ and g. Accordingly, $C(u,v)$ and $c(u,v)$ are c.d.f. and the density function of transformed variables $(U,V) = (F(X), G(Y))$. By deriving both sides of the (2.3) relation, we obtain the joint density function, i.e.

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = \upsilon(x) g(y) c\{F(x), G(y)\}.$$

So, the explicit product form of the conditional density function is obtained as follows:

$$\upsilon(y \mid x) = \frac{f(x,y)}{\upsilon(x)} = g(y) c\{F(x), G(y)\}. \tag{2.4}$$

2.4 Treshold Estimate

Using the product type formula (2.4) described above, a threshold method is used to construct the conditional density estimator.

1- a wavelet based nonparametric threshold estimate of the density g of Y , viz

$$\hat{g}(y) = \sum_{j \in Z} \hat{\theta}_{i_0,j} \phi_{i_0,j}(y) + \sum_{i=i_0}^{R} \sum_{j \in Z} \hat{\xi}_{ij} \psi_{i,j}(y),$$

where $\hat{\xi}_{ij} = \hat{\omega}_{i,j} I_{\left(|\hat{\omega}_{i,j}| > c\sqrt{n^{-1} \ln n}\right)}$ and $\hat{\theta}_{i_0,j},\ \hat{\omega}_{i,j}$ are defined as follow

$$\hat{\theta}_{i_0,j} = \frac{1}{n} \sum_{k=1}^{n} \phi_{i_0,j}(Y_k), \qquad \hat{\omega}_{i,j} = \frac{1}{n} \sum_{k=1}^{n} \psi_{i_0,j}(Y_k)$$

2- Given that $c(u,v)$ is the joint density of the transformed variables $(U,V) = (F(X), G(Y))$, it could be estimated in principle by the threshold estimator,

$$\hat{c}_{k'}(u,v) = \sum_{k} \hat{\alpha}_{i_0,j} \phi_{i_0 j}(u,v) + \sum_{i=i_1}^{i_2} \sum_{j \in Z} \hat{\beta}^{\epsilon}_{i,j} 1_{\{|\hat{\beta}^{\epsilon}_{i,j}| \geq d\lambda_n\}} \psi^{\epsilon}_{i,j}(u,v), \qquad u,v \in [0,1]. \tag{2.5}$$

where d is a large enough constant and $\lambda_n = \sqrt{\frac{\log(n)}{n}}$. So, the natural estimator of $\alpha_{j_0 k}$ and $\beta^{\epsilon}_{j,k}$ are given by

$$\hat{\alpha}_{j_0,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j_0,k}(F(X_i), G(Y_i)), \quad \hat{\beta}^{\epsilon}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \psi^{\epsilon}_{j,k}(F(X_i), G(Y_i)).$$

When the marginal distributions F and G are unknown, it is not possible to estimate these coefficients. At this time, it is suggested to use these unknown distributions functions by their corresponding empirical distributions functions F_n and G_n. The modified empirical coefficients are

$$\tilde{\alpha}_{j_0,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j_0,k}(F_n(X_i), G_n(Y_i)), \quad \tilde{\beta}^{\epsilon}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \psi^{\epsilon}_{j,k}(F_n(X_i), G_n(Y_i)). \tag{2.6}$$

where F_n and G_n are given by

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} I(X_i \leq x), \qquad G_n(y) = \frac{1}{n} \sum_{k=1}^{n} I(Y_i \leq y)$$

So, the copula density estimator based on the wavelet is presented as follows

$$\tilde{c}_{k'}(u,v)=\sum_{k}\tilde{\alpha}_{j_0,k}\phi_{j_0 k}(u,v)+\sum_{i=i_1}^{i_2}\sum_{j\in Z}\tilde{\beta}^{\epsilon}_{i,j}1_{\{|\tilde{\beta}^{\epsilon}_{i,j}|\geq d\lambda_n\}}\psi^{\epsilon}_{i,j}(u,v),\qquad u,v\in[0,1]. \tag{2.7}$$

where the smoothing parameters i_1 and i_2 satisfying $2^{i_1}\simeq\log_2^n$ and $2^{i_2}\simeq n(\log_2^n)^{-2}$.

3 Main Results

At the beginning of this section, we formulate some basic assumptions that are needed to prove the main results.

3.1 Assumptions

- **A1** There is some compact subset Ω_x such that for any $c>0$, we have

$$\sup_{y\in\Omega_x} f(x,y)\leq c$$

- **A2** There exists a known constant $c'>0$ such that $\upsilon(x)\geq c'>0$

These assumptions are only technical and necessary to prove our main theorem.

3.2 Main Results

In the theorems that follow, two upper bounds have been obtained for the convergence rate of the wavelet estimators of the bivariate density function and the conditional density function. In addition to these, two propositions have been presented that are used to prove these theorems.

Theorem 3.1. Suppose that **A1** hold and let $c\in B^s_{p,q}(M)$ for all $M>0$, $s>2/p$ and $p,q\geq 1$. Consider the estimator $\tilde{c}_{k'}(u,v)$ characterized by (2.7) with i_1 and i_2 satisfying $2^{i_1}\simeq\log_2^n$ and $2^{i_2}\simeq n(\log_2^n)^{-2}$. Then there exists a constant $C>0$ such that

$$E\left(\int_{[0,1]}|\ \tilde{c}_{k'}(u_n,v_n)-c(u,v)\ |^2\ dy\right)\leq C\,(n/\log n)^{-\frac{2s}{1+2s}}$$

Theorem 3.2. Suppose that **A1** and **A2** hold and let $c\in B^s_{p,q}(M)$ for all $M>0$, $s>2/p$ and $p,q\geq 1$. Consider the estimation of conditional density function $f(y\mid x)$ with estimates (2.5) and (2.7). Then for a large enough d, there exists a value of $C>0$ such that we have

$$E\left(\int_{[0,1]}|\ \hat{\upsilon}(y\mid x)-\upsilon(y\mid x)\ |^2\ dy\right)\leq C\,(n/\log n)^{-\frac{2s}{1+2s}}$$

Theorems 3.1 and 3.2 provide theoretical guarantees about the near optimal convergence rate for estimators (2.7) and (2.2) under mild boundedness assumptions. In addition, it should be noted that in Theorem 3.2, the value of $n^{-\frac{2s}{1+2s}}$ is the standard near optimal convergence rate for the nonlinear hard threshold wavelet estimator in the standard one-dimensional density estimation problem (see [2]).

4 Proof of Theorems

The proof of Theorem 3.1 is based on a suitable decomposition of the MISE and the statistical properties of (2.6) presented in Lemma 2 and Lemma 3 which are discussed in [10]. Also to prove Theorem 3.2, we follow the lines of [3], Proof of Theorem 3.1. In the proof process, we assume C as a constant value that varies from line to line.

Recall from (2.5) and (2.7) that $\hat{c}_{k'}$ and $\tilde{c}_{k'}$ are estimators of c based on the unobservable pseudo-data $(U, V) = (F(X_i), G(Y_i))$ and their approximations $(U_n, V_n) = (F_n(X_i), G_n(Y_i))$, respectively. The main ingredient of the proof is the decomposition:

$$\begin{aligned}\hat{v}(y \mid x) - v(y \mid x) &= \hat{g}(y)\tilde{c}_{k'}(u_n, v_n) - g(y)c(u, v) \\ &= \Big\{\hat{g}(y) - g(y)\Big\}\tilde{c}_{k'}(u_n, v_n) - g(y)\Big\{\tilde{c}_{k'}(u_n, v_n) - c(u, v)\Big\}\end{aligned}$$

Therefore, it can be concluded that

$$\|\hat{v}(y \mid x) - v(y \mid x)\|_2^2 \leq \|\hat{g}(y) - g(y)\|_2^2 \|\tilde{c}_{k'}(u_n, v_n)\|_2^2 - C\|\tilde{c}_{k'}(u_n, v_n) - c(u, v)\|_2^2 = K_1 + K_2$$

Now, according to Assumption A2 and using the result of Theorem 3.1, it can be shown

$$K_2 \leq C\,(n/\log n)^{-\frac{2s}{1+2s}} \tag{4.1}$$

Now, to find an upper bound for K_1, we continue to decompose it with the centrality at fixed locations

$$\begin{aligned}K1 =& \|\hat{g}(y) - g(y)\|_2^2 \|\tilde{c}_{k'}(u_n, v_n) - \hat{c}_{k'}(u_n, v_n)\|_2^2 \\ &+ \|\hat{g}(y) - g(y)\|_2^2 \|\hat{c}_{k'}(u, v) - c(u, v)\|_2^2 + \|\hat{g}(y) - g(y)\|_2^2 \|c(u, v)\|_2^2\end{aligned} \tag{4.2}$$

Since $\hat{g}$ is a block wavelet estimator for the density function g, it can be derived from [8] that

$$\|\hat{g}(y) - g(y)\|_2^2 = O\left((n/\log n)^{-\frac{2s}{1+2s}}\right) \tag{4.3}$$

It can also be shown from source [10], theorems 3.1 and 3.2 that

$$\|\tilde{c}_{k'}(u_n, v_n) - \hat{c}_{k'}(u_n, v_n)\|_2^2 \leq C\,(n/\log n)^{-\frac{2s}{1+2s}} \tag{4.4}$$

and

$$\|\hat{c}_{k'}(u, v) - c(u, v)\|_2^2 \leq C\,(n/\log n)^{-\frac{2s}{1+2s}} \tag{4.5}$$

Therefore, according to the results (4.2), (4.4) and (4.5), as well as the boundedness assumption on $c(u, v)$ and by Combining these three bounds together based on relation (4.3), we complete the proof of Theorem 3.2.

Acknowledgments

The author would like to acknowledge Gonbad Kavous University for the partial support of this research through a Discovery Research Grant with number (6.02.77).

Disclouser statement

No potential conflict of interest was reported by the author(s).

References

[1] D. Bosq, *Nonparametric Statistics for Stochastic Processes*, second edition, Springer, New York, 1998.

[2] C. Chesneau and H. Doosti, *A note on the adaptive estimation of a conditional continuous-discrete multivariate density by wavelet methods*, Chinese J. Math. **2** (2016), 1–8.

[3] O.P. Faugeras. *A quantile-copula approach to conditional density estimation*, J. Multivar. Anal. **100** (2009), 2083–2099.

[4] H. Joe, *Multivariate Models and Dependence Concepts*, Chapman & Hall, London, 1997

[5] G. Lugosi and L. Devroye, *Combinatorial Methods in Density Estimation*, Springer-Verlag, New York, 1987.

[6] S. Mallat, *A Wavelet Tour of Signal Processing, the Sparse Way*, third edition, Academic Press, 2009.

[7] D. Picard A. Tsybakov W. Hardle, and G. Kerkyacharian, *Wavelets, approximations, and statistical applications*, Lecture Notes in Statistics. Springer-Verlag, Berlin, 1998.

[8] E. Shirazi and Y.P. Chaubey, *Non-negative density estimation via wavelet block thresholding for biased data*, J. Statist. Theory Pract. **11** (2019), 13.

[9] E. Shirazi and O.P. Faugeras, *A new wavelet-based estimation of conditional density via block threshold method*, Commun. Statist. Theory Meth. **53** (2024), no. 22, 8155–8165.

[10] E. Shirazi, B. Ghanbari, and M. Yarmohammadi, *Wavelet block thresholding for copula density estimation under biased sampling*, J. Statist. Comput. Simul. **93** (2023), no. 14, 2512–2533.

[11] A. Sklar, *Fonctions de repartition a n dimensions et leurs marges*, Publ. Inst. Statist, Univ, Paris **8** (1959), 229–231.

[12] Q. Yao and J. Fan, *Nonlinear Time Series*, Springer, New York, 2005.