AN ADAPTIVE ESTIMATOR OF THE DENSITY
OF COMPONENTS OF A MIXTURE

UDC 519.21

D. I. POKHYL’KO

Abstract. Linear and nonlinear wavelet estimators of the density of components of a mixture are considered in the paper. The rate of convergence in the uniform metric and large deviation probabilities are obtained for linear estimators. The limit behavior of the threshold-based estimator is considered for the integral metric. An adaptive modification of the threshold-based estimator is constructed.

1. Introduction

Mixtures of probability distributions are used in statistics as a mathematical model of the distribution of real data. The classical mixture model is given by

\[ P\{\xi \in A\} = w^1 H_1(A) + w^2 H_2(A) + \cdots + w^M H_M(A), \]

where \( M \) is the number of components of a mixture, \( H_m, 1 \leq m \leq M, \) are given probability distributions, and \( w^m \) is the concentration of the component \( m \) in the mixture. These models are applied in the cases where

(i) every observation is known to belong to one of \( M \) different classes (populations, components) but the classes vary with observations,
(ii) the distribution of an observed characteristic (property) denoted by \( \xi \) is \( H_m \) for the component \( m, \) and
(iii) the probability that an observation belongs to the component \( m \) is \( w^m.\)

If the correct classification of the observations is unknown, we obtain the classical model of mixtures (see [4]).

We construct projective estimators of the density of components of the mixture by using a wavelet basis for the case where observations form a sample from a mixture with varying concentrations. Some applications of such models can be found in [4]. One can find further references in [4].

We assume that a sample \( \{\xi_j\}_{j=1}^N \) is formed by independent random variables with absolute continuous distributions. We also assume that the density of the distribution of \( \xi_j \) is given by

\[
 p^j(x) = \sum_{m=1}^{M} w_j^m p_m(x),
\]

where \( M \) is the number of components of a mixture, \( p_m \) is the density of the component \( m, \) and \( w_j^m \) is the probability that an observation \( j \) belongs to the component \( m \) (this implies

2000 Mathematics Subject Classification. Primary 62G07; Secondary 42C40.

Key words and phrases. Wavelets, mixture, estimator of the density, adaptive estimator, projective estimator.
that $w^m_j \geq 0$ for all $j, m$ and that $\sum_{m=1}^M w^m_j = 1$ for all $j$). Below we assume that the sample is of a fixed size $N$; therefore we omit the subscript $N$ if this creates no confusion (otherwise we write $N$ after a colon, like $w^m_{j:N}$).

We study the behavior of a nonlinear (namely, the threshold-based) estimator for the case where the densities of the components belong to some Besov space $B_{spq}$. In Section 1 we consider an estimator that requires the knowledge of the parameters $s$, $p$, $q$, and $\|s_m\|_\infty$, $m = 1, \ldots, M$. We also consider a version of this estimator that does not depend on $s$, $p$, and $q$ but still depends on $\|s_m\|_\infty$, $m = 1, \ldots, M$. We show that, up to a certain multiplicative factor (being bounded under some assumptions), the asymptotic behavior of this estimator in $L_p$ is the same as that of the estimator studied in [8] for the case of independent identically distributed observations. It is worthwhile mentioning that the asymptotic behavior of this estimator is the best up to a logarithmic factor as shown in [8] for the case of independent identically distributed observations.

In Section 2 we consider the convergence in the uniform metric of the linear estimator, a particular case of the nonlinear estimator. We estimate the expectation of the supremum of the error and the large deviation probability for the error. Using these results, we construct a statistic in Section 3 and substitute it in the estimator studied in Section 1 instead of the unknown value $\|s_m\|_\infty$. We also show that this substitution does not decrease the rate of convergence, which means that we obtain an adaptive estimator which does not depend on unknown parameters. An example exhibiting the behavior of the estimator for simulated data is discussed in Section 5.

Wavelet estimators for the density of components of a mixture are studied in [5]. The results of Sections 1 and 3 are generalizations of their counterparts in [5] for a more general class of spaces and norms. The result of Section 1 is a generalization of some results of [8] to the case of a mixture of distributions.

Linear and nonlinear wavelet estimators of the density of homogeneous observations are considered in [8] [9] [10]. Wavelet estimators of the density form a particular case of projective estimators whose properties are considered in [11] for homogeneous observations. The theory of wavelet analysis can be found in [2] [7] [9]. Various problems related to the analysis of mixtures are studied in [4]. The problem of estimation of the distribution functions of components of a mixture is investigated in [3]. Kernel estimators of the densities of components of mixtures are treated in [6].

In what follows, the symbol $\|\cdot\|_p$ stands for the $L_p$-norm, while the relation $2^{j(n)} \simeq g(n)$ means that $2^{j(n)} \leq g(n) < 2^{j(n)+1}$.

2. PARTIALLY ADAPTIVE ESTIMATOR: THE BEHAVIOR IN THE INTEGRAL METRIC

Let $W = (w^m_j)_{j=1, m=1}^{N,M}$ be a matrix with $M$ columns and $N$ rows and let

$$e^m = (0, \ldots, 0, 1, 0, \ldots, 0)^T$$

be the $m$-th basis vector in $\mathbb{R}^M$. Put

$$a^m = W\Gamma^{-1}e^m$$

where $\Gamma = N^{-1}W^TW$. We assume throughout the paper that $\det \Gamma > 0$ (more details concerning this condition can be found in [4]).

For all $x, y \in \mathbb{R}^N$ let

$$\langle x, y \rangle_N = \frac{1}{N} \sum_{i=1}^N x_i y_i, \quad \|x\|_N = \sqrt{\langle x, x \rangle_N}.$$  

If $w^m = (w^m_1, \ldots, w^m_M)^T$, $m = 1, \ldots, M$, then $\Gamma$ is the Gram matrix of the system $\{w^m, m = 1, \ldots, M\}$ for the scalar product $\langle \cdot, \cdot \rangle_N$.
Definitions of wavelets and Besov spaces can be found in [9]. The following expansion plays an important role in our further consideration. Let $\phi$ and $\psi$ be the $f$- and $m$-wavelets, respectively. Then for all $j_0 \in \mathbb{Z}$, every $p_m(x) \in L_2(\mathbb{R})$ can be expressed in terms of the following series (that converges in $L_2(\mathbb{R})$):

$$p_m = \sum_k \alpha_{j_0k}^m \phi_{j_0k} + \sum_{j \geq j_0} \sum_k \beta_{jk}^m \psi_{jk},$$

where $\alpha_{j_0k}^m = \int p_m(x) \phi_{j_0k}(x) \, dx$, $\beta_{jk}^m = \int p_m(x) \psi_{jk}(x) \, dx$, $\phi_{j_0k}(x) = 2^{j/2} \phi(2^j x - k)$, and $\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$.

The estimator $\tilde{p}_m$ of the density $p_m$ can be written as

$$\tilde{p}_m(x) = \sum_{k \in \mathbb{Z}} \tilde{\alpha}_{j_0k}^m \phi_{j_0k}(x) + \sum_{j = j_0}^{j_1} \sum_{k \in \mathbb{Z}} \tilde{\beta}_{jk}^m \psi_{jk}(x)$$

where $j_0$ and $j_1$ are parameters of the estimator to be chosen according to the properties of the density and the size of the sample. The variables $\tilde{\alpha}_{j_0k}^m$ and $\tilde{\beta}_{jk}^m$ are defined by

$$\tilde{\alpha}_{j_0k}^m = \frac{1}{N} \sum_{i=1}^{N} \phi_{jk}(\xi_i) \alpha_{i}^m, \quad \tilde{\beta}_{jk}^m = \frac{1}{N} \sum_{i=1}^{N} \psi_{jk}(\xi_i) \alpha_{i}^m, \quad \tilde{\beta}_{jk}^m = \tilde{\beta}_{jk}^m \mathbb{1}\{|\tilde{\beta}_{jk}| > t_j\}$$

where $\alpha_{i}^m$ is the coordinate $i$ of the vector $\alpha^m$ defined by [2] and $t_j$ is the parameter to be specified below. An unbiased estimator for distributions is constructed in [3] by using similar weighted sums. It is shown in [5] that $\tilde{\alpha}_{jk}$ and $\tilde{\beta}_{jk}$ are unbiased estimators for $\alpha_{jk}^m$ and $\beta_{jk}^m$. If $t_j = 0$ (no truncation), then $\tilde{p}_m = \sum_k \tilde{\alpha}_{jk+1} \phi_{j+1k}$ is the linear estimator at level $j_1 + 1$ (see [9]).

Put

$$\Pr_{j_0} p_m = \sum_k \alpha_{j_0k}^m \phi_{j_0k}, \quad \Pr_{j_1} p_m = \sum_{j = j_0}^{j_1} \sum_k \beta_{jk}^m \phi_{jk},$$

$$\tilde{\Pr}_{j_0} p_m = \sum_k \tilde{\alpha}_{j_0k}^m \phi_{j_0k}, \quad \tilde{\Pr}_{j_1} p_m = \sum_{j = j_0}^{j_1} \sum_k \tilde{\beta}_{jk}^m \phi_{jk}.$$

It is worthwhile mentioning that if

$$V_j = \text{cls}\{\phi_{jk}, k \in \mathbb{Z}\} \quad \text{and} \quad W_{j_1}^{j_0} = \text{cls}\{\psi_{jk}, j = j_0, \ldots, j_1, k \in \mathbb{Z}\}$$

where “cls” stands for the “closed linear span”, then $\Pr_{j_0} p_m$ and $\tilde{\Pr}_{j_0} p_m$ are projections of $p_m$ in $L_2(\mathbb{R})$ and $V_{j_0}$ in $W_{j_1}^{j_0}$, respectively. Put

$$\nu = sp - \frac{p' - p}{2}, \quad s' = s - 1/p + 1/p', \quad \alpha = \begin{cases} s/(1 + 2s), & \nu \geq 0, \\ s'/(1 + 2s - 2/p), & \nu < 0, \end{cases}$$

$\text{diam} A \overset{\text{def}}{=} \sup_{x,y \in A} |x - y|$ for all $A \subset \mathbb{R}$ and let

$$B_{spq}(M', T, C_{\infty}), \quad \{ f \in B_{spq}, \| f \|_{spq} \leq M', \| f \|_{\infty} \leq C_{\infty}, \text{diam supp } f \leq T, f \text{ is a density}\}$$

where $B_{spq}$ is the Besov space equipped with the norm $\| \cdot \|_{spq}$.

---

Here and in what follows
\[ \Omega_j \overset{\text{def}}{=} \left\{ k \in \mathbb{Z} : \supp \phi_j \cap \supp \left( \sum_m p_m \right) \neq \varnothing \right\} \]
\[ \cup \left\{ k \in \mathbb{Z} : \supp \phi_j \cap \supp \left( \sum_m p_m \right) \neq \varnothing \right\}. \]

**Theorem 2.1.** Let \( p' \geq p \geq 1 \), \( s - p^{-1} > 0 \), and let \( \phi, \psi \in C^r \), \( r > s \), be f- and m-wavelets, respectively, whose supports are compact. If \( t_j = K\sqrt{j/N} \), then there are positive constants \( C = C(s,p,q,M',C_{\infty}) \) and \( K_0 \) such that
\[
\sup_{p_m} E \| \hat{p}_m - p_m \|_p^{p'} \leq C \left( \max_{i=1,\ldots,N} |a_i^{m'}|^{p'} + 1 \right) \begin{cases} 
(\ln N)^{(1-\nu/(sp'))\alpha p'} N^{-\alpha p'}, & \nu > 0, \\
(\ln N)^{(\nu'/2-p/q')^+} (\ln N)^{\alpha p'}, & \nu = 0, \\
(\ln N)^{\alpha p'}, & \nu < 0,
\end{cases}
\]
for
\[
2^{j_0} \simeq \left( N(\ln N)^{\nu'/2-1(\nu\geq 0)} \right)^{1-2\alpha}, \quad 2^{j_1} \simeq \left( N(\ln N)^{-1(\nu\leq 0)} \right)^{\alpha/s'},
\]
and \( K \geq K_0 \) where \( x^+ \overset{\text{def}}{=} \max(x,0) \), \( \hat{p}_m \) is defined by (3), and the supremum is considered over all densities \( p_m \in B_{spq}(M',T,C_{\infty}) \), \( m = 1, \ldots, M \).

**Remark 2.1.** The latter result is an existence theorem. It says that there are parameters \( K_0, j_0, j_1 \) that guarantee a given rate of convergence for the estimator (3). These parameters are expressed in terms of some (unknown) parameters of the density, namely in terms of \( K = c \) where \( c = c(C_{\infty}) \) (see Lemma 2.1, equality (11)), \( \gamma = \gamma(s,p,q) \) (see equality (20)), \( j_0 = j_0(s,p,q) \), and \( j_1 = j_1(s,p,q) \).

The following theorem claims that there is an estimator that does not require us to know some unknown variables (namely, the variables \( s, p, q \)) to determine parameters \( \gamma, j_0, j_1 \); this estimator can be used if parameters assume arbitrary values in a certain set. Nevertheless the estimator relies on the parameter \( K \) which still depends on \( C_{\infty} \). For this reason, the estimator is called partially adaptive.

**Theorem 2.2** (Partially adaptive estimator). Fix an integer number \( r_0 > 0 \) and define the class
\[ S \overset{\text{def}}{=} \{ (s,p,q) : 1/p < s \leq r_0, 1 \leq p \leq p', 1 \leq q \leq \infty \}. \]
Let \( \phi, \psi \in C^{r_0+1} \) be the f- and m-wavelets whose supports are compact and let the densities \( p_m \), \( m = 1, \ldots, M \), belong to the space \( B_{spq}(M',T,C_{\infty}) \) where \( (s,p,q) \in S \). If \( t_j = K\sqrt{j/N} \), then there are constants \( C = C(s,p,q,M,C_{\infty}) \) and \( K_0 \) such that
\[
\sup_{p_m} E \| \hat{p}_m - p_m \|_p^{p'} \leq C \left( \max_{i=1,\ldots,N} |a_i^{m'}|^{p'} + 1 \right) \begin{cases} 
(\ln N)^{(\nu'/2-p/q')^+} (\ln N)^{\alpha p'}, & \nu = 0, \\
(\ln N)^{\alpha p'}, & \nu \neq 0,
\end{cases}
\]
for \( 2^{j_0} \simeq N^{1/(1+2r_0)} \), \( 2^{j_1} \simeq N/\ln N \), and \( K \geq K_0 \) where \( x^+ \overset{\text{def}}{=} \max(x,0) \), \( \hat{p}_m \) is defined by (3), and the supremum is evaluated over all densities \( p_m \in B_{spq}(M',T,C_{\infty}) \), \( m = 1, \ldots, M \).
Remark 2.2. The asymptotic behavior of the estimator coincides with that obtained in [9] for the case of independent identically distributed observations up to the factor

$$\left( \max_{i=1,\ldots,N} |a_i^m| \right)^{p'}.$$ 

Since $\Gamma_{lm} = \langle w^l, w^m \rangle_N$, $|\Gamma_{lm}| \leq 1$ for all $l$ and $m$, whence $|a_i^m| \leq M! / \det \Gamma$.

We recall some other results needed in the proofs of theorems.

**Proposition 2.1** ([9] Corollaries 9.1 and 9.2). If $\phi$ is an $m$-wavelet with a compact support and $\phi \in C^r$, then for all $s$, $0 < s < r$, all $p$ and $q$, $1 \leq p, q < \infty$, and for all $f \in B_{spq}$ the norm $\| \cdot \|_{spq}$ is equivalent to the following norm in the space of coefficients of the wavelet expansions:

$$\|f\|_{spq} = \|\alpha_0\|_p + \left( \sum_{j=0}^{\infty} \left( 2^{j(s+1/2-1/p)} \|\beta_j\|_p \right)^q \right)^{1/q}.$$ 

Moreover the following inclusions hold:

1. $B_{s',p,q} \subset B_{spq}$ if $s' > s$ or if $s' = s$ and $q' \leq q$.
2. $B_{spq} \subset B_{s',p',q}$ for $p' \geq p$, $s' = s - 1/p + 1/p'$, $1 \leq p \leq p' \leq \infty$.

**Proposition 2.2** ([9] Proposition 8.3). Let $\theta_\phi(x) \equiv \sum_{k \in \mathbb{Z}} |\phi(x-k)|$ where an $f$-wavelet $\phi$ is such that $\text{ess sup}_x \theta_\phi(x) < \infty$. Then

$$\frac{1}{\|\theta_\phi\|_\infty^{1/p} \|\phi\|_1^{1/q}} \|\lambda\|_p 2^j(1/2-1/p) \leq \left| \sum_{k} \lambda_k \phi_{jk}(x) \right|_p \leq \|\theta_\phi\|_\infty^{1/q} \|\phi\|_1^{1/p} \|\lambda\|_p 2^j(1/2-1/p)$$

for an arbitrary sequence $\lambda_k$, $k \in \mathbb{Z}$, such that $\|\lambda\|_p < \infty$ where $1 \leq p, q \leq \infty$ and $1/p + 1/q = 1$.

**Proposition 2.3** (Bernstein inequality [9] Corollary C.1). Let $Y_i$ be independent random variables such that $|Y_i| \leq M$ where $M$ is a nonrandom constant, $\mathbb{E}Y_i = 0$, and $b_N^2 \equiv \sum_{i=1}^{N} \text{Var} Y_i < \infty$. Then

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

for all $\lambda \geq 0$. In particular, if $\mathbb{E}Y_i^2 \leq \sigma^2$, then

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

for all $\lambda \geq 0$.

**Proposition 2.4** (A corollary of the Rosenthal inequality [9] Theorem C.2). Let $Y_1, \ldots, Y_N$ be independent random variables such that $\mathbb{E}Y_i = 0$, $\text{Var} Y_i \leq \sigma^2$, and $|Y_i| \leq A$ where $A$ is a nonrandom constant. Given a number $m \geq 1$, there exists a constant $c_m$ such that

$$\mathbb{E} \left( \sum_{i=1}^{n} Y_i \right)^m \leq c_m \begin{cases} \frac{\sigma^2 A^{m-2}}{N^{m/2}} & m > 2, \\ \frac{\sigma^m}{N^{m/2}} & 1 \leq m \leq 2. \end{cases}$$
Lemma 2.1. If \( j2^j \leq N \), \( \max_m \|p_m\|_\infty \leq C_\infty \), and \( \gamma \geq 1 \), then there exists a constant \( c = c(a^m, C_\infty) \) such that

\[
P \left\{ |\hat{\beta}_{jk}^m - \beta_{jk}^m| > \frac{c\gamma}{2} \sqrt{\frac{j}{N}} \right\} \leq 2^{-j\gamma}.
\]

Proof. Put

\[
Y_i = a^m_i(\psi_{jk}(X_i) - E \psi_{jk}(X_i)).
\]

Then \( \sum \operatorname{Var} Y_i \leq C_\infty \langle a^m, a^m \rangle_N \), and \( |Y_i| \leq 2 \cdot 2^{j/2} \max_k |a_k^m| \|\psi\|_\infty \).

The Bernstein inequality implies that

\[
\operatorname{Var} Y_i \leq \frac{\mathbb{E} \hat{\beta}_{jk}^m - \beta_{jk}^m|^2}{\gamma N} \leq \frac{(\frac{4}{3}\|\psi\|_\infty \max_k |a_k^m| \ln 2)^2 + \frac{8}{\gamma}C_\infty \langle a^m, a^m \rangle_N}{2\gamma N}.
\]

Since \( \gamma \geq 1 \) and \( j2^j \leq N \), the latter inequality holds if

\[
(11) \quad c = \frac{4}{3}\|\psi\|_\infty \max_k |a_k^m| \ln 2 + \left( \frac{4}{3}\|\psi\|_\infty \max_k |a_k^m| \ln 2 \right)^2 + \frac{8}{\gamma}C_\infty \langle a^m, a^m \rangle_N.
\]

Lemma 2.2. If \( N > 2^j \) and \( \max_k \|p_k\|_\infty \leq C_\infty \), then there exists a constant \( C = C(s, p, q, C_\infty) \) such that

\[
\mathbb{E} |\hat{\beta}_{jk}^l - \beta_{jk}^l|^m \leq C \max_k |a_k^l|^m N^{-m/2}
\]

and

\[
\mathbb{E} |\hat{\beta}_{jk}^l - \beta_{jk}^l|^m \leq C \max_k |a_k^l|^m N^{-m/2}
\]

for all \( m \geq 1 \). The proof for \( \mathbb{E} |\hat{\alpha}_{jk}^l - \alpha_{jk}^l|^m \) is similar.

Proposition 2.5 (Proved in [8]). Let

\[
\hat{f} = \sum_{j_0}^{j_1} \sum_k \hat{f}_{jk} \psi_{jk}
\]

where \( \hat{f}_{jk} \) are some random variables. Then

\[
\mathbb{E} \|\hat{f}\|_{p'}^{p'} \leq \begin{cases} C(p') \sum_{j_0}^{j_1} 2^{j(p'-2)/1} \sum_k \mathbb{E} |\hat{f}_{jk}|^{p'}, & 1 \leq p' \leq 2, \\ C(p')S(\beta a) \sum_{j_0}^{j_1} 2^{j(p'-2)/2} \sum_k \mathbb{E} |\hat{f}_{jk}|^{p'}, & p' > 2, \end{cases}
\]

for all \( \beta \) where

\[
a = \frac{p'}{p' - 1}, \quad S(\gamma) = \sum_{j_0}^{j_1} 2^{j(2\gamma)} \leq \begin{cases} c_2 \gamma^{\max j_0 j_1 \gamma}, & \gamma \neq 0, \\ |j_1 - j_0|, & \gamma = 0. \end{cases}
\]

Proof of Theorem 2.1. We follow the idea of the proof presented in the paper [8]. Estimator \([3]\) can be rewritten as \( \tilde{p}_m = \tilde{P}_{j_0} p_m + \tilde{P}_{j_0} \). Thus \( \mathbb{E} \|\tilde{p}_m - p_m\|_{p'}^{p'} \) does not exceed

\[
3^{p'} \left( \mathbb{E} \|\tilde{P}_{j_0} p_m - P_{j_0} p_m\|_{p'}^{p'} + \mathbb{E} \|\tilde{P}_{j_0} \tilde{P}_{j_0} p_m - P_{j_0} p_m\|_{p'}^{p'} + \mathbb{E} \|P_{j_1} p_m - P_{j_0} p_m\|_{p'}^{p'} \right).
\]
Applying Theorem 9.6 of [9] to the third term in (14), we get
\begin{equation}
\|p_m - \Pr_{j_0} p_m\|_{p'} \leq C(s, p, q, M')2^{-j_1 s'p'}.
\end{equation}

Since \(j_1\) is chosen appropriately, the rate of convergence of this term is as claimed in the theorem.

For the first term in (14), we get
\begin{equation}
E \left\| \Pr_{j_0} p_m - \Pr_{j_0} p_m \right\|_{p'}^{p'} \leq C' 2^{j_0 (p'/2 - 1)} \sum_{k \in \mathbb{Z}} E |\tilde{\alpha}_{i,k}^m - \alpha_{i,k}^m|^{p'} \leq C \max_i |a_i^m|^{p'} \left( \frac{2j_0}{N} \right)^{p'/2}
\end{equation}
in view of Propositions 2.2 and 2.4 and since the support of \(\phi\) is compact. Since \(j_0\) is chosen appropriately, the rate of convergence of this term is as claimed in the theorem.

Now we turn to the second term in (14) and define the following sets of indices:
\[\begin{align*}
\hat{B}_j &= \left\{ k : |\hat{\beta}_{j,k}^m| > K \sqrt{\frac{j}{N}} \right\}, \quad \hat{S}_j = \mathbb{Z} \setminus \hat{B}_j, \\
B_j &= \left\{ k : |\beta_{j,k}^m| > K \sqrt{\frac{j}{N}} \right\}, \quad S_j = \mathbb{Z} \setminus B_j, \\
B'_j &= \left\{ k : |\beta_{j,k}^m| > 2K \sqrt{\frac{j}{N}} \right\}, \quad S'_j = \mathbb{Z} \setminus B'_j.
\end{align*}\]

Then \(\Pr_{j_0} p_m - \Pr_{j_0} p_m = e_{ba} + e_{bb} + e_{ab} + e_{ss}\) where
\begin{equation}
e_{ba} = \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} (\hat{\beta}_{j,k}^m - \beta_{j,k}^m) \psi_{j,k} \mathbbm{1}\left\{ k \in \hat{B}_j \cap S_j \right\},
\end{equation}
\begin{equation}
e_{ab} = \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \beta_{j,k}^m \psi_{j,k} \mathbbm{1}\left\{ k \in \hat{S}_j \cap B'_j \right\},
\end{equation}
\begin{equation}
e_{bb} = \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} (\hat{\beta}_{j,k}^m - \beta_{j,k}^m) \psi_{j,k} \mathbbm{1}\left\{ k \in \hat{B}_j \cap B_j \right\},
\end{equation}
\begin{equation}
e_{ss} = \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \beta_{j,k}^m \psi_{j,k} \mathbbm{1}\left\{ k \in \hat{S}_j \cap S'_j \right\}.
\end{equation}

Now we use Proposition 2.5 to estimate the term \(e_{ba}\). Put
\[\hat{f}_{j,k} = |\hat{\beta}_{j,k}^m - \beta_{j,k}^m| \mathbbm{1}\left\{ k \in \hat{B}_j \cap S_j \right\}.
\]
Note that
\[\left\{ k \in \hat{B}_j \cap S_j \right\} \subset D_{jk} \overset{def}{=} \left\{ |\hat{\beta}_{j,k}^m - \beta_{j,k}^m| > (K/2) \sqrt{j/N} \right\}.
\]
The Hölder inequality implies that
\[\sum_{k \in \mathbb{Z}} E |\hat{f}_{j,k}|^{p'} \leq \sum_{k \in \mathbb{Z}} E |\beta_{j,k}^m - \beta_{j,k}^m|^{p'} 1_{D_{jk}} \leq \sum_{k \in \mathbb{Z}} \left( E |\beta_{j,k}^m - \beta_{j,k}^m|^{p'} \right)^{1/r} P(D_{jk})^{1/r'} \leq 2^{-j \gamma'/r'} C \max_i |a_i^m|^{p'} N^{-p'/2} \text{Card}(\Omega_j) \leq 2^{j(1-\gamma'/r')} C \max_i |a_i^m|^{p'} N^{-p'/2}\]
in view of Proposition 2.3 and Lemma 2.1, since the supports of the densities of the components are compact sets. Using Proposition 2.3 and choosing \( \gamma > \beta p'/2 \), we obtain

\[
\mathbb{E} \| e_{sb} \|_{p'}^{p'} \leq C 2^{j_0(p'/2 - \gamma / p')} N^{-p'/2} \max_i |a_i^m|^{p'} N^{-p'/2}.
\]

If \( r' = 3/2 \) and \( \gamma \geq p' \), the rate of convergence of this term is as claimed in the theorem.

To treat the term \( c_{sb} \), we apply Proposition 2.3 to \( \hat{f}_{jk} = \beta_{jk}^m \psi_{jk} \{ k \in \hat{S}_j \cap B_j' \} \). Since \( \{ k \in \hat{S}_j \cap B_j' \} \subset D_{jk} \), we use Lemma 2.1 and Propositions 2.1 and 2.4 and derive the inequality

\[
\sum_k \mathbb{E} |\beta_{jk} m|^{p'} P(D_{jk}) \leq \| \beta_{jk} m \|_{p'}^{p'} 2^{-\gamma j}.
\]

The latter sum is estimated in [8]. It is shown there that \( \mathbb{E} \| e_{sb} \|_{p'}^{p'} \leq C 2^{-j_0(\gamma + \beta p')} \). If \( \gamma \) is sufficiently large, for example, if

\[
\gamma = \gamma_0 = \max \left( \frac{\alpha}{1 - 2\alpha} - s', 1 \right) p',
\]

then the rate of convergence for the term \( e_{sb} \) is as claimed in the theorem.

The term \( e_{bb} \) is estimated by using Lemma 2.2 and applying Proposition 2.3 to \( \hat{f}_{jk} = |\beta_{jk} m - \beta_{jk}^m| \{ k \in \hat{B}_j \cap B_j' \} \):

\[
\sum_k \mathbb{E} |\hat{f}_{jk}|^{p'} \leq \max_i |a_i^m|^{p'} C N^{-p'/2} \sum_k \left| \frac{2\beta_{jk} m}{K} \sqrt{N} j \right|^{p'} \leq C \max_i |a_i^m|^{p'} \| \beta_{jk}^m \|_p^p j^{-p/2} N^{-(p'-p)/2} \leq C \max_i |a_i^m|^{p'} \| p_m \|_{sp, \infty} 2^{-j(s+1/2 - 1/p)p} j^{-p/2} N^{-(p'-p)/2}.
\]

The latter expression is estimated in [8]; thus \( \| e_{bb} \|_{p'}^{p'} \) has the rate of convergence as claimed in the theorem.

The case of mixtures does not require new ideas in the estimation of the term \( e_{sb} \) with \( \hat{f}_{jk} = |\beta_{jk} m - \beta_{jk}^m| \{ k \in \hat{S}_j \cap S'_j \} \) as compared to the case of pure samples, and it follows the lines in [8] where it is proved that the corresponding term is of the required order. We only note that

\[
\sum_{k \in \hat{S}_j \cap S'_j} |\beta_{jk} m|^{p'} \leq (2K)^{p'-p} \left( \frac{j}{N} \right)^{p'-p} \sum_{k \in \hat{B}_j} |\beta_{jk} m|^{p'}.
\]

\section*{Proof of Theorem 2.2}

Recall that \( 2^{j_0} \approx N^{1/(1+2\alpha)} \) and \( 2^{j_1} \approx N / \ln N \). Consider the numbers \( j_i(s, p, q) \) and \( j_i^0(s, p, q) \) defined by

\[
2^{j_0(s, p, q)} \approx \left( N(\ln N)^{-1(\nu > 0)} \right)^{1-2\alpha}, \quad 2^{j_1(s, p, q)} \approx \left( N(\ln N)^{-1(\nu \leq 0)} \right)^{\alpha / p'},
\]

\[
2^{j_i^0(s, p, q)} \approx \left( N(\ln N)^{(p'-p)/p + 1(\nu \geq 0)} \right)^{1-2\alpha}.
\]

Since \( j_i^0(s, p, q) \geq j_0 \) and \( j_1 \geq j_1(s, p, q) \), we have

\[
\mathbb{E} \left\| \hat{P}_{j_0} P_m - P_{j_0} P_m \right\|_{p'}^{p'} \leq C \left( \frac{2^{j_0}}{N} \right)^{p'/2} \leq C \left( \frac{2^{j_0^0(s, p, q)}}{N} \right)^{p'/2},
\]

\[
\mathbb{E} \left\| \hat{P}_{j_1} P_m - P_{j_1} P_m \right\|_{p'}^{p'} \leq C 2^{-j_1 s p'} \leq C 2^{-j_1(s, p, q)s p'}.
\]
We showed in the proof of the preceding theorem that both these terms are of the required order.

Now we split the term \( \mathbb{E} \|[\tilde{P}_{j'}_0]_n - \tilde{P}_{j_0} \|_{p'} \) into the same four parts as in the proof of the preceding theorem. The asymptotic behavior of the terms \( e_{sb} \) and \( e_{bs} \) is estimated in the same way as in the proof of Theorem 2.1. However, now we choose \( K = cp' r_0 \) (that is, \( \gamma \geq \gamma_0 = p' r_0 \)) to get the necessary rate of convergence. The rate of convergence of the term \( e_{ss} = \sum_{j_0} \sum_{k \in \tilde{B}_j \cap B_i} \beta_{jk} \psi_{jk} \) is estimated in [8] for the case of pure samples. The proof for the case of mixtures is the same. The same concerns the term

\[
e_{bb} = \sum_{j_0} \sum_{k \in B_i \cap B_j} (\beta_{jk} - \beta_{jk})
\]

with the exception that an extra factor \( \max_i |a_i|^p + 1 \) appears in view of Lemma 2.2 as compared to the case of pure samples considered in [8].

\[\square\]

3. Convergence of the linear estimator in the uniform metric

Consider the linear estimator

\[
p_t^s = \sum_{k \in \mathbb{Z}} \hat{\alpha}_{jk} \phi_{jk}
\]

where the numbers \( \hat{\alpha}_{jk} \) are defined by [4].

**Theorem 3.1.** Let \( \|e\|_\infty \leq C_\infty \). Then

\[
P \left\{ \sup_x |\tilde{P}_{j'}p_t(x) - \tilde{P}_j p_t(x) | > \lambda_0 \right\} \leq \text{Card}(\Omega_j) 2 \exp \left\{ - \frac{\lambda_0 2^{-j} N^j/\|\theta_\phi\|_{\infty} (N/\alpha)_{2^j} + 4/3 \lambda_0 \|\phi\|_{\infty} \max_i |a_i|^p \} \right\}
\]

for all \( \lambda_0 > 0 \) where \( \theta_\phi \) is defined by [4].

**Theorem 3.2.** Let \( \phi \in C^r, r \in \mathbb{N} \), be an \( f \)-wavelet and let \( \text{supp} \phi \) be a compact set. Then for all \( s, 0 < s < r + 1 \), for all \( p \geq 1 \), and for all \( \varepsilon > 0 \) there exists a constant \( C = C(\varepsilon, s, q, M', T, p) \) such that

\[
\sup_{p_t \in \mathcal{B}(s, q, M', T)} \mathbb{E} \left[ \|\tilde{P}_{j'}p_t - p_t \|_\infty^p \right] \\
\leq C \left( N^{-sjp} \max_i |a_i|^p \left( N^{-(1/2 - \varepsilon)} 2^{j(1/2 + \varepsilon)} \right) \right).
\]

**Lemma 3.1.** If \( f \in C^1(\mathbb{R}) \) is such that \( 0 < \|f'\|_\infty < \infty, \|f\|_\infty < \infty, \) and \( \|f\|_p < \infty \), then

\[
\|f\|_\infty \leq \sqrt{f^2} \|f\|_\infty \|f\|_p \]

for all \( p \geq 1 \).

**Proof.** For all \( \varepsilon \in (0, 1) \), there exists a point \( x_0 \) such that \( |f(x_0)| \geq \varepsilon \|f\|_\infty \). Thus \( |f(x)| \geq \varepsilon \|f\|_\infty /2 \) for all \( x \in (x_0 - \delta, x_0 + \delta) \) where \( \delta = \varepsilon \|f\|_\infty / (2 \|f'\|_\infty) \). Therefore

\[
\|f\|_p^p \geq \int_{x_0 - \delta}^{x_0 + \delta} \left( \frac{\varepsilon \|f\|_\infty}{2} \right)^p dx = \frac{\varepsilon^{p+1} \|f\|_\infty^{p+1}}{2^p \|f'\|_\infty^p},
\]

whence the required inequality follows by passing to the limit as \( \varepsilon \to 1 \). \( \square \)
Proof of Theorem 3.1 Since
\[ \sup_x |\tilde{P}_{rj} p_i - P_{rj} p_i| \leq \sup_x \sum_k |\phi_{jk}(x)||\hat{a}_{jk}^l - \alpha_j^l| \leq 2^{l/2} \max_k |\hat{a}_{jk}^l - \alpha_j^l| \|\theta_{\phi}\|_\infty, \]
we obtain the inclusion
\[ \left\{ \sup_x |\tilde{P}_{rj} p_i - P_{rj} p_i| \geq M' \|\theta_{\phi}\|_\infty 2^{l/2} \right\} \subset \bigcup_{k \in \Omega_j} \left\{ |\hat{a}_{jk}^l - \alpha_j^l| \geq M' \right\}. \]
Thus
\[ (27) \quad P \left\{ \sup_x |\tilde{P}_{rj} p_i - P_{rj} p_i| \geq M' \|\theta_{\phi}\|_\infty 2^{l/2} \right\} \leq \sum_{k \in \Omega_j} P \left\{ |\hat{a}_{jk}^l - \alpha_j^l| \geq M' \right\}. \]
An estimate for the probability \( P \{ |\hat{a}_{jk}^l - \alpha_j^l| \geq M' \} \) follows from the Bernstein inequality. Put \( Y_i = a_j^l(\phi_{jk}(X_i) - \mathbb{E}\phi_{jk}(X_i)) \). It is clear that \( \mathbb{E} Y_j = 0, \sum Y_i \leq C_\infty \langle a^l, a^l \rangle_N \), and \( |Y_i| \leq 2 \cdot 2^{l/2} \max_k |a_k^l| \|\phi\|_\infty \). Then
\[ P \left\{ |\hat{a}_{jk}^l - \alpha_j^l| \geq M' \right\} \leq 2 \exp \left\{ \frac{-NM'^2}{2(C_\infty \langle a^l, a^l \rangle_N + 2 \cdot 2^{l/2} \max_k |a_k^l| \|\phi\|_\infty / 3) \right\}. \]
Using the constant \( M' = \lambda_0 \|\theta_{\phi}\|_\infty^{-1} 2^{-2l/2} \) in inequality (27), we prove bound (25). \( \square \)

Proof of Theorem 3.2 We have
\[ \mathbb{E} \left\| \tilde{P}_{rj} p_i - p_i \right\|_\infty^p \leq C(p) \left( \left\| P_{rj} p_i - p_i \right\|_\infty^p + \mathbb{E} \left\| \tilde{P}_{rj} p_i - P_{rj} p_i \right\|_\infty^p \right). \]
To estimate the first term, we use Theorem 9.3 of [9] and show that the assumptions of the theorem (namely, \( \phi \in C^r, r \in \mathbb{N} \), and supp \( \phi \) is a compact set) imply that \( \|P_{rj} p_i - p_i\|_\infty^p \leq C 2^{-jp^r} \).
Now we estimate the second term. It is shown in [5] that
\[ \left\| \left( \tilde{P}_{rj} p_i - P_{rj} p_i \right) \right\|_\infty^p \leq C 2^{2j} \|a_i\|_N \quad \text{and} \quad \left\| \tilde{P}_{rj} p_i - P_{rj} p_i \right\| \leq C 2^j \|a_i\|_N. \]
We apply Lemma 3.1 and Jensen’s inequality to prove that
\[ \mathbb{E} \left\| \tilde{P}_{rj} p_i - P_{rj} p_i \right\|_\infty^p \leq C 4^{jp^r/(m+1)} \|a_i\|_N^{p/(m+1)} \left( \mathbb{E} \left\| \tilde{P}_{rj} p_i - P_{rj} p_i \right\|_m^p \right)^{p/(m+1)} \]
for all \( m \geq \max(1, (p - 1)) \). It follows from [16] that
\[ \mathbb{E} \left\| \tilde{P}_{rj} p_i - P_{rj} p_i \right\|_m^p \leq C(m) \max_i |a_i|^m \left( \frac{2^j}{N} \right)^{m/2}. \]
Since \( \|a_i\|_N \leq \max_i |a_i|^i \), we get the inequality
\[ \|a_i\|_N^{p/(m+1)} \max_i |a_i|^{p/(m+1)} \leq \max_i |a_i|^{p}. \]
Hence
\[ \mathbb{E} \left\| \tilde{P}_{rj} p_i - P_{rj} p_i \right\|_\infty^p \leq C \max_i |a_i|^p \left( N^{-m/(2(m+1))} 2^{j(2/(m+1)+m/(2(m+1)))} \right)^p. \]
Therefore, given \( \varepsilon > 0 \), there exists a sufficiently large number \( m = m(\varepsilon) \) (thus there exists a number \( C(\varepsilon) = C(\varepsilon) \)) such that
\[ (28) \quad \mathbb{E} \left\| \tilde{P}_{rj} p_i - P_{rj} p_i \right\|_\infty^p \leq C(\varepsilon) \max_i |a_i|^p \left( N^{-(1/2-\varepsilon)} 2^{j(1/2+\varepsilon)} \right)^p. \]
Remark 3.1. Choosing $j$ such that $2^j \simeq N^{(1-2s)/(2s+1+\varepsilon)}$, the order of the right hand side of (26) is $N^{-ps(1-2s)/(1+2s+\varepsilon)}$. If we fix $m = p$ and choose $j$ such that

$$2^j \simeq p/(2s(p+1)+p+4),$$

then the order becomes $N^{-p^2s/(2s(p+1)+p+4)}$.

4. An Adaptive Estimator

Recall that the estimator considered in Theorem 2.1 is not a true adaptive estimator, since it relies on unknown parameters $s$ and $p$ to determine $j_0$ and $j_1$ and $C_\infty$ to evaluate $c$ needed to construct the parameter $K$ involved in the definition of $t_j$. The estimator considered in Theorem 2.2 does not use $s$ and $p$. However $C_\infty$ still is needed. Below we consider an estimator based on a certain statistic $\hat{C}_{\infty}^{j,\eta}$ used instead of $C_\infty$.

Let $\eta > 0$ be a number. Put

$$\hat{C}_{\infty}^{j,\eta} = \max_m \| \hat{P}_m \|_{\infty} + 2\eta,$$

(29) $\hat{K} = \gamma \max \left( \frac{4}{3} \| \psi \| \max_k |a^m_k| \ln 2 \right.

$$+ \sqrt{\left( \frac{4}{3} \| \psi \| \max_k |a^m_k| \ln 2 \right)^2 + \frac{8}{\gamma} \hat{C}_{\infty}^{j,\eta} (a^m, a^m, N, 1)}. $$

Theorem 4.1. Let

$$S \overset{\text{def}}{=} \sup_N \max_{i,l} |d^l_{i,N}| + 1 < \infty,$$

$p' \geq p \geq 1$, $s-1/p > 0$, and let $\phi, \psi \in C^r$, $r > s$, be an $f$- and an $m$-wavelet, respectively, whose supports are compact sets. Let $t_j = \hat{K} \sqrt{\frac{j}{N}}$ where $\hat{K} = \hat{K}(\hat{C}_{\infty}^{j,\eta}, \gamma_0, m)$ is defined by (29).

$$\eta > 0, \quad \gamma_0 = \max \left( \left( \frac{\alpha}{1-2\alpha} - s' \right), 1 \right) p', \quad \hat{C}_{\infty}^{j,\eta} \simeq N^{2/(s-1/p+5)}.$$

Then there are constants $C = C(s, p, q, M', C_\infty)$ and $N_0 > 0$ such that

$$\sup_{p_m} \mathbb{E} \| \hat{p}_m - p_m \|_{p'} \leq C \left( \max_{i=1, \ldots, N} |a^m_i| p' + 1 \right) \left\{ \begin{array}{ll} (\ln N)^{(1-v/s'p)\alpha p'} & \nu > 0, \\
(\ln N)^{\nu/2-\nu/q^*}(\ln N)^{\nu/p_0} & \nu = 0, \\
(\ln N)^{\nu/p_0} & \nu < 0, \end{array} \right. $$

(30)

for all $N > N_0$ if

$$2^{j_0} \simeq \left( N(\ln N)^{(p'/p)^{\frac{\nu}{2}}(\nu \geq 0)} \right)^{1-2\alpha}, \quad 2^{j_1} \simeq \left( N(\ln N)^{-1(\nu \leq 0)} \right)^{\alpha/s'}$$

(31)

where $x^* \overset{\text{def}}{=} \max(x, 0)$, $\hat{p}_m$ is defined by (23), and the supremum is evaluated with respect to all densities $p_m \in B_{s,p}^{\nu}(M', T, C_\infty)$, $m = 1, \ldots, M$.

Theorem 4.2 (Adaptive estimator). Fix an integer number $r_0 > 0$, a real number $\mu > 0$, and define the class

$$S_\mu \overset{\text{def}}{=} \{ (s, p, q) : 1/p + p \leq s \leq r_0, 1 \leq p \leq p', 1 \leq q \leq \infty \}.$$

Let $S \overset{\text{def}}{=} \sup_N \max_{i,l} |d^l_{i,N}| + 1 < \infty$, $p' \geq p \geq 1$, and let $\phi, \psi \in C^{r_0+1}$ be an $f$- and an $m$-wavelet, respectively, whose supports are compact sets. Assume that the densities $p_m$, $m = 1, \ldots, M$, belong to the space $B_{s,p}^{\nu}(M', T)$ where $(s, p, q) \in S_\mu$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Proof. We have
\[
\sup_{p_m} \mathbb{E} \| \tilde{p}_m - p_m \|_p^p \leq C \left( \max_{i=1, \ldots, N} \left| a_i^m \right| + 1 \right) \left\{ \left( \ln N \right)^{p'/2 - p/q} \left( \frac{\ln N}{N} \right)^{\nu' p'}, \quad \nu = 0, \right. \\
\left. \left( \frac{\ln N}{N} \right)^{\nu' p'}, \quad \nu \neq 0, \right\}
\]
for \( 2^{\nu} \approx N^{1/(1+2\nu)}, \) \( 2^{\nu} \approx N/\ln N, \) and for all \( N \geq N_0 \) where \( x^+ \equiv \max(x, 0), \) \( \tilde{p}_m \) is defined by (3), and the supremum is evaluated over all densities \( p_m \in B_{s}'(M', T, C_{\infty}), \) \( m = 1, \ldots, M. \)

Remark 4.1. Note that the estimator considered in Theorem 4.2 has the same rate of convergence as the right hand side of (7). The definition of the estimator considered in Theorem 4.2 does not require the knowledge of the parameters \( s, p, q, \) and \( C_{\infty} \) (this explains the name “adaptive” we use for this estimator); however inequality (22) holds starting with some \( N_0 \) that depends on all of these parameters.

Lemma 4.1. Let the assumptions of Theorem 4.2 hold and let \( s'' = s - 1/p \geq \mu > 0. \) If \( 2^f \approx N^{2/(8\mu+5)} \) (this corresponds to the case of \( \varepsilon = 1/4 \) in Theorem 3.2, then for all \( \gamma > 0 \) and for all \( \eta > 0 \) there exists a number \( N_0 \) such that
\[
P \left\{ \tilde{C}^{\gamma, \eta} < C_{\infty} \right\} \leq N^{-\gamma}
\]
for all \( N > N_0. \)

Proof. We have
\[
P \left\{ \tilde{C}^{\gamma, \eta} < C_{\infty} \right\} \leq \sum_{m=1}^{M} P \left\{ \left\| p_m^* - p_m \right\|_{\infty} \geq 2\eta \right\}
\]
Since \( B_{s''} \subset B_{s'', q} \subset B_{s''}, \) Theorem 3.2 implies that \( \left\| \text{Pr}_j p_m - p_m \right\|_{\infty} \leq C2^{-j\mu} \) and that there exists a number \( N_{01} \) such that
\[
\left\| \text{Pr}_j p_m - p_m \right\|_{\infty} \leq \eta
\]
for all \( m = 1, \ldots, M \) and for all \( N \geq N_{01}. \) Thus
\[
\left\{ \left\| p_m^* - p_m \right\|_{\infty} > 2\eta \right\} \subset \left\{ \left\| \tilde{\text{Pr}}_j p_m - \text{Pr}_j p_m \right\|_{\infty} > \eta \right\}
\]
for all \( N \geq N_{01}. \) By Theorem 3.1, the probability of the latter event does not exceed
\[
C2^{j/2} \left( \frac{\eta^2 2^{-j} N/\left\| \phi \right\|_{\infty} (a_i, a_i^t) \right) \leq C2^{j/2} \left( \frac{\eta^2 2^{-j} N/2C_{\infty} \left\| \phi \right\|_{\infty} + \frac{M}{2} \eta \right) \right) .
\]
Since \( 2^{-j} N \approx N^{1-2/(8\mu+5)}, \) there exists a number \( N_{02} \) such that
\[
P \left\{ \left\| \tilde{\text{Pr}}_j p_m - \text{Pr}_j p_m \right\|_{\infty} > \eta \right\} \leq N^{-\gamma}/M
\]
for all \( N \geq N_{02}. \)

Proof of Theorem 4.1. We consider the changes in the proof caused by the substitution of \( \tilde{C}_{\infty} \) for \( C_{\infty} \) in the definition of the estimator. Note that \( \{ \tilde{K} < K \} \subset \{ \tilde{C}_{\infty} < C_{\infty} \}. \) It is clear that the substitution mentioned above does not influence the reasoning concerning...
the terms $\mathbb{E}[|\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}|^{p}]$ and $\mathbb{E}[\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}]$, although the behavior of the term $\mathbb{E}[|\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}|^{p}]$ must be studied separately. As before we define the sets

$$
\hat{B}_j = \left\{ k : |\hat{\beta}_{jk}^{m}| > \tilde{K} \sqrt{\frac{j}{N}} \right\}, \quad \hat{S}_j = \mathbb{Z} \setminus \hat{B}_j,
$$

$$
B_j = \left\{ k : |\beta_{jk}^{m}| > \tilde{K} / 2 \sqrt{\frac{j}{N}} \right\}, \quad S_j = \mathbb{Z} \setminus B_j,
$$

$$
B_j' = \left\{ k : |\beta_{jk}^{m}| > 2 \tilde{K} \sqrt{\frac{j}{N}} \right\}, \quad S_j' = \mathbb{Z} \setminus B_j'.
$$

We also define the terms $e_{bs}$, $e_{bb}$, $e_{sx}$, and $e_{sb}$ according to obvious analogs of equalities (17)–(18). Put

$$
\hat{D}_{jk} = \left\{ |\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}| > \tilde{K} / 2 \sqrt{\frac{j}{N}} \right\}, \quad D_{jk} = \left\{ |\beta_{jk}^{m} - \beta_{jk}^{m}| > K / 2 \sqrt{\frac{j}{N}} \right\}.
$$

We start with the term $e_{bs}$. As in the proof of the preceding theorem,

$$
\{ k \in \hat{B}_j \cap S_j \} \subset \hat{D}_{jk}.
$$

Thus

$$
\sum_{k \in \mathbb{Z}} \mathbb{E}[|\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}|^{p}] \mathbb{I} \{ k \in \hat{B}_j \cap S_j \} \leq \sum_{k \in \mathbb{Z}} \mathbb{E}[|\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}|^{p}] \mathbb{I} \hat{D}_{jk} = \sum_{k \in \mathbb{Z}} \mathbb{E}[|\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}|^{p}] \mathbb{I} \hat{D}_{jk} \mathbb{I} \{ \hat{C}_{\infty} \geq C_{\infty} \} + \sum_{k \in \mathbb{Z}} \mathbb{E}[|\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}|^{p}] \mathbb{I} \hat{D}_{jk} \mathbb{I} \{ \hat{C}_{\infty} < C_{\infty} \}
$$

$$
\leq \sum_{k \in \mathbb{Z}} \mathbb{E}[|\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}|^{p}] \mathbb{I} \hat{D}_{jk} + \sum_{k \in \mathbb{Z}} \mathbb{E}[|\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}|^{p}] \mathbb{I} \{ \hat{C}_{\infty} < C_{\infty} \}.
$$

The order of this term is as required if $\mathbb{P} \{ \hat{C}_{\infty} < C_{\infty} \} \leq N^{-m}$.

To estimate the term $e_{bs}$, we consider the sum $\sum_{k} |\beta_{jk}^{m}|^{p} \mathbb{P}(\hat{D}_{jk})$. Taking into account the inclusion $\hat{D}_{jk} \subset D_{jk} \cup \{ \hat{C}_{\infty} < C_{\infty} \}$, we prove that the order of this term is as required if $\mathbb{P} \{ \hat{C}_{\infty} < C_{\infty} \} \leq N^{-m}$.

We apply the inequality

$$
\sum_{k \in \mathbb{Z}} (\hat{\beta}_{jk}^{m} - \beta_{jk}^{m}) \psi_{jk} \mathbb{I} \{ k \in \hat{B}_j \cap B_j \} \leq C^{-p/2} \sum_{k \in \mathbb{Z}} \frac{2 \beta_{jk}^{m}}{K} \sqrt{\frac{N}{j}}
$$

to estimate the term $e_{bb}$. Since $\tilde{K} \geq 1$, the order of this term coincides with that of the corresponding term in Theorem 4.4.

The substitution of $\tilde{K}$ for $K$ does not change the order of the term $e_{sx}$, since Theorem 4.4 implies that for all $b > 0$ there exists a constant $C = C(b)$ such that

$$
\mathbb{E}[|C_{\infty}^{b} - \eta|^{b}] \leq C(b) ( (C_{\infty} + \eta)^{b} + N^{-b} \frac{(s-1/p)^2}{2(s-1/p)+1/p} ).
$$

Therefore $\mathbb{E}[|\hat{K}|^{t}] < \infty$ for all $t > 0$. To get the inequality in the case of $b \in (0, 1)$, one uses the Lyapunov inequality.

**Proof of Theorem 4.4** The proof follows the lines of that of Theorem 4.2 with the substitution of $\tilde{K}$ for $K$. In doing so, we apply the method used in the proof of the preceding theorem.
Figure 1. Averaged $\|\cdot\|_2^2$ error of the threshold-based estimator and linear estimators at various levels. Logarithmic scale.

5. Results of simulation

Consider a mixture of two components. Assume that the density of the first component is of the form $p_1(x) = 8f(8x - 0.075)$ where $f(x) = 16\int_0^x g(y)\,dy$ and

$$g(y) \overset{\text{def}}{=} 2y\mathbb{1}\{y \in [0, 0.25]\} + (1 - 2y)\mathbb{1}\{y \in [0.25, 0.75]\} + (2y - 1)\mathbb{1}\{y \in [0.75, 1]\}.$$

Note that this density has the Lipschitz derivative. The density of the second component coincides with the shifted first density, namely

$$p_2(x) = p_1(x - 1/16).$$

We also assume that $w_{j_0}^{1,N} \overset{\text{def}}{=} j/N$ and $w_{j_1}^{1,N} = 1 - j/N$. Indicated with diamonds, stars, and squares in Figure 1 are the results of simulation for linear estimators at levels 3, 5, and 8, respectively. Indicated with triangles in Figure 1 are the results for the threshold-based estimator at levels $j_0 = 3$ and $j_1 = 7$ (that correspond to the linear estimator at level 3 with all rejected $\hat{\beta}_{jk}$ and that at level 8 with all remaining $\hat{\beta}_{jk}$, respectively).

The simulation is based on the wavelet basis $D_{12}$ (see [9]). The sizes of the sample are indicated on the $x$-axis in Figure 1. The logarithm of $\|\hat{p}_1 - p_1\|_2^2$ averaged over 100 experiments is indicated on the $y$-axis.

The density $p_1$ is approximated by the wavelet expansions in the basis $D_{12}$ with the accuracy

$$\|p_1 - \text{Pr}_3 p_1\|_2^2 = 4.450, \quad \|p_1 - \text{Pr}_4 p_1\|_2^2 = 2.183, \quad \|p_1 - \text{Pr}_5 p_1\|_2^2 = 0.006,$$

$$\|p_1 - \text{Pr}_6 p_1\|_2^2 \leq 10^{-3}.$$

Thus the error of the linear estimator at level 3 is almost completely explained by the nonrandom component $\|p_1 - \text{Pr}_3 p_1\|_2^2$ (in other words, by the error of approximation).
This is not the case for the linear estimator at level 8; namely the error of the approximation is almost zero but the variance of the random component $E\|\hat{Pr}_8p_1 - Pr_8p_1\|^2$ is large enough.

The best results are obtained for the linear estimator at level 5: the error of approximation is small and the variance is not large. However the best level is not known in actual applied problems even if the parameters $s, p, a, n, d, r, q$ and $κ$ are known. This problem can be partially solved with the help of the adaptive estimator. Its behavior is as follows in the present example: if the size $N$ of the sample is small, then $t_j = K\sqrt{j/N}$ are big, whence $β_jk = 0$. If $N$ grows, then $t_j$ decreases and the estimator starts to involve estimators $β_jk$ for which $β_jk$ have big absolute values. This decreases essentially the error of approximation ($N \in [1000, 8000]$ in Figure 1). Therefore the $β_jk$ with large absolute values remain in the estimator and this decreases the error of approximation and does not increase essentially the variance (the threshold-based estimator becomes better than the linear estimator at level 8 for $N \geq 4000$).

The results for the threshold-based estimator for $N \in [10000, 15000]$ are not much worse than those of the linear estimator at level 5: the error of the threshold-based estimator is $\|\hat{p}_1 - p_1\|^2 \approx 0.027$, while the error for the linear estimator at level 5 is $\|p^*_1 - p_1\|^2 \approx 0.016$.

The above consideration is summed up as follows. If the size of a sample is large, then the threshold-based estimator shows better results than the linear estimators at the corresponding levels. It is worthwhile mentioning that these results are not worse essentially than those for the linear estimator at the optimal level which is not known a priori.

6. CONCLUDING REMARKS

The rate of convergence of the linear estimator in the uniform metric is considered in this paper. A nonlinear nonadaptive estimator of the density of components of a mixture is constructed and its rate of convergence is obtained in the integral metric. Moreover we constructed a statistic that can be used in place of the unknown parameter in the definition of the adaptive estimator. We show that this substitution does not decrease the rate of convergence. The simulated data shows that the properties of the threshold-based estimator are better than those of the linear estimator in some cases.

BIBLIOGRAPHY


**Department of Probability Theory and Mathematical Statistics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Volodymyrs'ka Street, 64, Kyiv 01033, Ukraine**

*E-mail address*: pdi2004@mail.ru

Received 27/JUN/2005

Translated by S. KVASKO