HAL
open science

Block thresholding for a density estimation problem with a change-point

Christophe Chesneau

To cite this version:

Christophe Chesneau. Block thresholding for a density estimation problem with a change-point. 2008. hal-00268393

HAL Id: hal-00268393

https://hal.science/hal-00268393v1

Preprint submitted on 1 Apr 2008

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

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

Block thresholding for a density estimation problem with a change-point

Christophe Chesneau

Laboratoire de Mathématiques Nicolas Oresme,
Université de Caen Basse-Normandie,
Campus II, Science 3,
14032 Caen, France.
http://www.chesneau-stat.com *chesneau@math.unicaen.fr*

Abstract

We consider a density estimation problem with a change-point. We develop an adaptive wavelet estimator constructed from a block thresholding rule. Adopting the minimax point of view under the $\mathbb{L}^p$ risk (with $p \geq 1$) over Besov balls, we prove that it is near optimal.

Key words: Density estimation, Wavelets, Block thresholding.

1991 MSC: 62G07

1 MOTIVATIONS

The standard density estimation problem can be formulated as follows. Let $n \in \mathbb{N}^*$ and $(X_i)_{i\in\{1,...,n\}}$ be an i.i.d. sample from a distribution with density function f. The goal is to estimate the density function f based on the sample. In the statistical literature, various estimation techniques have been studied. We refer to the books of Devroye and Györfi [5], Silverman [13], Efromovich [6], Härdle et al. [9] and Tsybakov [14].

In this paper, we consider a different density estimation problem inspired from models for control charts (type Shewhart). Let $n \in \mathbb{N}^*$ and $T_n \in \mathbb{N}^*$. Let $(X_{i,r})_{(i,r)\in\{1,...,n\}\times\{1,...,T_n\}}$ be independent random variables. For each $r \in \{1, ..., T_n\}$, $(X_{i,r})_{i\in\{1,...,n\}}$ is an i.i.d. sample from a distribution with density function f_r. We suppose that there exists $\eta_n \in \{1, ..., T_n - 1\}$ such that

- for any $r \in \{1, ..., \eta_n\}$, we have $f_r = f$,

- for any $r \in \{\eta_n + 1, ..., T_n\}$, we have $f_r = h \neq f$ and $\mathbb{E}(X_{1,\eta_n}) \neq \mathbb{E}(X_{1,\eta_n+1})$.

The integer η_n and the function f are unknown. The goal is to estimate f from $(X_{i,r})_{(i,r)\in\{1,...,n\}\times\{1,...,T_n\}}$.

We make the two following assumptions.

($H1$) We assume that $\lim_{n\to\infty} n^{-1} \log T_n = 0$. Thus T_n and, a fortiori, η_n, can be really greater than n.

($H2$) We assume that, for any $r \in \{1, ..., T_n\}$, $X_{1,r}(\Omega) = [0, 1]$ and that there exists a known constant $K > 0$ such that $\sup_{x\in[0,1]} |f(x)| \leq K < \infty$.

To estimate f, one can only use the variables $(X_{i,1})_{i\in\{1,...,n\}}$ and take a standard density estimator (kernel, wavelets, ...). However, if we consider all the variables $(X_{i,r})_{(i,r)\in\{1,...,n\}\times\{1,...,\eta_n\}}$, we gain informations on f. Its estimation can be significantly improved. This motivates the construction of a plug-in estimator of f described as follows. Firstly, we estimate η_n via $(X_{i,r})_{(i,r)\in\{1,...,n\}\times\{1,...,T_n\}}$. Let $\widehat{\eta}_n$ be the corresponding estimator. Then, we estimate f by a density estimator $\widetilde{f}_n$ constructed from $(X_{i,r})_{(i,r)\in\{1,...,n\}\times\{1,...,\widehat{\eta}_n\}}$. In this study, we adopt the wavelet methodology. The considered estimator uses a $\mathbb{L}^p$ version of the local block thresholding rule known under the name of BlockShrink. It has been initially developed for the standard density estimation under $\mathbb{L}^2$ risk by Hall et al. [8, 7] and recently improved by Cai and Chicken [1]. The $\mathbb{L}^p$ version of this thresholding rule, more general, has been introduced by Picard and Tribouley [12].

To measure the performances of $\widetilde{f}_n$, we consider the minimax approach under the $\mathbb{L}^p$ risk with $p \geq 1$ (not only $p = 2$) over wide range of smoothness spaces: the Besov balls. We aim to evaluate the smallest bound w_n such that

$$\sup_{f\in B^s_{\pi,r}(M)} \mathbb{E}\left(\|\widetilde{f}_n - f\|_p^p\right) \leq w_n,$$

where, for any $m \in \mathbb{L}^p([0, 1])$, $\|m\|_p^p = \int_0^1 |m(x)|^p dx$ and $B^s_{\pi,r}(M)$ is the Besov ball (to be defined in Section 2). In this study, we prove that w_n is of the form

$$w_n = C(n\eta_n)^{-\alpha}(\log(n\eta_n))^\beta,$$

where C is a constant independent of f and n, $\alpha \in]0, 1[$, $\beta \in [0, 1[$ and α, β only depend on s, π, r, s and p. It is near optimal in the minimax sense. The proof is based on several auxiliary results including one proved by Chesneau [3]. The originality of w_n resides in the presence of η_n in its expression: more η_n is large, more w_n is small. This illustrates the fact that our estimator takes into account all the pertinent observations for the estimation of f.

The rest of the paper is organized as follows. In Section 2, we present wavelets

and Besov balls. The estimators are defined in Section 3. Section 4 is devoted to the main result. The proofs are postponed in Section 5.

2 WAVELETS AND BESOV BALLS

We consider an orthonormal wavelet basis generated by dilations and translations of a compactly supported "father" wavelet ϕ and a compactly supported "mother" wavelet ψ. For the purposes of this paper, we use the periodized wavelet bases on the unit interval. For any $x \in [0,1]$, any integer j and any $k \in \{0, \ldots, 2^j - 1\}$, let $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ be the elements of the wavelet basis and

$$\phi_{j,k}^{per}(x) = \sum_{l \in \mathbb{Z}} \phi_{j,k}(x - l), \qquad \psi_{j,k}^{per}(x) = \sum_{l \in \mathbb{Z}} \psi_{j,k}(x - l),$$

their periodized version. There exists an integer τ such that the collection ζ defined by $\zeta = \{\phi_{\tau,k}^{per}, k = 0, ..., 2^\tau - 1;\ \psi_{j,k}^{per}, \quad j = \tau, ..., \infty,\ k = 0, ..., 2^j - 1\}$ constitutes an orthonormal basis of $\mathbb{L}^2([0,1])$. In what follows, the superscript "*per*" will be suppressed from the notations for convenience. For any integer $l \geq \tau$, a function $f \in \mathbb{L}^2([0,1])$ can be expanded into a wavelet series as

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

where $\alpha_{l,k} = \int_0^1 f(t)\phi_{l,k}(t)dt$ and $\beta_{j,k} = \int_0^1 f(t)\psi_{j,k}(t)dt$. For further details about wavelet bases on the unit interval, we refer to Cohen et al. [4].

Let us now define the Besov balls. Let $M \in (0, \infty)$, $s \in (0, \infty)$, $\pi \in [1, \infty]$ and $r \in [1, \infty]$. Let us set $\beta_{\tau-1,k} = \alpha_{\tau,k}$. We say that a function f belongs to the Besov balls $B_{\pi,r}^s(M)$ if and only if there exists a constant $M^* > 0$ such that the associated wavelet coefficients satisfy

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

For a particular choice of parameters s, π and r, these sets contain the Hölder and Sobolev balls. See Meyer [10].

3 ESTIMATOR

For any $\kappa > 0$, set

$$\mathcal{A}_n(\kappa) = \left\{ r \in \{1, ..., T_n\}; \quad \left| \sum_{i=1}^{n} (X_{i,r} - X_{i,r+1}) \right| \geq \kappa \sqrt{n \log(nT_n)} \right\}. \tag{3.1}$$

We estimate η_n by the random integer

$$\widehat{\eta}_n(\kappa) = \min \mathcal{A}_n(\kappa). \tag{3.2}$$

This estimator satisfies $\mathbb{P}\left(\widehat{\eta}_n(\kappa) = \eta_n\right) \geq 1 - 2\eta_n(nT_n)^{-h(\kappa)}$, where $h(\kappa) = \kappa^2/(32 + 3^{-1}8\kappa)$. See Proposition 4.1 below. A suitable value for κ will be specified later.

We are now in the position to describe the considered estimator of f. As mentioned in Section 1, it can be viewed as a generalization of the $\mathbb{L}^p$ version of the "BlockShrink estimator" initially developed under $\mathbb{L}^2$ risk by Hall et al. [8, 7] and Cai and Chicken [1]. For its $\mathbb{L}^p$ form, see Picard and Tribouley [12].

Let $p \geq 1$ and $u \in \{1, ..., T_n\}$. Let $j_1(u)$ and $j_2(u)$ be the integers defined by $j_1(u) = \lfloor ((p \vee 2)/2) \log_2(\log(nu)) \rfloor$ and $j_2(u) = \lfloor \log_2(nu/\log(nu)) \rfloor$. Here, $p \vee 2 = \max(p, 2)$ and the quantity $\lfloor a \rfloor$ denotes the whole number part of a. For any $j \in \{j_1(u), ..., j_2(u)\}$, set $L(u) = \lfloor (\log(nu))^{(p \vee 2)/2} \rfloor$ and $A_j(u) = \{1, ..., \lfloor 2^j L(u)^{-1} \rfloor\}$. For any $K \in A_j(u)$, we consider the set

$$U_{j,K}(u) = \{k \in \{0, ..., 2^j - 1\}; \, (K-1)L(u) \leq k \leq KL(u) - 1\}.$$

For any $u \in \{1, ..., T_n\}$, we define $\widehat{f}_n(x; u)$, $x \in [0, 1]$, by

$$\begin{aligned} \widehat{f}_n(x; u) = & \sum_{k=0}^{2^{j_1(u)}-1} \widehat{\alpha}_{j_1(u),k}(u) \phi_{j_1(u),k}(x) \\ & + \sum_{j=j_1(u)}^{j_2(u)} \sum_{K \in A_j(u)} \sum_{k \in U_{j,K}(u)} \widehat{\beta}_{j,k}(u) 1_{\left\{\widehat{b}_{j,K}(u) \geq dn^{-1/2}\right\}} \psi_{j,k}(x), \end{aligned} \tag{3.3}$$

where d is a constant independent of f and n, $\widehat{b}_{j,K}(u) = (L^{-1} \sum_{k \in U_{j,K}(u)} |\widehat{\beta}_{j,k}(u)|^p)^{1/p}$,

$$\widehat{\alpha}_{j_1(u),k}(u) = (nu)^{-1} \sum_{i=1}^{n} \sum_{r=1}^{u} \phi_{j_1(u),k}(X_{i,r}), \qquad \widehat{\beta}_{j,k}(u) = (nu)^{-1} \sum_{i=1}^{n} \sum_{r=1}^{u} \psi_{j,k}(X_{i,r}). \tag{3.4}$$

We finally consider the estimator

$$\widetilde{f}_n(x) = \widehat{f}_n\left(x; \widehat{\eta}_n(\kappa_*)\right), \tag{3.5}$$

where $\widehat{\eta}_n(\kappa_*)$ is defined by (3.2), $\kappa_* = t^{-1}(2p+1)$ and t is the function defined by $t(x) = x^2/(8 + 3^{-1}4x)$, $x \in [0, \infty[$.

Let us mention that $\widetilde{f}_n$ does not require any a priori knowledge on f and η_n in its construction. It is adaptive.

Remark. For any $u \in \{1, ..., T_n\}$, the sets $A_j(u)$ and $U_{j,K}(u)$ are chosen such that $\cup_{K \in A_j(u)} U_{j,K}(u) = \{0, ..., 2^j - 1\}$, $U_{j,K}(u) \cap U_{j,K'}(u) = \emptyset$ for any $K \neq K'$ with K, $K' \in A_j(u)$, and $|U_{j,K}(u)| = L(u)$.

4 RESULTS

4.1 MAIN RESULT

Theorem 4.1 below determines the rates of convergence achieved by the estimator $\widetilde{f}_n$ under the $\mathbb{L}^p$ risk (with $p \geq 1$) over Besov balls.

Theorem 4.1 *Consider the density model described in Section* 1 *and the estimator* $\widetilde{f}_n$ *defined by (3.5). Then there exists a constant* $C > 0$ *such that, for any* $\pi \in [1, \infty]$, $r \in [1, \infty]$, $s \in (1/\pi, \infty)$, *and* d *and* n *large enough, we have*

$$\sup_{f \in B^s_{\pi,r}(M)} \mathbb{E}\left(\|\widetilde{f}_n - f\|_p^p\right) \leq C\omega_n,$$

where

$$\omega_n = \begin{cases} (n\eta_n)^{-\alpha_1 p}(\log(n\eta_n))^{\alpha_1 p 1_{\{p>\pi\}}}, & \textit{when} \quad \epsilon > 0, \\ ((n\eta_n)^{-1}\log(n\eta_n))^{\alpha_2 p}(\log(n\eta_n))^{(p-\pi/r)+1_{\{\epsilon=0\}}}, & \textit{when} \quad \epsilon \leq 0, \end{cases}$$

with $\alpha_1 = s/(2s+1)$, $\alpha_2 = (s - 1/\pi + 1/p)/(2(s - 1/\pi) + 1)$ *and* $\epsilon = \pi s + 2^{-1}(\pi - p)$.

Now, let us briefly discuss the optimal nature of ω_n. Using standard lower bound techniques, we can prove that there exists a constant $c > 0$ such that

$$\inf_{\widetilde{f}} \sup_{f \in B^s_{\pi,r}(M)} \mathbb{E}\left(\|\widetilde{f} - f\|_p^p\right) \geq cv_n,$$

where $\inf_{\widetilde{f}}$ denotes the infimum over all the possible estimators $\widetilde{f}$ of f and $v_n = (n\eta_n)^{-\alpha_1 p}$ when $\epsilon > 0$ and $v_n = ((n\eta_n)^{-1}\log(n\eta_n))^{\alpha_2 p}$ when $\epsilon \leq 0$. The proof is similar to the proof of the lower bound for the standard density estimation problem with $n\eta_n$ i.i.d. variables. Further details can be found in the book of Härdle et al. [9] (Section 10.4). Since ω_n is equal to v_n up to logarithmic terms, it is near optimal in the minimax sense. Moreover, one can show that it is better than those achieved by the conventional term-by-term thresholding estimators (hard, soft,...). The main difference is for the case $\{\pi \geq p\}$ where there is no extra logarithmic term.

The proof of Theorem 4.1 is based on several auxiliary results.

Statistical properties satisfied by the set $\mathcal{A}_n(\kappa)$ and the estimator $\widehat{\eta}_n(\kappa)$ are presented in Proposition 4.1 below.

Proposition 4.1 *Let $\kappa > 0$, $\mathcal{A}_n(\kappa)$ be defined by (3.1) and $\widehat{\eta}_n(\kappa)$ be defined by (3.2). Then*

- *we have*
$$\mathbb{P}\left(\eta_n \in \mathcal{A}_n(\kappa)\right) \geq 1 - 2(nT_n)^{-h(\kappa)},$$
where $h(\kappa) = \kappa^2/(32 + 3^{-1}8\kappa)$.
- *for any $m \in \{1, ..., T_n\} - \{\eta_n\}$, we have*
$$\mathbb{P}\left(m \in \mathcal{A}_n(\kappa)\right) \leq 2(nT_n)^{-t(\kappa)},$$
where $t(\kappa) = \kappa^2/(8 + 3^{-1}4\kappa)$.
- *we have*
$$\mathbb{P}\left(\widehat{\eta}_n(\kappa) = \eta_n\right) \geq 1 - 2\eta_n(nT_n)^{-h(\kappa)},$$
where $h(\kappa) = \kappa^2/(32 + 3^{-1}8\kappa)$.

Proposition 4.2 below investigates the performances of the non adaptive estimator $\widehat{f}_n(x;\eta_n)$ via the minimax approach under the $\mathbb{L}^p$ risk (with $p \geq 1$) over Besov balls.

Proposition 4.2 *Consider the density model described in Section 1 and the estimator $\widehat{f}_n(x;\eta_n)$ defined by (3.3). Then there exists a constant $C > 0$ such that, for any $\pi \in [1, \infty]$, $r \in [1, \infty]$, $s \in (1/\pi, \infty)$, and d and n large enough, we have*
$$\sup_{f \in B^s_{\pi,r}(M)} \mathbb{E}\left(\|\widehat{f}_n(.;\eta_n) - f\|_p^p\right) \leq C\omega_n,$$
where ω_n is defined as in Theorem 4.1.

By definition of η_n, the variables $(X_{i,r})_{(i,r)\in\{1,...,n\}\times\{1,...\eta_n\}}$ are i.i.d. with probability density function f. By $(H2)$, we have $\sup_{x\in[0,1]} |f(x)| \leq K < \infty$. Therefore, the proof of Proposition 4.2 is similar to those of Chesneau [2, Theorem 4.1] with $n\eta_n$ i.i.d. variables (and the weight function $w(x) = 1$). It is a consequence of a general result proved by Chesneau [3, Theorem 4.2]. The crucial points are some statistical properties satisfied by the estimators $\widehat{\alpha}_{j,k}(u)$ and $\widehat{\beta}_{j,k}(u)$ defined by (3.4) (moment inequality and concentration inequality). For the particular case $p = 2$, we refer to Cai and Chicken [1, Theorem 1].

5 PROOFS

In this section, C denotes a positive constant which can take different values for each mathematical term. It is independent of f and n.

Proof of Theorem 4.1. We have

$$\mathbb{E}\left(\|\widetilde{f}_n - f\|_p^p\right) = \sum_{u=1}^{T_n} \mathbb{E}\left(\|\widehat{f}_n(.;u) - f\|_p^p 1_{\{\widehat{\eta}_n(\kappa_*)=u\}}\right) = A + B, \tag{5.1}$$

where

$$A = \mathbb{E}\left(\|\widehat{f}_n(.;\eta_n) - f\|_p^p 1_{\{\widehat{\eta}_n(\kappa_*)=\eta_n\}}\right)$$

and

$$B = \sum_{u\in\{1,\ldots,T_n\}-\{\eta_n\}} \mathbb{E}\left(\|\widehat{f}_n(.;u) - f\|_p^p 1_{\{\widehat{\eta}_n(\kappa_*)=u\}}\right).$$

The upper bound for A. It follows from Proposition 4.2 that, if $f \in B^s_{\pi,r}(M)$,

$$A \leq \mathbb{E}\left(\|\widehat{f}_n(.;\eta_n) - f\|_p^p\right) \leq C\omega_n, \tag{5.2}$$

where ω_n is the desired rate of convergence.

The upper bound for B. By $(H2)$, we have $\sup_{x\in[0,1]} |f(x)| \leq K < \infty$. Using the elementary inequalities : $\|\widehat{f}_n(.;u) - f\|_p^p \leq (\|\widehat{f}_n(.;u)\|_p + \|f\|_p)^p \leq 2^{p-1}(\|\widehat{f}_n(.;u)\|_p^p + \|f\|_p^p) \leq 2^{p-1}(\|\widehat{f}_n(.;u)\|_p^p + K^p)$, we obtain

$$B \leq 2^{p-1} \sum_{u\in\{1,\ldots,T_n\}-\{\eta_n\}} \mathbb{E}\left(\left(\|\widehat{f}_n(.;u)\|_p^p + K^p\right) 1_{\{\widehat{\eta}_n(\kappa_*)=u\}}\right). \tag{5.3}$$

Since $\|\phi_{j,k}\|_p = \|\phi\|_p 2^{j(1/2-1/p)}$ and $\|\psi_{j,k}\|_p = \|\psi\|_p 2^{j(1/2-1/p)}$, the Minkowski inequality yields

$$\begin{aligned}
\|\widehat{f}_n(.;u)\|_p^p &\leq \Bigg(\sum_{k=0}^{2^{j_1(u)}-1} \left|\widehat{\alpha}_{j_1(u),k}(u)\right| \|\phi_{j_1(u),k}\|_p \\
&+ \sum_{j=j_1(u)}^{j_2(u)} \sum_{K\in A_j(u)} \sum_{k\in U_{j,K}(u)} \left|\widehat{\beta}_{j,k}(u)\right| 1_{\{\widehat{b}_{j,K}(u)\geq dn^{-1/2}\}} \|\psi_{j,k}\|_p \Bigg)^p \\
&\leq C\Bigg(\sum_{k=0}^{2^{j_1(T_n)}-1} \left|\widehat{\alpha}_{j_1(u),k}(u)\right| 2^{j_1(T_n)(1/2-1/p)} + \sum_{j=j_1(1)}^{j_2(T_n)} \sum_{k=0}^{2^j-1} \left|\widehat{\beta}_{j,k}(u)\right| 2^{j(1/2-1/p)} \Bigg)^p.
\end{aligned}$$

Since the wavelet basis is compactly supported, we have $\sup_{x\in[0,1]} |\phi_{j,k}(x)| \leq$

$\|\phi\|_\infty 2^{j/2}$ and $\sup_{x\in[0,1]}|\psi_{j,k}(x)| \leq \|\psi\|_\infty 2^{j/2}$. It follows that $\left|\widehat{\alpha}_{j_1(u),k}(u)\right| \leq (nu)^{-1}\sum_{i=1}^n\sum_{r=1}^u \left|\phi_{j_1(u),k}(X_{i,r})\right| \leq \sup_{x\in[0,1]}|\phi_{j_1(u),k}(x)| \leq C2^{j_1(u)/2} \leq C2^{j_1(T_n)/2}$ and $\left|\widehat{\beta}_{j,k}(u)\right| \leq (nu)^{-1}\sum_{i=1}^n\sum_{r=1}^u |\psi_{j,k}(X_{i,r})| \leq \sup_{x\in[0,1]}|\psi_{j,k}(x)| \leq C2^{j/2}$. This yields

$$\begin{aligned}\|\widehat{f}_n(.;u)\|_p^p &\leq C\left(2^{j_1(T_n)(2-1/p)} + \sum_{j=j_1(1)}^{j_2(T_n)} 2^{j(2-1/p)}\right)^p \leq C2^{j_2(T_n)(2p-1)} \\ &\leq C(nT_n)^{2p-1}. \end{aligned} \tag{5.4}$$

Combining (5.3) and (5.4), and using the fact that, for any $u \in \{1, ..., T_n\}$, we have $\{\widehat{\eta}_n(\kappa_*) = u\} \subseteq \{u \in \mathcal{A}_n(\kappa_*)\}$, we obtain

$$\begin{aligned} B &\leq C(nT_n)^{2p-1} \sum_{u\in\{1,...,T_n\}-\{\eta_n\}} \mathbb{P}(\widehat{\eta}_n(\kappa_*) = u) \\ &\leq C(nT_n)^{2p} \sup_{u\in\{1,...,T_n\}-\{\eta_n\}} \mathbb{P}(u \in \mathcal{A}_n(\kappa_*)). \end{aligned}$$

The second point of Proposition 4.1 and the definition of κ_* give

$$B \leq C(nT_n)^{2p}(nT_n)^{-t(\kappa_*)} \leq C(nT_n)^{2p-t(\kappa_*)} \leq C(nT_n)^{-1} \leq C\omega_n, \tag{5.5}$$

where ω_n is the desired rate of convergence. Putting the inequalities (5.1), (5.2) and (5.5) together, this completes the proof of Theorem 4.1.

□

Proof of Proposition 4.1. First of all, let us present the Bernstein inequality. See, for instance, Petrov [11].

Lemma 5.1 (Bernstein's inequality) *Let $(Y_i)_{i\in\mathbb{N}^*}$ be independent random variables such that, for any $n \in \mathbb{N}^*$ and any $i \in \{1, ..., n\}$, we have $\mathbb{E}(Y_i) = 0$ and $|Y_i| \leq M < \infty$. Then, for any $\lambda > 0$, and any $n \in \mathbb{N}^*$, we have*

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

where $d_n^2 = \sum_{i=1}^n \mathbb{E}(Y_i^2)$.

- *Proof of the first point.* We have

$$\mathbb{P}\left(\eta_n \in \mathcal{A}_n(\kappa)\right) = \mathbb{P}\left(\left|\sum_{i=1}^n (X_{i,\eta_n} - X_{i,\eta_n+1})\right| \geq \kappa\sqrt{n\log(nT_n)}\right).$$

Set $\mu_n = \mathbb{E}(X_{1,\eta_n}) - \mathbb{E}(X_{1,\eta_n+1})$. By definition of η_n, we have $\mathbb{E}(X_{1,\eta_n}) \neq \mathbb{E}(X_{1,\eta_n+1})$. By $(H1)$, we have $\lim_{n\to\infty} n^{-1}\log T_n = 0$. Therefore, for a large enough n, we have $|\mu_n| \geq 2\kappa\sqrt{n^{-1}\log(nT_n)}$. Using the triangular inequality, we obtain the inclusion $\left\{|\sum_{i=1}^n (X_{i,\eta_n} - X_{i,\eta_n+1} - \mu_n)| \leq \kappa\sqrt{n\log(nT_n)}\right\} \subseteq \left\{|\sum_{i=1}^n (X_{i,\eta_n} - X_{i,\eta_n+1})| \geq \kappa\sqrt{n\log(nT_n)}\right\}$. It follows that

$$\begin{aligned}\mathbb{P}\left(\eta_n \in \mathcal{A}_n(\kappa)\right) &\geq \mathbb{P}\left(\left|\sum_{i=1}^n (X_{i,\eta_n} - X_{i,\eta_n+1} - \mu_n)\right| \leq \kappa\sqrt{n\log(nT_n)}\right)\\ &= 1 - \mathbb{P}\left(\left|\sum_{i=1}^n (X_{i,\eta_n} - X_{i,\eta_n+1} - \mu_n)\right| \geq \kappa\sqrt{n\log(nT_n)}\right).\end{aligned} \tag{5.6}$$

By $(H2)$, we have, for any $i \in \{1,...,n\}$, $X_{i,\eta_n}(\Omega) = X_{i,\eta_n+1}(\Omega) = [0,1]$. Therefore, for any $i \in \{1,...,n\}$, $|X_{i,\eta_n} - X_{i,\eta_n+1} - \mu_n| \leq |X_{i,\eta_n}| + |X_{i,\eta_n+1}| + \mathbb{E}(|X_{1,\eta_n}|) + \mathbb{E}(|X_{1,\eta_n+1}|) \leq 4$. Hence $\sum_{i=1}^n \mathbb{E}\left((X_{i,\eta_n} - X_{i,\eta_n+1} - \mu_n)^2\right) \leq 16n$. The Bernstein inequality applied to the independent, uniformly bounded and centered random variables $(X_{i,\eta_n} - X_{i,\eta_n+1} - \mu_n)_{i\in\{1,...,n\}}$ implies that

$$\begin{aligned}&\mathbb{P}\left(\left|\sum_{i=1}^n (X_{i,\eta_n} - X_{i,\eta_n+1} - \mu_n)\right| \geq \kappa\sqrt{n\log(nT_n)}\right)\\ &\leq 2\exp\left(-\frac{\left(\kappa\sqrt{n\log(nT_n)}\right)^2}{2\left(16n + \frac{4\kappa\sqrt{n\log(nT_n)}}{3}\right)}\right)\\ &= 2\exp\left(-\frac{\kappa^2\log(nT_n)}{2\left(16 + \frac{4\kappa}{3}\sqrt{n^{-1}\log(nT_n)}\right)}\right).\end{aligned}$$

Since, by $(H1)$, for n large enough, $n^{-1}\log(nT_n) \leq 1$, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^n (X_{i,\eta_n} - X_{i,\eta_n+1} - \mu_n)\right| \geq \kappa\sqrt{n\log(nT_n)}\right) \leq 2(nT_n)^{-h(\kappa)}, \tag{5.7}$$

where $h(\kappa) = \kappa^2/(32 + 3^{-1}8\kappa)$.

Putting (5.6) and (5.7) together, we obtain

$$\mathbb{P}\left(\eta_n \in \mathcal{A}_n(\kappa)\right) \geq 1 - 2(nT_n)^{-h(\kappa)}.$$

This proved the first point of Proposition 4.1.

• *Proof of the second point.* We have, for any $m \in \{1, ..., T_n\} - \{\eta_n\}$,

$$\mathbb{P}\left(m \in \mathcal{A}_n(\kappa)\right) = \mathbb{P}\left(\left|\sum_{i=1}^{n}(X_{i,m} - X_{i,m+1})\right| \geq \kappa\sqrt{n\log(nT_n)}\right). \tag{5.8}$$

By $(H2)$, we have, for any $i \in \{1, ..., n\}$ and any $m \in \{1, ..., T_n\}$, $X_{i,m}(\Omega) = X_{i,m+1}(\Omega) = [0, 1]$. Therefore, for any $i \in \{1, ..., n\}$ and any $m \in \{1, ..., T_n\} - \{\eta_n\}$, $|X_{i,m} - X_{i,m+1}| \leq |X_{i,m}| + |X_{i,m+1}| \leq 2$. Hence $\sum_{i=1}^{n} \mathbb{E}\left((X_{i,m} - X_{i,m+1})^2\right) \leq 4n$. The Bernstein inequality applied to the independent, uniformly bounded and centered random variables $(X_{i,m} - X_{i,m+1})_{i\in\{1,...,n\}}$ implies that

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{n}\left(X_{i,m} - X_{i,m+1}\right)\right| \geq \kappa\sqrt{n\log(nT_n)}\right) \\
&\leq 2\exp\left(-\frac{\left(\kappa\sqrt{n\log(nT_n)}\right)^2}{2\left(4n + \frac{2\kappa\sqrt{n\log(nT_n)}}{3}\right)}\right) \\
&= 2\exp\left(-\frac{\kappa^2\log(nT_n)}{2\left(4 + \frac{2\kappa}{3}\sqrt{n^{-1}\log(nT_n)}\right)}\right).
\end{aligned}$$

Since, by $(H1)$, for n large enough, $n^{-1}\log(nT_n) \leq 1$, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\left(X_{i,m} - X_{i,m+1}\right)\right| \geq \kappa\sqrt{n\log(nT_n)}\right) \leq 2(nT_n)^{-t(\kappa)}, \tag{5.9}$$

where $t(\kappa) = \kappa^2/(8 + 3^{-1}4\kappa)$.

It follows from (5.8) and (5.9) that, for any $m \in \{1, ..., T_n\} - \{\eta_n\}$,

$$\mathbb{P}\left(m \in \mathcal{A}_n(\kappa)\right) \leq 2(nT_n)^{-t(\kappa)}.$$

• *Proof of the third point.* It follows from the Bonferroni inequality that

$$\begin{aligned}
\mathbb{P}\left(\widehat{\eta}_n(\kappa) = \eta_n\right) &= \mathbb{P}\left(\cap_{m=1}^{\eta_n-1}\{m \notin \mathcal{A}_n(\kappa)\} \cap \{\eta_n \in \mathcal{A}_n(\kappa)\}\right) \\
&\geq \sum_{m=1}^{\eta_n-1}\mathbb{P}(m \notin \mathcal{A}_n(\kappa)) + \mathbb{P}(\eta_n \in \mathcal{A}_n(\kappa)) - (\eta_n - 1).
\end{aligned}$$

The two first points give

$$
\begin{aligned}
&\mathbb{P}\left(\widehat{\eta}_n(\kappa) = \eta_n\right) \\
&\geq (\eta_n - 1)(1 - 2(nT_n)^{-t(\kappa)}) + 1 - 2(nT_n)^{-t(\kappa)} - (\eta_n - 1) \\
&= 1 - 2(\eta_n - 1)(nT_n)^{-t(\kappa)}) - 2(nT_n)^{-h(\kappa)} \geq 1 - 2\eta_n(nT_n)^{-h(\kappa)}.
\end{aligned}
$$

This completes the proof of Proposition 4.1.

□

Acknowledgment. The author would like to thank Professor Joseph Ngatchou-Wandji and Professor Jean Lejeune for their helpful comments.

References

[1] Cai, T. and Chicken, E. (2005). Block thresholding for density estimation: local and global adaptivity. *Journal of Multivariate Analysis*, 95:76–106.

[2] Chesneau, C. (2007). Wavelet block thresholding for density estimation in the presence of bias. *Preprint LMNO. Submitted (http://hal.archives-ouvertes.fr/docs/00/17/09/89/PDF/cislatex.pdf).*

[3] Chesneau, C. (2007). Wavelet estimation via block thresholding: A minimax study under the lp risk. *Statistica Sinica (to appear).*

[4] Cohen, A., Daubechies, I., Jawerth, B., and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24(1):54–81.

[5] Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L^1 View*. Wiley, New York.

[6] Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory and Applications*. Springer, New York.

[7] Hall, P., Kerkyacharian, G., and Picard, D. (1998). Block thresholding rules for curve estimation using kernel and wavelet methods. *Ann. Statist.*, 26:942–962.

[8] Hall, P., Kerkyacharian, G., and Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators. *Statist. Sinica*, 9(1):33–49.

[9] Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*. vol. 129 of Lecture Notes in Statistics. Springer-Verlag, New York.

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

[11] Petrov, V. V. (1995). *Limit Theorems of Probability Theory*. Clarendon Press, Oxford.

[12] Picard, D. and Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann.Statist.*, 28(1):298–335.

[13] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
[14] Tsybakov, A. (2004). *Introduction à l'estimation nonparamétrique*. Springer Verlag.