isid/ms/2011/18
September 12, 2011
http://www.isid.ac.in/~statmath/eprints

Nonparametric estimation of two dimensional continuous-discrete density function by wavelets with an application to competing risks

Christophe Chesneau
Isha Dewan
Hassan Doosti

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Nonparametric estimation of two dimensional continuous-discrete density function by wavelets with an application to competing risks

Christophe Chesneau, Isha Dewan and Hassan Doosti
LMNO, Université de Caen Basse-Normandie, France,
Indian Statistical Institute, New Delhi, India
Tarbiat Moallem University, Tehran, Iran

January 9, 2012

Abstract

We consider the estimation of a two dimensional continuous-discrete density function with applications to competing risks. We construct two new wavelet estimators (non-adaptive and adaptive) for the joint density function taking into account this special continuous-discrete structure. The rates of convergence of the proposed estimators are established under the $\mathbb{L}_2$ risk over Besov balls. Our main result proves that our adaptive wavelet estimator (based on hard thresholding) attains a sharp rate of convergence. A simulation study illustrates the usefulness of the proposed estimators.

Key words and phrases: Density estimation, competing risks, wavelets, adaptivity, hard thresholding.

1 Introduction

Probability density function (pdf) estimation plays an important role in statistical data analysis and has application to many fields. Since the pdf characterizes several population characteristics - it is of interest to estimate the density function. Among the most widely used techniques of density estimation is the "kernel method" developed by Rosenblatt (1956), Parzen (1962), Cacoullos (1966), Van Ryzin (1969), Prakasa Rao (1983), among others. The techniques have also been extended to estimation of multivariate density function.

Many a times the bivariate random variable of interest is of a special type - one marginal is discrete and the other is continuous. Such random variables arise in survival analysis where one studies the failure/censoring time and an indicator variable indicating whether the event was failure or censoring. Or one could consider the length of marriage (continuous random variable) along with the reason of end of marriage - divorce or death (discrete random variable). Ahmed and Cerrito (1994) considered nonparametric estimation using kernels of the joint pdf of a random vector (X, Y) where X is continuous and Y is discrete. They studied the basic properties of the density estimator they proposed. Li and Racine (2003) defined an estimator for joint pdfs over mixed discrete and continuous variables. They derived the rate of convergence and also established the asymptotic normality of the resulting kernel density estimator. Li *et. al.* (2009) used the earlier results and considered the problem of testing for equality of two density functions over mixed discrete and continuous variables.

In the last twenty years, the subject of nonparametric density estimation has been enriched by considerable mathematical advance in the theory of wavelets. The early papers on density estimation using wavelets by Doukhan and Leon (1990), Walter (1992) and Kerkyacharian and Picard (1992) dealt with linear estimators in *i.i.d.* setting and established rates of convergence for the $\mathbb{L}_p$ risk. Walter and Ghorai (1992) discussed the advantages and disadvantages of wavelet based density estimation. The reader can refer to Härdle *et al.* (1998) and Vidakovic (1999) for a detailed coverage of wavelet theory in statistics. Chaubey *et al.* (2011) gives details of work done in wavelet based linear density estimation in the last two decades.

In this paper we consider a wavelet-based density estimation for a random vector (X, Y) where X is continuous and Y is discrete. We investigate the estimation of such a two dimensional density in the framework of competing risks. We develop two new wavelet estimators: a linear estimator and a non-linear adaptive hard thresholding estimator. We evaluate their performance by taking the $\mathbb{L}_2$ risk over Besov balls. We prove that the hard thresholding wavelets attains the rate of convergence $(\ln n/n)^{2s/(2s+1)}$ where s denotes the smoothness parameter of the density. It is interesting to note that this rate is the standard one for the one dimensional hard thresholding wavelet estimator (see Donoho *et al.* (1996)). A short simulation study illustrates the performances of the proposed estimators.

The paper is organized as follows. In section 2 we introduce the competing risks set up. The two dimensional wavelets are stated in section 3. We introduce our wavelet estimators applicable to competing risks data and investigate some of their theoretical and practical properties in Sections 4 and 5, respectively. Section 6 contains the details of the proofs and is followed by an appendix.

2 Competing risks

Consider a series system consisting of m components. The system fails as soon as the first component fails. Information consists of the system failure time and a marker indicating failure of which component resulted in system failure. Thus, competing risks data consists of the failure time of a unit (say T) and the cause of failure δ where $T > 0$ and $\delta = j$ with $j \in \{1, \ldots, m\}$. Or T could be the number of years spent in an unemployment registry and δ indicates the job that the individual got. Crowder (2001) gives a comprehensive review of the theory of competing risks. The joint distribution of (T, δ) is given by the sub-distribution function

$$F(t, j) = P(T \leq t, \delta = j),\ j = 1, 2, \ldots, m.$$

Note that $\sum_{j=1}^{m} F(t, j) = F(t)$, where $F(t)$ is the distribution function of the failure time T. We set $p_j = P(\delta = j)$, the probability that the individual fails due to jth failure. Since an individual fails due to only one of the m risks we have $\sum_{j=1}^{m} p_j = 1$.

The joint pdf of (T, δ) is given by the sub-density function

$$f(t, j) = \frac{\partial}{\partial t} F(t, j), \ \ j = 1, 2, \ldots, m.$$

Note that $\sum_{j=1}^{m} f(t, j) = g(t)$, where g is the density function of the failure time T.

The cause specific failure rate of (T, δ) is given by

$$h(t, j) = \frac{f(t, j)}{1 - F(t)},\ j = 1, 2, \ldots, m.$$

Notice that $\sum_{j=1}^{m} h(t, j) = h(t)$, where $h(t)$ is the failure rate of T.

The sub-distribution functions, the sub-density functions and the cause specific failure rates defined above can be estimated on the basis of n $i.i.d.$ observations of (T, δ) denoted by $(T_1, \delta_1), \ldots, (T_n, \delta_n)$. Deshpande and Purohit (2005) review several test procedures for testing of equality of sub-distribution functions and equality of cause specific hazard rates.

In what follows we consider estimators for the sub-density function based on wavelets and study their properties.

We formulate the following assumptions:

- The support of f is included into $[-a, a]^2$ for a $a > 0$.

- There exists a constant $C > 0$ such that $\sup_{x \in \mathbb{R}} g(x) \leq C$.

3 Two dimensional wavelets

In this section we consider two dimensional wavelets that will be useful in constructing nonparametric estimators for the sub-density function $f(t, \delta)$.

Let $a > 0$, ϕ be a compactly supported "father" wavelet and ψ be a compactly supported "mother" wavelet ψ. For any $x \in \mathbb{R}$ and any integer j, let

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

Then there exists an integer τ and a set of consecutive integers D_j with cardinality proportional to 2^j such that the collection $\{\phi_{\tau,k}, k \in D_\tau;\ \psi_{j,k}, \quad j = \tau, \ldots, \infty,\ k \in D_j\}$ forms an orthonormal basis of $\mathbb{L}_2([-a, a])$ (the space of square integrable two dimensional functions on $[-a, a]$).

For the purpose of this paper, we consider the tensor product wavelet basis on $\mathbb{L}_2([-a, a]^2)$. Let us briefly recall the construction of such a basis (see, for instance, Vidakovic (1999), Antoine *et. al.* (2004) and Mallat (2009)). Let us define the tensor-product wavelets Φ, Ψ^1, Ψ^2 and Ψ^3 as

$$\Phi(x, y) = \phi(x)\phi(y), \quad \Psi^1(x, y) = \psi(x)\phi(y), \quad \Psi^2(x, y) = \phi(x)\psi(y), \quad \Psi^3(x, y) = \psi(x)\psi(y).$$

For any orientation $\ell \in \{1, 2, 3\}$, scale $j \geq \tau$ and spatial location $\mathbf{k} = (k_1, k_2) \in \Lambda_j = D_j^2$, we define the translated and scaled versions

$$\Phi_{j,\mathbf{k}}(x, y) = 2^j \Phi(2^j x - k_1, 2^j y - k_2), \qquad \Psi_{j,\mathbf{k},\ell}(x, y) = 2^j \Psi^\ell(2^j x - k_1, 2^j y - k_2).$$

For any integer $J_0 \geq \tau$, any function $f \in \mathbb{L}_2([-a,a]^2)$ can be expanded into a wavelet series

$$f(x,y) = \sum_{\mathbf{k}\in\Lambda_{J_0}} \alpha_{J_0,\mathbf{k}} \Phi_{J_0,\mathbf{k}}(\mathbf{x}) + \sum_{\ell=1}^{3}\sum_{j=J_0}^{\infty}\sum_{\mathbf{k}\in\Lambda_j} \beta_{j,\mathbf{k},\ell}\Psi_{j,\mathbf{k},\ell}(x,y),$$

where

$$\alpha_{J_0,\mathbf{k}} = \int\int f(x,y)\Phi_{J_0,\mathbf{k}}(x,y)dxdy, \qquad \beta_{j,\mathbf{k},\ell} = \int\int f(x,y)\Psi_{j,\mathbf{k},\ell}(x,y)dxdy$$

are the wavelet coefficients of f.

We say that a function f in $\mathbb{L}_2([-a,a]^2)$ belongs to the two dimensional Besov ball $\mathbf{B}^s_{p,r}(M)$ if and only if there exists a constant M_*, depending on M, such that the wavelet coefficients of f satisfy

$$\left(\sum_{\ell=1}^{3}\sum_{j=\tau}^{\infty}\left(2^{j(s+1-2/p)}\left(\sum_{\mathbf{k}\in\Lambda_j}|\beta_{j,\mathbf{k},\ell}|^p\right)^{1/p}\right)^r\right)^{1/r} \leq M_* < \infty,$$

with a smoothness parameter $s > 0$, and the norm parameters: $0 < p \leq \infty$ and $0 < r \leq \infty$. See Mallat (2009) for details.

4 Estimators

The most standard wavelet approach to estimate a density function $f \in \mathbb{L}_2([-a,a]^2)$ from n i.i.d. observations $(X_1,Y_1),\ldots,(X_n,Y_n)$ from continuous (X,Y) is to consider the following linear estimator:

$$\hat{f}^L(x,y) = \sum_{\mathbf{k}\in\Lambda_{J_0}} \hat{\alpha}_{J_0,\mathbf{k}}\Phi_{J_0,\mathbf{k}}(x,y),$$

where $(x,y) \in [-a,a]^2$, J_0 is a "suitable integer" and

$$\hat{\alpha}_{j,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n}\Phi_{j,\mathbf{k}}(X_i,Y_i). \tag{4.1}$$

Tribouley (1995) and Masry (1997) discussed their statistical properties, viz, uniform rates of almost sure convergence, optimality in the minimax sense, etc.

In our competing risks framework, since δ is a discrete random variable, the wavelet coefficients estimators (4.1) are not really adapted. For this reason, following the idea of Dong and Jiang (2009), we consider the following unknown step function:

$$f_*(t,\delta) = \sum_{v=1}^{m} p_v f(t|\delta_1 = v)\mathbf{1}_{\{v-1/2\leq \delta < v+1/2\}},$$

where $\mathbf{1}_{\mathcal{D}}$ is the indicator function over the set $\mathcal{D}$ and $f(t|\delta_1 = v)$ is the density of T conditionally to $\{\delta_1 = v\}$. Note that, for any $j \in \{1,\dots,m\}$, $f_*(t,j) = f(t,j)$.

We consider the following estimators of the unknown wavelet coefficients of f_*: for any $j \geq \tau$, any $k \in \Lambda_j$ and any $\ell \in \{1,2,3\}$,

$$\hat{\alpha}_{j,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n}\int_{\delta_i-1/2}^{\delta_i+1/2} \Phi_{j,\mathbf{k}}(T_i,\delta)d\delta, \qquad \hat{\beta}_{j,\mathbf{k},\ell} = \frac{1}{n}\sum_{i=1}^{n}\int_{\delta_i-1/2}^{\delta_i+1/2} \Psi_{j,\mathbf{k},\ell}(T_i,\delta)d\delta. \tag{4.2}$$

The choice of these estimators is motivated by the following statistical results. The proofs are postponed to Section 5.

Lemma 4.1 *Let $j \geq \tau$, $\mathbf{k} \in \Lambda_j$, $\ell \in \{1,2,3\}$, $\hat{\alpha}_{j,\mathbf{k}}$ and $\hat{\beta}_{j,\mathbf{k},\ell}$ be* (4.2)*. Then*

$$E(\hat{\alpha}_{j,\mathbf{k}}) = \alpha_{j,\mathbf{k}}, \qquad E(\hat{\beta}_{j,\mathbf{k},\ell}) = \beta_{j,\mathbf{k},\ell}.$$

Lemma 4.2 *Let $j \geq \tau$ such that $2^j \leq n$, $\mathbf{k} \in \Lambda_j$, $\ell \in \{1,2,3\}$, $\hat{\alpha}_{j,\mathbf{k}}$ and $\hat{\beta}_{j,\mathbf{k},\ell}$ be* (4.2)*. Then there exists a constant $C > 0$ such that*

$$E((\hat{\alpha}_{j,\mathbf{k}} - \alpha_{j,\mathbf{k}})^4) \leq C\frac{1}{n^2}2^{-2j}, \qquad E((\hat{\beta}_{j,\mathbf{k},\ell} - \beta_{j,\mathbf{k},\ell})^4) \leq C\frac{1}{n^2}2^{-2j}.$$

The proof of Lemma 4.2 is based on the Rosenthal inequality (see the appendix).

An immediate consequence of Lemmas 4.1 and 4.2 is

$$V(\hat{\alpha}_{j,\mathbf{k}}) \leq C\frac{2^{-j}}{n}, \qquad V(\hat{\beta}_{j,\mathbf{k},\ell}) \leq C\frac{2^{-j}}{n}.$$

Lemma 4.3 *Let $j \geq \tau$ such that $2^j \leq n/\ln n$, $\mathbf{k} \in \Lambda_j$, $\ell \in \{1,2,3\}$, $\hat{\alpha}_{j,\mathbf{k}}$ and $\hat{\beta}_{j,\mathbf{k},\ell}$ be* (4.2)*. Then there exists a constant $\kappa > 0$ such that*

$$P\left(|\hat{\alpha}_{j,\mathbf{k}} - \alpha_{j,\mathbf{k}}| \geq \frac{\kappa}{2}2^{-j/2}\sqrt{\frac{\ln n}{n}}\right) \leq 2\frac{1}{n^2}, \quad P\left(|\hat{\beta}_{j,\mathbf{k},\ell} - \beta_{j,\mathbf{k},\ell}| \geq \frac{\kappa}{2}2^{-j/2}\sqrt{\frac{\ln n}{n}}\right) \leq 2\frac{1}{n^2}.$$

The proof of Lemma 4.3 is based on the Bernstein inequality (see the appendix).

We investigate the estimation of f_* via two different wavelet estimators: a linear estimator denoted by $\hat{f}^L$ and a hard thresholding estimator denoted $\hat{f}^H$.

Under the a priori assumption that $f_* \in \mathbf{B}^s_{p,r}(M)$ with $p \geq 2$, we define the linear estimator $\hat{f}^L$ by

$$\hat{f}^L(t,\delta) = \sum_{\mathbf{k}\in\Lambda_{J_0}} \hat{\alpha}_{J_0,\mathbf{k}} \Phi_{J_0,\mathbf{k}}(t,\delta), \tag{4.3}$$

where $\hat{\alpha}_{J_0,\mathbf{k}}$ is given in (4.2) and J_0 is an integer satisfying

$$\frac{1}{2} n^{1/(2s+1)} < 2^{J_0} \leq n^{1/(2s+1)}.$$

For a review on the density estimation via the linear wavelet estimator, we refer to Chaubey *et al.* (2011).

Now let us suppose that the smoothness of f_* is not known a priori. We define the hard thresholding estimator $\hat{f}^H$ by

$$\hat{f}^H(t,\delta) = \sum_{\mathbf{k}\in\Lambda_\tau} \hat{\alpha}_{\tau,\mathbf{k}} \Phi_{\tau,\mathbf{k}}(t,\delta) + \sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j} \hat{\beta}_{j,\mathbf{k},\ell} \mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|\geq\kappa\theta_j\right\}} \Psi_{j,\mathbf{k},\ell}(t,\delta), \tag{4.4}$$

where $\hat{\alpha}_{\tau,\mathbf{k}}$ and $\hat{\beta}_{j,\mathbf{k},\ell}$ are as in (4.2), J_1 is the integer satisfying

$$\frac{1}{2}\frac{n}{\ln n} < 2^{J_1} \leq \frac{n}{\ln n},$$

κ is a large enough constant (the one exhibited in Lemma 4.3) and θ_j is the "threshold level dependent" defined by

$$\theta_j = 2^{-j/2}\sqrt{\frac{\ln n}{n}}.$$

The feature of the hard thresholding estimator is to only estimate the "large" unknown wavelet coefficients of f_* because they are the ones which contain the main characteristics of f_* (singularities etc). Details on the hard thresholding estimator for standard statistical models can be found in Härdle *et al.* (1998).

5 Results

5.1 Theoretical results

Theorems 5.1 and 5.2 below investigate the mean integrated square error of the linear estimator $\hat{f}^L(t,\delta)$ and the hard threshold estimator $\hat{f}^H(t,\delta)$.

Theorem 5.1 (Upper bound for $\hat{f}^L$) *Consider the model described in Section 2. Suppose that $f_* \in \mathbf{B}^s_{p,r}(M)$ with $s > 0$, $p \geq 2$ and $r \geq 1$. Let $\hat{f}^L$ be as defined (4.3). Then there exists a constant $C > 0$ such that*

$$E\left(\int\int\left(\hat{f}^L(t,\delta) - f_*(t,\delta)\right)^2 dtd\delta\right) \leq Cn^{-2s/(2s+1)}.$$

Theorem 5.2 (Upper bound for $\hat{f}^H$) *Consider the model described in Section 2. Let $\hat{f}^H$ be as defined in (4.4). Suppose that $f_* \in \mathbf{B}^s_{p,r}(M)$ with $r \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 2/p\}$. Then there exists a constant $C > 0$ such that*

$$E\left(\int\int\left(\hat{f}^H(t,\delta) - f_*(t,\delta)\right)^2 dtd\delta\right) \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

The proof of Theorem 5.2 is based on a suitable decomposition of the mean integrated square error and the statistical properties of (4.2) presented in Lemmas 4.1, 4.2 and 4.3.

Theorem 5.2 shows that, besides being adaptive, $\hat{f}^H$ attains a rate of convergence close to the one of $\hat{f}^L$.

Note that $(\ln n/n)^{2s/(2s+1)}$ is the standard rate of convergence for the hard thresholding wavelet estimator in the standard one dimensional density estimation problem (see Donoho *et al.* (1996)). It is faster than $(\ln n/n)^{2s/(2s+2)}$ which corresponds to the one attained by the standard two dimensional wavelet hard thresholding estimator (see e.g. Delyon and Jusditsky (1996)). This difference can be explained by the original construction of our estimator entirely adapted to the "continuous-discrete structure" of our density estimation problem.

A possible improvement of our procedure will be to consider another thresholding rule as the BlockJS one (see e.g. Cai (1999) and, for the multidimensional case, Chesneau *et al.* (2010)). However, several technical difficulties arise and it is not immediatly clear how to solve them. This needs further investigations that we leave for a future work.

5.2 A Simulation Study

In this section we study the performance of the estimators (4.3) and (4.4) and the smoothed version of linear estimator (4.3) after local linear regression (see e.g. Fan (1992)). This smooth version of linear wavelet estimator was proposed by Ramirez and Vidakovic (2010). We use the Daubechies-Lagarias algorithm, explained in detail in Ramirez and Vidakovic (2010). The codes are written in Matlab and are adopted from Ramirez and Vidakovic (2010).

We generate $n = 200$ deta points $T_i, i = 1, ..., n$, from Beta(2,3), distribution. The discrete random sample are generated from Binomial(1, t_i), so it is simple to see the common two dimensional density function is

$$f(t, \delta) = 12t^{1+\delta}(1-t)^{3-\delta},$$

where $\delta \in \{0, 1\}$ and $t \in [0, 1]$.

In this simulation study, we used Daubechies's compactly supported Symmlet 4 (see Daubechies (1992), p.198) and Coiflet 1 (see Daubechies (1992), p. 258) and primary resolution level $J_0 = 5$.

Figures (1) in the first column concerns the estimation of $f(t, 0)$ via the linear wavelet estimator, the wavelet estimator after thresholding and the smooth version of the linear one. Each figure depicts the true value of the sub-density function along with the estimators mentioned above. Similarly the figures (1) in the second column give the corresponding estimators for $f(t, 1)$ for $n = 200$.

All the figures illustrate the good performances of our proposed linear and nonlinear estimators. Both the linear wavelet estimator and the linear wavelet estimator after thresholding appear pretty close because J_0 has been chosen to have the best result in view of f. Note that the hard thresholding one has no tuning parameter, it is entirely adaptive.

In each case the estimators are close to the true sub-density functions except for a few initial values.

6 Proofs

Notations. For the sake of simplicity, for any $v \in \{1, \ldots, m\}$, we set

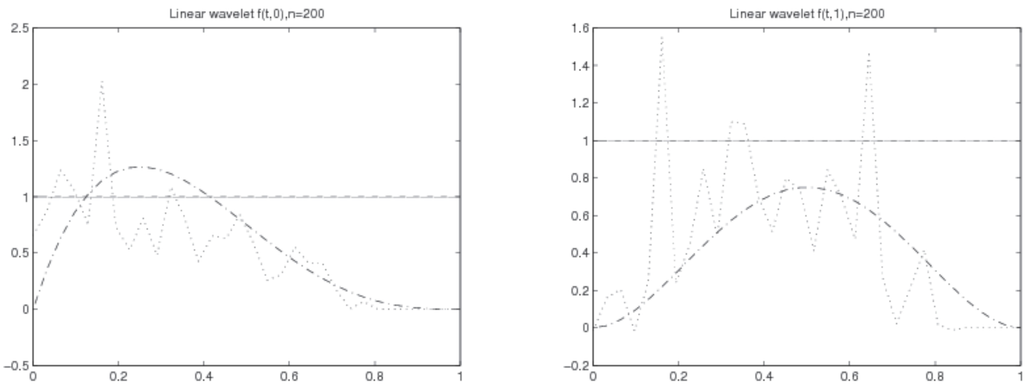
$$c_{j,k_2}(v) = \int_{v-1/2}^{v+1/2} \phi_{j,k_2}(\delta)d\delta. \tag{6.1}$$

Note that, for any $i \in \{1, \ldots, n\}$,

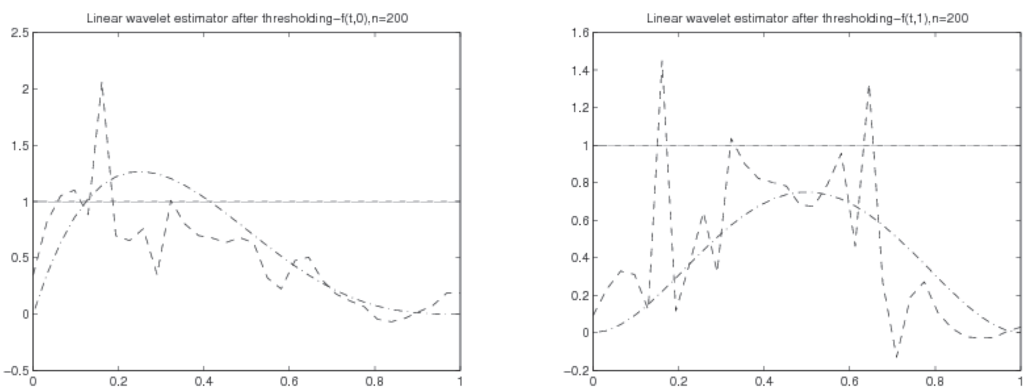
$$\int_{\delta_i-1/2}^{\delta_i+1/2} \Phi_{j,\mathbf{k}}(T_i, \delta)d\delta = \phi_{j,k_1}(T_i)c_{j,k_2}(\delta_i)$$

and we can write

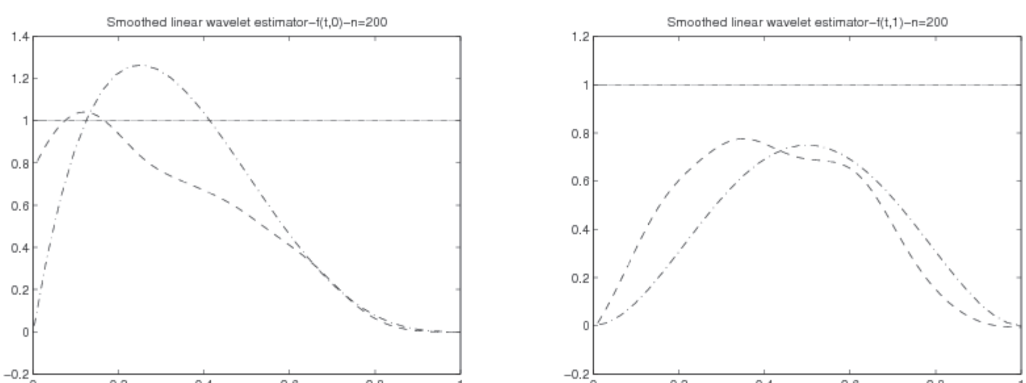
$$\hat{\alpha}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} \phi_{j,k_1}(T_i)c_{j,k_2}(\delta_i).$$



(a) Wavelet linear estimator for $f(t, \delta)$ with $\delta = 0$ and $\delta = 1$



(b) Wavelet hard thresholding estimator for $f(t, \delta)$ with $\delta = 0$ and $\delta = 1$



(c) Smooth version of our wavelet linear estimator for $f(t, \delta)$ with $\delta = 0$ and $\delta = 1$

Figure 1: Original (dashed-point) and estimated densities (dashed) using our wavelet estimators from $n = 200$ samples (T, δ).

In the following, the quantity C denotes a generic constant that does not depend on j, k and n. Its value may change from one term to another and may depends on ϕ or ψ.

Proof of Lemma 4.1. Let c_{j,k_2} be (6.1). We have

$$\begin{aligned} &E(\phi_{j,k_1}(T_1)c_{j,k_2}(\delta_1)) = E(E(\phi_{j,k_1}(T_1)c_{j,k_2}(\delta_1)|\delta_1)) \\ &= E(c_{j,k_2}(\delta_1)E(\phi_{j,k_1}(T_1)|\delta_1)) = E\left(c_{j,k_2}(\delta_1)\int \phi_{j,k_1}(t)f(t|\delta_1)dt\right) \\ &= \sum_{v=1}^{m} c_{j,k_2}(v)\int \phi_{j,k_1}(t)f(t|\delta_1 = v)dtp_v \\ &= \int\int \sum_{v=1}^{m} p_v f(t|\delta_1 = v)1_{\{v-1/2\leq\delta<v+1/2\}}\phi_{j,k_1}(t)\phi_{j,k_2}(\delta)dtd\delta \\ &= \int\int f_*(t,\delta)\phi_{j,k_1}(t)\phi_{j,k_2}(\delta)dtd\delta = \alpha_{j,\mathbf{k}}. \end{aligned}$$

Therefore

$$E(\hat{\alpha}_{j,\mathbf{k}}) = \alpha_{j,\mathbf{k}}.$$

Proceeding in a similar fashion, we prove that $E(\hat{\beta}_{j,\mathbf{k},\ell}) = \beta_{j,\mathbf{k},\ell}$.

•

Proof of Lemma 4.2. Let c_{j,k_2} be as in (6.1). Set, for any $i \in \{1,\ldots,n\}$,

$$\xi_i = \phi_{j,k_1}(T_i)c_{j,k_2}(\delta_i) - \alpha_{j,\mathbf{k}}.$$

Note that, since $(T_1,\delta_1),\ldots,(T_n,\delta_n)$ are *i.i.d.*, $\xi_1,\ldots,\xi_n$ are *i.i.d.*. Moreover, using Lemma (4.1), we have $E(\xi_1) = 0$. The Rosenthal inequality (see the appendix) yields

$$E((\hat{\alpha}_{j,\mathbf{k}} - \alpha_{j,\mathbf{k}})^4) = E\left(\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)^4\right) \leq C\left(\frac{1}{n^3}E(\xi_1^4) + \frac{1}{n^2}(E(\xi_1^2))^2\right). \tag{6.2}$$

For any $u \in \{2,4\}$, we have

$$E(\xi_1^u) \leq 2^u E((\phi_{j,k_1}(T_1))^u(c_{j,k_2}(\delta_1))^u).$$

By the change of variable $y = 2^j x - k_2$, we obtain

$$|c_{j,k_2}(\delta_1)| \leq \int |\phi_{j,k_2}(x)|dx = 2^{-j/2}\int |\phi(x)|dx = C2^{-j/2}. \tag{6.3}$$

Since g (the density of T_1) is bounded from above, ϕ is compactly supported and $\int(\phi_{j,k_1}(x))^2dx = 1$, we have

$$\begin{aligned} E((\phi_{j,k_1}(T_1))^u) &= \int(\phi_{j,k_1}(x))^u g(x)dx \leq C\int(\phi_{j,k_1}(x))^u dx \\ &\leq C2^{j(u-2)/2}\int(\phi_{j,k_1}(x))^2dx = C2^{j(u-2)/2}. \end{aligned}$$

Therefore

$$E(\xi_1^u) \leq C2^{j(u-2)/2}2^{-ju/2} = C2^{-j}. \tag{6.4}$$

It follows from (6.2), (6.4) and the fact that $2^j \leq n$

$$E((\hat{\alpha}_{j,\mathbf{k}} - \alpha_{j,\mathbf{k}})^4) \leq C\left(\frac{1}{n^3}2^{-j} + \frac{1}{n^2}2^{-2j}\right) \leq C\frac{1}{n^2}2^{-2j}.$$

•

Proof of Lemma 4.3. Let c_{j,k_2} be as in (6.1). Set, for any $i \in \{1,\dots,n\}$,

$$\xi_i = \phi_{j,k_1}(T_i)c_{j,k_2}(\delta_i) - \alpha_{j,\mathbf{k}}.$$

Then

- since $(T_1,\delta_1),\dots,(T_n,\delta_n)$ are *i.i.d.*, $\xi_1,\dots,\xi_n$ are *i.i.d.*,
- using Lemma (4.1), we have $E(\xi_1) = 0$,
- using (6.3), $|\xi_1| \leq |\phi_{j,k_1}(T_1)||c_{j,k_2}(\delta_1)| + |\alpha_{j,\mathbf{k}}| \leq C2^{j/2}2^{-j/2} + C \leq C$,
- using (6.4), we have $E(\xi_1^2) \leq C2^{-j}$.

It follows from the Bernstein inequality (see the appendix) and $2^j \leq n/\ln n$ that

$$\begin{aligned} &P\left(|\hat{\alpha}_{j,\mathbf{k}} - \alpha_{j,\mathbf{k}}| \geq \frac{\kappa}{2}2^{-j/2}\sqrt{\frac{\ln n}{n}}\right) = P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i\right| \geq \frac{\kappa}{2}2^{-j/2}\sqrt{\frac{\ln n}{n}}\right) \\ \leq\ & 2\exp\left(-C\frac{n\kappa^2\theta_j^2}{\mathbb{E}\left(\xi_i^2\right) + \kappa\theta_j}\right) = 2\exp\left(-C\frac{\kappa^2 2^{-j}\ln n}{2^{-j} + \kappa 2^{-j./2}\sqrt{\ln n/n}}\right) \\ \leq\ & 2\exp\left(-C\frac{\kappa^2\ln n}{1 + \kappa 2^{j./2}\sqrt{\ln n/n}}\right) \leq 2\frac{1}{n^{h(\kappa)}}, \end{aligned}$$

where $h(\kappa) = Ck^2/(1+\kappa)$. Since $\lim_{\kappa\to\infty} h(\kappa) = \infty$, there exists a $\kappa > 0$ such that $h(\kappa) = 2$.

Proceeding in a similar fashion, we prove that
$P\left(|\hat{\beta}_{j,\mathbf{k},\ell} - \beta_{j,\mathbf{k},\ell}| \geq (\kappa/2)2^{-j/2}\sqrt{\ln n/n}\right) \leq 2/n^2$.
This completes the proof of Lemma 4.3.

•

Proof of Theorem 5.1. Since f_* is a square integrable function, we can write

$$f_*(t,\delta) = \sum_{\mathbf{k}\in\Lambda_{J_0}} \alpha_{J_0,\mathbf{k}}\Phi_{J_0,\mathbf{k}}(t,\delta) + \sum_{\ell=1}^{3}\sum_{j=J_0}^{\infty}\sum_{\mathbf{k}\in\Lambda_j} \beta_{j,\mathbf{k},\ell}\Psi_{j,\mathbf{k},\ell}(t,\delta),$$

where $\alpha_{J_0,k} = \int\int f_*(t,\delta)\Phi_{J_0,k}(t,\delta)dtd\delta$ and $\beta_{j,\mathbf{k},\ell} = \int\int f_*(t,\delta)\Psi_{j,\mathbf{k},\ell}(t,\delta)dtd\delta$.

Therefore

$$\hat{f}^L(t,\delta) - f_*(t,\delta) = \sum_{\mathbf{k}\in\Lambda_{J_0}} \left(\hat{\alpha}_{J_0,\mathbf{k}} - \alpha_{J_0,\mathbf{k}}\right)\Phi_{J_0,\mathbf{k}}(t,\delta) - \sum_{\ell=1}^{3}\sum_{j=J_0}^{\infty}\sum_{\mathbf{k}\in\Lambda_j} \beta_{j,\mathbf{k},\ell}\Psi_{j,\mathbf{k},\ell}(t,\delta).$$

Using the orthonormality of the wavelet basis, Lemma (4.2), $\mathbf{B}^s_{p,r}(M) \subseteq \mathbf{B}^s_{2,\infty}(M)$ and the fact $p \geq 2$, we obtain

$$\begin{aligned}
&E\left(\int\int\left(\hat{f}^L(t,\delta) - f_*(t,\delta)\right)^2 dtd\delta\right)\\
&= \sum_{\mathbf{k}\in\Lambda_{J_0}} E\left((\hat{\alpha}_{J_0,\mathbf{k}} - \alpha_{J_0,\mathbf{k}})^2\right) + \sum_{\ell=1}^{3}\sum_{j=J_0}^{\infty}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell}\\
&= \sum_{\mathbf{k}\in\Lambda_{J_0}} V\left(\hat{\alpha}_{J_0,\mathbf{k}}\right) + \sum_{\ell=1}^{3}\sum_{j=J_0}^{\infty}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell}\\
&\leq C\left(2^{2J_0}\frac{2^{-J_0}}{n} + 2^{-2J_0 s}\right) \leq Cn^{-2s/(2s+1)}.
\end{aligned}$$

This ends the proof of Theorem 5.1.

•

Proof of Theorem 5.2. Since f_* is a square integrable function, we can write

$$f_*(t,\delta) = \sum_{\mathbf{k}\in\Lambda_\tau} \alpha_{\tau,\mathbf{k}}\Phi_{\tau,\mathbf{k}}(t,\delta) + \sum_{\ell=1}^{3}\sum_{j=\tau}^{\infty}\sum_{\mathbf{k}\in\Lambda_j} \beta_{j,\mathbf{k},\ell}\Psi_{j,\mathbf{k},\ell}(t,\delta),$$

where $\alpha_{\tau,\mathbf{k}} = \int\int f_*(t,\delta)\Phi_{\tau,\mathbf{k}}(t,\delta)dtd\delta$ and $\beta_{j,\mathbf{k},\ell} = \int\int f_*(t,\delta)\Psi_{j,\mathbf{k},\ell}(t,\delta)dtd\delta$.

Therefore

$$\begin{aligned}
&\hat{f}^H(t,\delta) - f_*(t,\delta)\\
&= \sum_{\mathbf{k}\in\Lambda_\tau}(\hat{\alpha}_{\tau,\mathbf{k}} - \alpha_{\tau,\mathbf{k}})\Phi_{\tau,\mathbf{k}}(t,\delta) + \sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j}\left(\hat{\beta}_{j,\mathbf{k},\ell}\mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|\geq\kappa\theta_j\right\}} - \beta_{j,\mathbf{k},\ell}\right)\Psi_{j,\mathbf{k},\ell}(t,\delta)\\
&- \sum_{\ell=1}^{3}\sum_{j=J_1+1}^{\infty}\sum_{\mathbf{k}\in\Lambda_j}\beta_{j,\mathbf{k},\ell}\Psi_{j,\mathbf{k},\ell}(t,\delta).
\end{aligned}$$

The orthonormality of the wavelet basis gives

$$\begin{aligned}
E\left(\int\int\left(\hat{f}^H(t,\delta) - f_*(t,\delta)\right)^2 dtd\delta\right) &= \sum_{\mathbf{k}\in\Lambda_\tau}E\left((\hat{\alpha}_{\tau,\mathbf{k}} - \alpha_{\tau,\mathbf{k}})^2\right)\\
+ \sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j}E\left(\left(\hat{\beta}_{j,\mathbf{k},\ell}\mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|\geq\kappa\theta_j\right\}} - \beta_{j,\mathbf{k},\ell}\right)^2\right) &+ \sum_{\ell=1}^{3}\sum_{j=J_1+1}^{\infty}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell}. \qquad (6.5)
\end{aligned}$$

Let us now bound these three terms.
Using Lemma 4.2, we obtain

$$\sum_{\mathbf{k}\in\Lambda_\tau}E\left((\hat{\alpha}_{\tau,\mathbf{k}} - \alpha_{\tau,\mathbf{k}})^2\right) \leq C2^{2\tau}\frac{2^{-\tau}}{n} \leq C\frac{1}{n} \leq Cn^{-2s/(2s+1)}. \qquad (6.6)$$

For $r \geq 1$ and $p \geq 2$, we have $\mathbf{B}^s_{p,r}(M) \subseteq \mathbf{B}^s_{2,\infty}(M)$. So

$$\sum_{\ell=1}^{3}\sum_{j=J_1+1}^{\infty}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell} \leq C\sum_{j=J_1+1}^{\infty}2^{-2js} \leq C2^{-2J_1 s} \leq C\left(\frac{\ln n}{n}\right)^{2s} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

For $r \geq 1$ and $p \in [1,2)$, we have $\mathbf{B}^s_{p,r}(M) \subseteq \mathbf{B}^{s+1-2/p}_{2,\infty}(M)$. Since $s > 2/p$, we have $s+1-2/p > s/(2s+1)$. Therefore

$$\begin{aligned}
\sum_{\ell=1}^{3}\sum_{j=J_1+1}^{\infty}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell} &\leq C\sum_{j=J_1+1}^{\infty}2^{-2j(s+1-2/p)} \leq C2^{-2J_1(s+1-2/p)}\\
&\leq C\left(\frac{\ln n}{n}\right)^{2(s+1-2/p)} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.
\end{aligned}$$

Hence, for $r \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 2/p\}$, we have

$$\sum_{\ell=1}^{3}\sum_{j=J_1+1}^{\infty}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \qquad (6.7)$$

Let us now bound the second term of (6.5). We can write

$$\sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j} E\left(\left(\hat{\beta}_{j,\mathbf{k},\ell}\mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|\geq\kappa\theta_j\right\}}-\beta_{j,\mathbf{k},\ell}\right)^2\right)=A_1+A_2+A_3+A_4, \tag{6.8}$$

where

$$A_1=\sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j} E\left(\left(\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}\right)^2\mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|\geq\kappa\theta_j\right\}}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|<\kappa\theta_j/2\right\}}\right),$$

$$A_2=\sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j} E\left(\left(\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}\right)^2\mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|\geq\kappa\theta_j\right\}}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|\geq\kappa\theta_j/2\right\}}\right),$$

$$A_3=\sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j} E\left(\beta_{j,\mathbf{k},\ell}^2\mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|<\kappa\theta_j\right\}}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|\geq 2\kappa\theta_j\right\}}\right)$$

and

$$A_4=\sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j} E\left(\beta_{j,\mathbf{k},\ell}^2\mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|<\kappa\theta_j\right\}}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|< 2\kappa\theta_j\right\}}\right).$$

Bounds for A_1 and A_3.

Using elementary inequalities, we establish the following three inclusion results:
$\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|<\kappa\theta_j,\ |\beta_{j,\mathbf{k},\ell}|\geq 2\kappa\theta_j\right\}\subseteq\left\{|\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}|>\kappa\theta_j/2\right\}$,
$\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|\geq\kappa\theta_j,\ |\beta_{j,\mathbf{k},\ell}|<\kappa\theta_j/2\right\}\subseteq\left\{|\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}|>\kappa\theta_j/2\right\}$ and
$\left\{|\hat{\beta}_{j,\mathbf{k},\ell}|<\kappa\theta_j,\ |\beta_{j,\mathbf{k},\ell}|\geq 2\kappa\theta_j\right\}\subseteq\left\{|\beta_{j,\mathbf{k},\ell}|\leq 2|\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}|\right\}$.
Therefore

$$\max(A_1,A_3)\leq C\sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j} E\left(\left(\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}\right)^2\mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}|>\kappa\theta_j/2\right\}}\right).$$

The Cauchy-Schwarz inequality combined with Lemma 4.2 and 4.3 gives

$$\begin{aligned}&E\left(\left(\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}\right)^2\mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}|>\kappa\theta_j/2\right\}}\right)\\ &\leq\ \left(E\left(\left(\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}\right)^4\right)\right)^{1/2}\left(P\left(|\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}|>\kappa\theta_j/2\right)\right)^{1/2}\leq C2^{-j}\frac{1}{n^2}.\end{aligned}$$

So, using $2^{J_1}\leq n$,

$$\max(A_1,A_3)\leq C\frac{1}{n^2}\sum_{j=\tau}^{J_1}2^{2j}2^{-j}\leq C\frac{1}{n^2}2^{J_1}\leq C\frac{1}{n}\leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \tag{6.9}$$

Bound for A_2.

By Lemma 4.2 we get

$$E\left(\left(\hat{\beta}_{j,\mathbf{k},\ell}-\beta_{j,\mathbf{k},\ell}\right)^2\right)\leq C\frac{2^{-j}}{n}.$$

Hence

$$A_2\leq C\frac{1}{n}\sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}2^{-j}\sum_{\mathbf{k}\in\Lambda_j}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|>\kappa\theta_j/2\right\}}.$$

Let J_2 be the integer defined by

$$\frac{1}{2}\left(\frac{n}{\ln n}\right)^{1/(2s+1)}<2^{J_2}\leq\left(\frac{n}{\ln n}\right)^{1/(2s+1)}. \tag{6.10}$$

We can bound A_2 as

$$A_2\leq A_{2,1}+A_{2,2},$$

where

$$A_{2,1}=C\frac{1}{n}\sum_{\ell=1}^{3}\sum_{j=\tau}^{J_2}2^{-j}\sum_{\mathbf{k}\in\Lambda_j}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|>\kappa\theta_j/2\right\}}$$

and

$$A_{2,2}=C\frac{1}{n}\sum_{\ell=1}^{3}\sum_{j=J_2+1}^{J_1}2^{-j}\sum_{\mathbf{k}\in\Lambda_j}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|>\kappa\theta_j/2\right\}}.$$

We have

$$A_{2,1}\leq C\frac{1}{n}\sum_{j=\tau}^{J_2}2^{-j}2^{2j}\leq C\frac{1}{n}2^{J_2}\leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

For $r\geq 1$ and $p\geq 2$, since $\mathbf{B}^s_{p,r}(M)\subseteq\mathbf{B}^s_{2,\infty}(M)$,

$$\begin{aligned}A_{2,2}&\leq C\frac{1}{n}\sum_{\ell=1}^{3}\sum_{j=J_2+1}^{J_1}2^{-j}\frac{1}{\theta_j^2}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell}\leq C\sum_{\ell=1}^{3}\sum_{j=J_2+1}^{\infty}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell}\\&\leq C\sum_{j=J_2+1}^{\infty}2^{-2js}\leq C2^{-2J_2s}\leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.\end{aligned}$$

For $r \geq 1$, $p \in [1, 2)$ and $s > 2/p$, since $\mathbf{B}^s_{p,r}(M) \subseteq \mathbf{B}^{s+1-2/p}_{2,\infty}(M)$ and $(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} A_{2,2} &\leq C\frac{1}{n}\sum_{\ell=1}^{3}\sum_{j=J_2+1}^{J_1} 2^{-j}\frac{1}{\theta_j^p}\sum_{\mathbf{k}\in\Lambda_j}|\beta_{j,\mathbf{k},\ell}|^p \\ &\leq C\left(\frac{\ln n}{n}\right)^{(2-p)/2}\sum_{j=J_2+1}^{\infty} 2^{-j(1-p/2)}2^{-j(s+1-2/p)p} \\ &\leq C\left(\frac{\ln n}{n}\right)^{(2-p)/2} 2^{-J_2(s+1/2-1/p)p} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

So, for $r \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1, 2)$ and $s > 2/p\}$, we have

$$A_2 \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \tag{6.11}$$

Bound for A_4

We have

$$A_4 \leq \sum_{\ell=1}^{3}\sum_{j=\tau}^{J_1}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|<2\kappa\theta_j\right\}}.$$

Let J_2 be the integer satisfying (6.10). We can bound A_4 as

$$A_4 \leq A_{4,1} + A_{4,2},$$

where

$$A_{4,1} = \sum_{\ell=1}^{3}\sum_{j=\tau}^{J_2}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|<2\kappa\theta_j\right\}}, \qquad A_{4,2} = \sum_{\ell=1}^{3}\sum_{j=J_2+1}^{J_1}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell}\mathbf{1}_{\left\{|\beta_{j,\mathbf{k},\ell}|<2\kappa\theta_j\right\}}.$$

We have

$$A_{4,1} \leq C\sum_{j=\tau}^{J_2}2^{2j}\theta_j^2 = C\frac{\ln n}{n}\sum_{j=\tau}^{J_2}2^{2j}2^{-j} \leq C\frac{\ln n}{n}2^{J_2} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

For $r \geq 1$ and $p \geq 2$, since $\mathbf{B}^s_{p,r}(M) \subseteq \mathbf{B}^s_{2,\infty}(M)$, we have

$$A_{4,2} \leq \sum_{\ell=1}^{3}\sum_{j=J_2+1}^{\infty}\sum_{\mathbf{k}\in\Lambda_j}\beta^2_{j,\mathbf{k},\ell} \leq C\sum_{j=J_2+1}^{\infty}2^{-2js} \leq C2^{-2J_2 s} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

For $r \geq 1$, $p \in [1, 2)$ and $s > 2/p$, since $\mathbf{B}^s_{p,r}(M) \subseteq \mathbf{B}^{s+1-2/p}_{2,\infty}(M)$ and

$(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} A_{4,2} &\leq C \sum_{\ell=1}^{3} \sum_{j=J_2+1}^{J_1} \theta_j^{2-p} \sum_{\mathbf{k}\in\Lambda_j} |\beta_{j,\mathbf{k},\ell}|^p \\ &= C \left(\frac{\ln n}{n}\right)^{(2-p)/2} \sum_{\ell=1}^{3} \sum_{j=J_2+1}^{J_1} 2^{-j(1-p/2)} \sum_{\mathbf{k}\in\Lambda_j} |\beta_{j,\mathbf{k},\ell}|^p \\ &\leq C \left(\frac{\ln n}{n}\right)^{(2-p)/2} \sum_{j=J_2+1}^{\infty} 2^{-j(1-p/2)} 2^{-j(s+1-2/p)p} \\ &\leq C \left(\frac{\ln n}{n}\right)^{(2-p)/2} 2^{-J_2(s+1/2-1/p)p} \leq C \left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

Consequently, for $r \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 2/p\}$, we have

$$A_4 \leq C \left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \tag{6.12}$$

It follows from (6.8), (6.9), (6.11) and (6.12) that

$$\sum_{\ell=1}^{3} \sum_{j=\tau}^{J_1} \sum_{\mathbf{k}\in\Lambda_j} E\left(\left(\hat{\beta}_{j,\mathbf{k},\ell} \mathbf{1}_{\left\{|\hat{\beta}_{j,\mathbf{k},\ell}| \geq \kappa\theta_j\right\}} - \beta_{j,\mathbf{k},\ell}\right)^2\right) \leq C \left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \tag{6.13}$$

Putting (6.5), (6.6), (6.7) and (6.13) together, we have, for $r \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 2/p\}$,

$$E\left(\int\int \left(\hat{f}^H(t,\delta) - f_*(t,\delta)\right)^2 dt d\delta\right) \leq C \left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

This completes the proof of Theorem 5.2.

•

Appendix

Here we state the two inequalities that have been used for proving the results in earlier section.

Lemma 6.1 (Rosenthal's inequality) *Let n be a positive integer, $p \geq 2$ and $U_1, \ldots, U_n$ be n zero mean independent random variables such that $\sup_{i\in\{1,\ldots,n\}} \mathbb{E}(|U_i|^p) < \infty$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left|\sum_{i=1}^{n} U_i\right|^p\right) \leq C\left(\sum_{i=1}^{n} \mathbb{E}\left(|U_i|^p\right) + \left(\sum_{i=1}^{n} \mathbb{E}\left(U_i^2\right)\right)^{p/2}\right).$$

Lemma 6.1 can be found in Rosenthal (1970).

Lemma 6.2 (Bernstein's inequality) *Let n be a positive integer and $U_1, \ldots, U_n$ be n zero mean independent random variables such that there exists a constant $M > 0$ satisfying $\sup_{i\in\{1,\ldots,n\}} |U_i| \leq M < \infty$. Then, for any $\lambda > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} U_i\right| \geq \lambda\right) \leq 2\exp\left(-\frac{\lambda^2}{2\left(\sum_{i=1}^{n} \mathbb{E}\left(U_i^2\right) + \lambda M/3\right)}\right).$$

Lemma 6.2 can be found in Petrov (1995).

Acknowledgements: We thanks the referee for his suggestions which lead to the improved version of the paper.

References

[1] Ahmed, I.A. and Cerrito, P.B. (1994), Nonparametric estimation of joint discrete-continuous probability densities with applications. *Journal of Statistical Planning and Inference*, **41**, 349-364.

[2] Antoine, J.P., Murenzi, R., Vandergheynst, P. and Ali, S.T. (2004). *Two-dimensional Wavelets and their relatives*, Cambrige, United Kingdom.

[3] Cai T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *The Annals of Statistics*, 27, 898-924.

[4] Chaubey, Y. P., Chesneau, C. and Doosti, H. (2011). Wavelet Linear Density Estimation: A review. *Journal of the Indian Society of Agricultural Statistics*, 65, (2), 169-179.

[5] Chesneau C., Fadili M.J. and Starck J.-L. (2010). Stein Block Thresholding For Image Denoising , *Applied and Computational Harmonic Analysis*, 28, 1, 67-88.

[6] Cacoullos, T. (1966). Estimation of a multivariate density. *Ann. Instit. Statist. Math.*, **18**, 179-189.

[7] Crowder, M.J. (2001). *Classical Competing Risks*. Chapman and Hall.

[8] Daubechies, I. (1992). *Ten Lectures on Wavelets*, CBMS-NSF regional conferences series in applied mathematics. SIAM, Philadelphia.

[9] Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators. *Applied Computational Harmonic Analysis*, 3, 215–228.

[10] Deshpande, J.V. and Purohit, S.G. (2005). *Life-time data: statistical models and methods* . World Scientific, New Jersey.

[11] Dong, J. and Jiang, R. (2009). Multinomial probability estimation by wavelet thresholding, *Communication in Statistics, Theory and Methods*, **38**, 1486-1507.

[12] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics*, 23, 508-539.

[13] Doukhan, P. and Leon, J.R. (1990). Une note sur la dé viation quadratique déstimateurs de densités par projections orthogonales. .*C.R. Acad. Sci. Paris,* série I, Math. **310**, 425–430.

[14] Fan, J. (1992). Design-adaptative nonparametric regression. *Journal of the American Statistical Association*, 87, 998-1004.

[15] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximations, and Statistical Applications,* Lecture Notes in Statistics, **129**, Springer- Verlag, Berlin.

[16] Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces. *Stat. Prob. Lett.*, **13**, 15–24.

[17] Li, Q. and Racine, J.S. (2003). Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis*,**86**, 266-292.

[18] Li, Q., Maasoumi, E. and Racine J.S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics*, **148**, 186-200.

[19] Mallat, S. (2009). *A wavelet tour of signal processing, the sparse way*, Academic Press, Third Edition.

[20] Masry, E. (1997). Multivariate probability density estimation by wavelet methods: strong consistency and rates for stationary time series. *Stochastic Processes and Their Application*, **67**, 177-193.

[21] Parzen, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist.*, **33**, 1065-1076.

[22] Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford: Clarendon Press.

[23] Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic Press, New York.

[24] Ramirez, P. and Vidakovic, B. (2010). Wavelet density estimation for stratified size-biased sample. *Journal of Statistical Planning and Inference*, 140, 2, 419-432.

[25] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of density functions. *Ann. Math. Stat.*, **27**, 832–837.

[26] Rosenthal, H.P. (1970). On the subspaces of $\mathbb{L}_p$ $(p \geq 2)$ spanned by sequences of independent random variables, *Israel Journal of Mathematics*, **8**, 273-303.

[27] Tribouley, K. (1995). Practical estimation of multivariate densities using wavelet methods. *Statistica Neerlandica.* **49**, 41–62.

[28] Vanryzin, J. (1969). On strong consistency of density estimates. *Ann. Math. Statist.*, **40**, 1765-1771.

[29] Vidakovic, B. (1999). *Statistical Modeling by Wavelets,* Wiley, New York.

[30] Walter, G. (1992). Approximation of the delta function by wavelets, J. Approx. Theory, **3**, 329-343.

[31] Walter, G. and Ghorai, J. (1992). Advantages and disadvantages of density estimation with wavelets. *Proceedings of the 24th Symp. on the Interface,* Ed. H. Joseph Newton, Interface FNA, VA, 24, 234–343.