Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 346 (2008) 461–466
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics

Nonparametric estimation of the density of the regression noise

Sandra Plancade

MAP5, UMR 8145, Université Paris Descartes, 45, rue des Saints-Pères, 75270 Paris cedex 06, France

Received 12 February 2008; accepted 25 February 2008

Presented by Paul Deheuvels

Abstract

This Note presents an estimator of the density of the error in a homoscedastic regression model, based on model selection methods, and propose a bound for the quadratic integrated risk. ***To cite this article: S. Plancade, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

Résumé

Estimation non-paramétrique de la loi des erreurs dans un modèle de régression. Cette Note présente un estimateur de la loi de l'erreur dans un modèle de régression homoscédastique, basé sur des techniques de sélection de modèle, et propose une majoration du risque quadratique intégré. ***Pour citer cet article : S. Plancade, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

Version française abrégée

On considère le modèle de regression hétéroscédastique suivant :

$$Y_i = b(X_i) + \epsilon_i, \quad i = -n, \ldots, n, \tag{1}$$

où les $\{X_i\}$ sont des variables i.i.d. (indépendantes identiquement distribuées) et les $\{\epsilon_i\}$ sont des variables i.i.d. indépendantes des $\{X_i\}$. Dans cette note nous proposons un estimateur pour la densité f_ϵ de ϵ_1, à partir d'un échantillon $\{X_i, Y_i\}$. Soient les deux échantillons indépendants suivants :

$$Z^- = \{X_i, Y_i\}_{i=-n,\ldots,-1}; \qquad Z^+ = \{X_i, Y_i\}_{i=1,\ldots,n}. \tag{2}$$

A partir de l'échantillon Z^-, on calcule un estimateur $\hat{b}_{\hat{m}}$ de b par une technique de sélection de modèle avec pénalité, basée sur un contraste des moindres carrés. On définit alors les résidus relatifs au deuxième échantillon : $\hat{\epsilon}_i := Y_i - \hat{b}_{\hat{m}}(X_i)\}_{i=1,\ldots,n}$. On applique ensuite une sélection de modèle basée sur un contraste de densité aux $\hat{\epsilon}_i$. On obtient un estimateur f^* de la densité des $\hat{\epsilon}_i$ qui constitue notre estimateur de f_ϵ.

En s'appuyant sur des résultats de convergence des estimateurs de densité et de fonction de régression (présentés dans la section 4), on obtient la majoration suivante pour le risque intégré :

E-mail address: sandra.plancade@math-info.univ-paris5.fr.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2008.02.021

$$\mathbb{E}\big[\|f^* - f_\epsilon\|^2\big] \leqslant C_1 \inf_{m\in\mathcal{M}_n}\left[\inf_{t\in m}\|b-t\|^2 + \sigma^2\frac{D_m}{n}\right] + C_2 \inf_{m\in\Sigma_n}\left[\inf_{t\in m}\|f-t\|^2 + \frac{D_m}{n}\right] \tag{3}$$

où $\mathcal{M}_n$ et Σ_n sont les ensembles de modèles pour l'estimation respectivement de la fonction de régression et de la densité des résidus, et $\|\cdot\|$ désigne la norme L^2 pour la mesure de Lebesgue. En particulier, si b appartient à une classe de Besov plus régulière que f_ϵ, alors l'estimateur f^* atteint la vitesse de convergence minimax (en un sens que l'on précisera dans la dernière section).

1. Introduction

We consider the homoscedastic regression framework (1), where the $\{X_i\}$ are i.i.d. variables with density f_X and the $\{\epsilon_i\}$ are i.i.d.centered variables with density f_ϵ and unknown variance σ^2. Moreover we assume that the $\{X_i\}$ and the $\{\epsilon_i\}$ are independent sequences. The aim of this paper is to propose an estimator for the error density f_ϵ based on the sample $\{(X_i, Y_i)\}_{i=-n,\dots,n}$ drawn from model (1).

There exists a vast literature devoted to density estimation, but the difficulty here is that the errors $\{\epsilon_i\}$ are unobserved. A usual method consists in estimating the regression function b, then computing the residuals $\hat{\epsilon}_i = Y_i - \hat{b}(X_i)$ which are approximations of the ϵ_i's, and finally applying a density estimation procedure to these residuals. Some results have been established using kernels, but the estimators are not adaptive. In a recent article [4], Efromovich proposes an adaptive estimator for the error density, based on a density estimator proposed by Pinsker. His estimator proves very powerful, with a convergence rate that reaches the oracle, but as the risk bound estimation relies on asymptotic developments, these results are mainly asymptotic. Moreover, it requires stronger hypothesis of regularity on f_ϵ than we need. Finally, the estimator we propose proves to be easily computable.

Our estimator is constructed following the above method, using penalized linear model selection to estimate the regression function and the density. More precisely, we split the sample into two independent samples as in (2). From Z^-, an estimate $\hat{\sigma}_n^2$ of the variance σ^2 of ϵ_1 is computed. From the same sample, we build a model selection estimator $\hat{b}_{\hat{m}}$ based on a set of models $\mathcal{M}_n$ and a penalty function of order $\hat{\sigma}_n^2(D_m/n)$. Then we compute the residuals from the second sample: $\hat{\epsilon}_i = Y_i - \hat{b}_{\hat{m}}(X_i)$, $i = 1,\dots,n$. Lastly, a model selection over a set of models Σ_n gives an estimate f^* of the density $f_{\hat{\epsilon}}$ of the residuals: this is our estimator for f_ϵ.

We prove that the quadratic risk of the estimator f^* is bounded by the sum of the oracles for the estimation of b in $\mathcal{M}_n$ and for the estimation of f in Σ_n when the ϵ_i are observed, that is the inequality (3) where $\|t\|^2 = \int t(x)^2\,\mathrm{d}x$ denotes the L^2-norm for Lebesgue measure. Notably, if b belongs to a more regular Besov space than f_ϵ (in a sense we will be more precise later), our estimator reaches the minimax rate of convergence.

2. Notations and estimators

For a density g with respect to the Lebesgue measure, we denote $\|t\|_g^2 := \int t^2(x)g(x)\,\mathrm{d}x$, and for $t \in L^2(g)$, $A \subset L^2(g)$, $d_g(t,A) := \inf_{s\in A}\|s-t\|_g$. We define the empirical norm with respect to the sample Z^-, and the scalar product associated: for every function t, for every $U \in \mathbb{R}^n$, $\|t\|_n^2 = (1/n)\sum_{i=-n}^{-1} t^2(X_i)$ and $\langle t, U\rangle_n = (1/n)\sum_{i=-n}^{-1} t(X_i)U_i$. Throughout the paper, C_i denotes a nonnegative constant that may vary from line to line. We consider the regression framework (1) and build our estimators as follows. We consider two sets of linear models: $\mathcal{M}_n$ and Σ_n. For each model m of $\mathcal{M}_n$ or Σ_n, we denote by D_m its dimension.

Let $\gamma_n(t) = (1/n)\sum_{i=-n}^{-1}(Y_i - t(X_i))^2$ denote the least-squares contrast. For each $m \in \mathcal{M}_n$, we denote by $\hat{b}_m$ the minimizer of γ_n over m. Then, let V_n be a $[n/2]$-dimensional space which contains S_n (the largest space of the collection, defined in Section 3) and $\hat{\sigma}_n^2 = \frac{n}{n-[n/2]}\inf_{t\in V_n}\|t(X) - Y\|_n^2$ an estimator of σ^2 (where $t(X) = (t(X_1,\dots,t(X_n)))$. We select $\hat{m}$ in $\mathcal{M}_n$ as:

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n}\big[\gamma_n(\hat{b}_m) + \mathrm{pen}_1(m)\big]$$

with $\mathrm{pen}_1 = \delta\hat{\sigma}_n^2(D_m/n)$, δ being a constant. We consider the estimator b^* defined as follows:

$$b^* = \hat{b}_{\hat{m}} \quad \text{if } \|\hat{b}_{\hat{m}}\| \leqslant k_n \quad \text{and} \quad b^* = 0 \text{ otherwise,}$$

with $k_n = 2\exp(\ln^2 n)$ (other values for k_n are possible, but this one is convenient for proofs).

We compute the residuals using the second sample: $\hat{\epsilon}_i = Y_i - b^*(X_i)$ for $i = 1, \dots, n$. Let $\nu_n(t) = \|t\|^2 - (2/n)\sum_{i=1}^n t(\hat{\epsilon}_i)$. For each $m \in \Sigma_n$, we denote $\hat{f}_m$ the minimizer of ν_n over m. We select a model $\tilde{m} \in \Sigma_n$ as:

$$\tilde{m} = \arg\min_{m \in \Sigma_n} \left[\nu_n(\hat{f}_m) + \mathrm{pen}_2(m)\right]$$

with a penalty function $\mathrm{pen}_2(m)$ of order D_m/n. Finally, we consider the estimator of f: $f^* = \hat{f}_{\tilde{m}}$.

3. Assumptions

We consider two types of assumptions: the assumptions on the sets of models which determine the allowed models, and the assumptions on the unknown functions.

- H_{frame}: f_X has a known compact support equal to $[0, 1]$, and is upper bounded by $m_1 < \infty$ and lower bounded by $m_0 > 0$. The density f_ϵ has a known compact support $[-1, 1]$, and is Lipschitz on $[-1, 1]$ with a Lipschitz constant $\mathrm{Lip}(f_\epsilon)$. The errors ϵ_i are centered.
- H_{mod}: All models of $\mathcal{M}_n$ (resp. Σ_n) are included in a global model $S_n \subset L^2 \cap L^\infty([0, 1])$ of dimension $N_n \leqslant n^{1/2-d}/\ln^2 n$ for some $d > 0$ (resp. $T_n \subset L^2 \cap L^\infty([-1, 1])$ of dimension $L_n \leqslant n$). Furthermore, the collections $\mathcal{M}_n$ and Σ_n are polynomial, that is there exists some nonnegative constants Γ and R such that $|\{m \in \mathcal{M}_n$ (resp. Σ_n): $D_m = n\} \leqslant \Gamma D^R$. Finally, there exist constants K and K' such that:

$$\|t\|_\infty \leqslant K\sqrt{N_n}\,\|t\|, \quad \forall t \in S_n \quad \text{and} \quad \|s\|_\infty \leqslant K'\sqrt{D_m}\,\|t\|, \quad \forall m' \in \Sigma_n, s \in m'.$$

4. Tools

Theorem 4.1. *Assume that H_{frame}–H_{mod} hold. Consider the penalty function $\mathrm{pen}_1(m) = \delta\hat{\sigma}_n^2(D_m/n)$ for some $\delta > 1$, then*:

$$\mathbb{E}\left[\|b - b^*\|_{f_X}^2\right] \leqslant C \inf_{m \in \mathcal{M}_n}\left[\inf_{t \in m}\|b - t\|^2 + \sigma^2\frac{D_m}{n}\right] + \frac{C'}{n}$$

where C depends on (δ, m_1) and C' depends on $(\sigma, m_0, m_1, \mathbb{E}(\epsilon_1^4))$ and the polynomial structure of $\mathcal{M}_n$.

Proof of Theorem 4.1. We give a sketch of proof based on Baraud's work [1] who proves a similar result when σ is known, adding some elements of proof from [2] to include the case when σ is unknown. Let $\alpha > 1$, $\beta > 1$ and $\gamma > 1$ such that $\delta = \alpha\beta\gamma$. The first step consists in splitting the risk into three terms:

$$\begin{aligned}\mathbb{E}\left[\|b - b^*\|_{f_X}^2\right] &= \mathbb{E}\left[\|b - b^*\|_{f_X}^2 1_A 1_{\|\hat{b}_{\hat{m}}\| < k_n}\right] + \mathbb{E}\left[\|b - b^*\|_{f_X}^2 1_A 1_{\|\hat{b}_{\hat{m}}\| > k_n}\right] + \mathbb{E}\left[\|b - b^*\|_{f_X}^2 1_{A^c}\right] \\ &= E_1 + E_2 + E_3\end{aligned}$$

where $A = \{\omega \in \Omega\colon\ \forall t \in S_n, \|t\|_{f_X}^2 - \|t\|_n^2 \leqslant a\|t\|_{f_X}^2\}$, for some $a \in \left]0, 1 - \frac{2}{\alpha}\right[$. The terms E_2 and E_3 are upper bounded almost like in Baraud [1] (p. 140): $E_2 \leqslant C\|b\|_{f_X}^2(\|b\|_{f_X}^2 + \sigma^2)/k_n^2$ where C is a universal constant and $E_3 \leqslant 2(\|b\|_{f_X} + k_n^2)P(A^c) \leqslant C'/k_n^2$ with C' depending on m_0, m_1, K and $\|b\|_{f_X}$. We upper bound E_1 following the classical method based on Talagrand Inequality. First of all, we notice that, for every $t, s \in S_n$: $\gamma_n(t) - \gamma_n(s) = \|b - t\|_n^2 - \|b - s\|_n^2 - 2\langle\epsilon, t - s\rangle_n$. Then the definition of $\hat{b}_{\hat{m}}$ leads to:

$$\|\hat{b}_{\hat{m}} - b\|_n^2 \leqslant \|b_m - b\|_n^2 + 2\langle\epsilon, \hat{b}_{\hat{m}} - b\rangle_n + \mathrm{pen}_1(m) - \mathrm{pen}_1(\hat{m}) \tag{4}$$

for every $m \in \mathcal{M}_n$ and $b_m \in m$. Moreover, let $B_{m+\hat{m}}^X = \{t \in m + \hat{m}, \|t\|_{f_X} = 1\}$:

$$\begin{aligned}2\langle\epsilon, \hat{b}_{\hat{m}} - b_m\rangle_n &\leqslant 2\|\hat{b}_{\hat{m}} - b_m\|_{f_X}\left\langle\epsilon, \frac{\hat{b}_{\hat{m}} - b_m}{\|\hat{b}_{\hat{m}} - b_m\|_{f_X}}\right\rangle_n \leqslant \frac{1}{\alpha}\|\hat{b}_{\hat{m}} - b_m\|_{f_X}^2 + \alpha\left\langle\epsilon, \frac{\hat{b}_{\hat{m}} - b_m}{\|\hat{b}_{\hat{m}} - b_m\|_{f_X}}\right\rangle_n^2 \\ &\leqslant \frac{1+\theta}{\alpha}\|\hat{b}_{\hat{m}} - b\|_{f_X}^2 + \frac{1}{\alpha}\left(1 + \frac{1}{\theta}\right)\|b - b_m\|_{f_X}^2 + \alpha\sup_{t \in B_{m+\hat{m}}^X}\left\{\langle\epsilon, t\rangle_n^2\right\}\end{aligned} \tag{5}$$

for all $\theta > 0$. Finally, a decomposition similar to the one of Lemma 5.1 [1] establishes that, for every $\nu > 0$:

$$\mathbb{E}\big[\|\hat{b}_{\hat{m}} - b\|_n^2 1_A\big] \geqslant \frac{1-a}{1+\nu}\mathbb{E}\big[\|\hat{b}_{\hat{m}} - b\|_{f_X}\big] - C_1(a,\theta,\nu)d^2_{f_X}(b,m). \tag{6}$$

Gathering inequalities (4), (5) and (6), and taking the expectation over A, for a suitable choice of (a,θ,ν) (that is $(1-a)/(1+\nu) - (1+\theta)/\alpha > 0$), we bound $\mathbb{E}[\|b - \hat{b}_{\hat{m}}\|^2_{f_X} 1_A]$ by:

$$C_1\Big\{\|b - b_m\|^2_{f_X} + \mathbb{E}\big[(\mathrm{pen}(m) - \mathrm{pen}(\hat{m}) + \alpha p(m,\hat{m}))1_A\big] + \mathbb{E}\Big[\Big(\sup_{t\in B^X_{m+\hat{m}}} \langle\epsilon,t\rangle_n^2 - p(m,\hat{m})\Big)_+\Big]\Big\} \tag{7}$$

for every function $p : \mathcal{M}_n \times \mathcal{M}_n \to \mathbb{R}_+$. Let $E_4 = \mathbb{E}[(\sup_{t\in B^X_{m+\hat{m}}} \langle\epsilon,t\rangle_n^2 - p(m,\hat{m}))_+]$. We bound E_4 by using an integrated version of Talagrand Inequality, enunciated for example in [6] (Section 6, Lemma 5). We consider $p(m,m') = \beta\sigma^2(D_m + D_{m'})/n$. Then, as:

$$\mathbb{E}\Big[\sup_{t\in B^X_{m+\hat{m}}} \langle\epsilon,t\rangle_n^2\Big] \leqslant \frac{D_m + D_{m'}}{n}\sigma^2, \qquad \sup_{\|t\|_{f_X}=1} \mathrm{Var}\big(t(X_1)\epsilon_1\big) \leqslant \sigma^2, \qquad \sup_{\|t\|_{f_X}=1} \big\|t(X_1)\epsilon_1\big\|_\infty \leqslant K'\sqrt{N_n}.$$

We get:

$$E_4 \leqslant \sum_{m'\in\mathcal{M}_n} C_1\left\{\frac{\sigma^2}{n}\exp[-\bar{\kappa} D_{m'}] + \frac{4K^2}{m_0^2}\frac{1}{n}\exp\left[-\frac{\overline{\kappa'}\sigma m_0}{2K}n^{1/4}\right]\right\} \leqslant \frac{C_2}{n}. \tag{8}$$

It remains to bound $E_5 := \mathbb{E}[(\mathrm{pen}_1(m) - \mathrm{pen}_1(\hat{m}) + \alpha p(m,\hat{m}))1_A]$. Let $B = \{\hat{\sigma}_n^2 \geqslant \frac{1}{\gamma}\sigma^2\}$, then $P(B^c) \leqslant C_3/n$, where C_3 depends on γ and $\mathbb{E}[\epsilon_1^4]$. In addition $\mathbb{E}[\hat{\sigma}_n^2] \leqslant \sigma^2 + 2\inf_{t\in V_n}\|b-t\|^2_{f_X} \leqslant \sigma^2 + \inf_{t\in m}\|b-t\|^2_{f_X}$. These results derive from a claim in [2] (p. 487), by considering the conditional probability on the designs (X_i). Hence:

$$E_5 \leqslant \mathbb{E}\left[2\gamma\hat{\sigma}_n^2\frac{D_m}{n} + \alpha\beta\sigma^2\frac{D_{\hat{m}} + D_m}{n}1_{B^c}\right] \leqslant 2\,\mathrm{pen}_1(m) + 4\delta\inf_{t\in m}\|t - b\|^2_{f_X} + 2\delta\sigma^2\frac{C_3}{n}. \tag{9}$$

The result follows by gathering (7), (8) and (9). □

We use the following result for density estimation via model selection given by Massart [7] (p. 216):

Theorem 4.2. *Let* $(U_1,\dots,U_n)$ *be a sample from a density* g *with respect to Lebesgue measure, with support* $[-1,1]$. *Let* Σ_n *be a set of models satisfying* H_{mod}. *For every* $m \in \Sigma_n$, *let* $\hat{g}_m$ *be the minimizer of* $\nu_n(t) := \|t\|^2 - 2/n\sum_{i=1}^n t(U_i)$ *over* m. *Let* $\beta > 1$, *and* $\mathrm{pen}_2(m) = \beta K'(D_m/n)$ *where* K' *is defined in* H_{mod}, *and* $\hat{m} = \arg\min_{m\in\Sigma_n}[\nu_n(\hat{g}_m) + \mathrm{pen}_2(m)]$, *then*:

$$\mathbb{E}\big[\|g - \hat{g}_{\hat{m}}\|^2\big] \leqslant C(K',\beta)\inf_{m\in\Sigma_n}\left[d^2(g,m) + \frac{D_m}{n}\right] + \frac{C'(K',\beta,\Gamma)\max(1,\|g\|^{4+2R})}{n}.$$

5. Main results

Using Proposition 4.1 and Theorem 4.2, we obtain our main result:

Theorem 5.1. *Assume that* H_{frame} *and* H_{mod} *holds. Consider* $pen_1(m) = \delta\hat{\sigma}_n\frac{D_m}{n}$ *and* $\mathrm{pen}_2(m) = \beta K'\frac{D_m}{n}$ *with* $\delta > 0$ *and* $\beta > 0$. *Then*:

$$\mathbb{E}\big[\|f^* - f_\epsilon\|^2_{f_X}\big] \leqslant C\inf_{m\in\mathcal{M}_n}\left[d^2(b,m) + \sigma^2\frac{D_m}{n}\right] + C'\inf_{m\in\Sigma_n}\left[d^2(f,m) + \frac{D_m}{n}\right] + \frac{C''}{n}$$

where C *depends on* $(K',\beta,\delta,\mathrm{Lip}(f))$, C' *on* (K',β) *and* C'' *on* $(K',\beta,\Gamma,\|f\|_{f_X})$.

Remark. Let $\mathcal{B}^\alpha_\infty(L_p)$ denote the Besov space of parameter $\alpha > 0$, $p > 0$ and ∞, $\|\cdot\|_{\alpha,p}$ the corresponding semi-norm (see [3]) and for $R > 0$, $\mathcal{B}_{\alpha,p,\infty}(R) = \{f \in \mathcal{B}^\alpha_\infty(L_p)\colon\ \|f\|_{\alpha,p} \leqslant R\}$. Suppose that models are generated by either trigonometric polynomials, or piecewise polynomials , or wavelet families (see [1], Section 2 for more details). If there exists $0 < \alpha \leqslant \alpha'$, and $(R, R') \in (\mathbb{R}_+^*)^2$ such that $f_\epsilon \in \mathcal{B}_{\alpha,2,\infty}(R)$ and $b \in \mathcal{B}_{\alpha',2,\infty}(R')$, then the estimator f^* of f_ϵ reaches the minimax rate of convergence for a density function over $\mathcal{B}_{\alpha,2\infty}$.

Indeed, classical results about approximation (see [5] for wavelets and [3] for piecewise and trigonometric polynomials) prove that $\inf_{t\in m} \|f_\epsilon - t\| \leqslant C D_m^{-\alpha}$ and $\inf_{t\in m'} \|b - t\| \leqslant C' D_{m'}^{-\alpha'}$. Then as $\alpha \leqslant \alpha'$, a simple computation leads to:

$$\inf_{m\in\mathcal{M}_n}\left[d^2(b,m) + \Sigma^2\frac{D_m}{n}\right] + \inf_{m\in\Sigma_n}\left[d^2(f,m) + \sigma^2\frac{D_m}{n}\right] \leqslant C_1\left(n^{-2\alpha/(2\alpha+1)} + n^{-2\alpha'/(2\alpha'+1)}\right) \leqslant C_2 n^{-2\alpha/(2\alpha+1)}.$$

In addition, it is well known (see for example [5]) that the minimax rate for a density estimation over $\mathcal{B}_{\alpha,2,\infty}(R)$ is of order $n^{-2\alpha/(2\alpha+1)}$, which proves the result.

Proof of the Theorem 5.1. Let f^- denote the density of $\hat{\epsilon}_1$, conditionally to Z^-. The proof relies on the following facts: conditionally to Z^-, $\mathbb{E}(\|f^* - f^-\|^2)$ is bounded with Theorem 4.2. Then, f^- being the density of $\epsilon_1 + (b - b^*)(X_i)$, it can be expressed as the convolution of f and the density of $(b - b^*)(X_1)$ so conditionally to Z^-, $\|f^- - f\|$ is bounded in function of $\|b - b^*\|$. Finally, $\mathbb{E}(\|b - b^*\|^2)$ is bounded by means of Proposition 4.1.

More precisely: $\mathbb{E}(\|f^* - f_\epsilon\|^2) \leqslant 2\mathbb{E}(\|f^* - f^-\|^2) + 2\mathbb{E}(\|f^- - f_\epsilon\|^2)$. Now, given Z^-, we have:

$$\begin{aligned}\mathbb{E}\left(\|f^* - f^-\|^2 | Z^-\right) &\leqslant C_1 \inf_{m\in\Sigma_n}\left[\inf_{t\in m}\|f^- - t\|^2 + \mathrm{pen}_2(m)\right] + C_2\max\left(1, \|f^-\|^{4+2R}\right)/n \\ &\leqslant C_1 \inf_{m\in\Sigma_n}\left[2\inf_{t\in m}\|f_\epsilon - t\|^2 + 2\|f_\epsilon - f^-\|^2 + \mathrm{pen}_2(m)\right] + C_2\max\left(1, \|f^-\|^{4+2R}\right)/n \\ &= C_1\|f_\epsilon - f^-\|^2 + C_2 \inf_{m\in\Sigma_n}\left[\inf_{t\in m}\|t - f_\epsilon\|^2 + \mathrm{pen}_2(m)\right] + C_3\max\left(1, \|f^-\|^{4+2R}\right)/n.\end{aligned}$$

We derive that:

$$\mathbb{E}\left[\|f^* - f_\epsilon\|^2\right] \leqslant C_1 \inf_{m\in\Sigma_n}\left[\inf_{t\in m}\|f_\epsilon - t\|^2 + \mathrm{pen}_2(m)\right] + C_2\mathbb{E}\left[\|f_\epsilon - f^-\|^2\right] + \frac{C'}{n}\mathbb{E}\left[\|f^-\|^{4+2R}\right]. \tag{10}$$

Now as $f^-(y) = \int_{x=0}^1 f_\epsilon(y - (b - b^*)(x)) f_X(x)\,\mathrm{d}x$, Jensen inequality gives, as a bound of $\|f_\epsilon - f^-\|^2 = \int_{y\in\mathbb{R}} (f^-(y) - f_\epsilon(y))^2\,\mathrm{d}y$:

$$\begin{aligned}&\int_{y\in\mathbb{R}}\left[\int_{x=0}^1 \left(f_\epsilon\left(y - (b - b^*)(x)\right) - f_\epsilon(y)\right) f_X(x)\,\mathrm{d}x\right]^2\mathrm{d}y \\ &\quad\leqslant \int_{y\in\mathbb{R}}\int_{x=0}^1 \left[f_\epsilon\left(y - (b - b^*)(x)\right) - f_\epsilon(y)\right]^2 f_X(x)\,\mathrm{d}x\,\mathrm{d}y \\ &\quad= \int_{x=0}^1\left(\int_{y\in\mathbb{R}}\left[f_\epsilon\left(y - (b - b^*)(x)\right) - f_\epsilon(y)\right]^2\mathrm{d}y\right) f_X(x)\,\mathrm{d}x.\end{aligned}$$

But f_ϵ has a support in $[0, 1]$, so $[f_\epsilon(y - (b - b^*)(x)) - f_\epsilon(y)] = 0$ for all $y \notin [0, 1] \cup [(b - b^*)(x), (b - b^*)(x) + 1] := U(x)$, and $\int 1_{U(x)}\,\mathrm{d}x \leqslant 2$. Hence

$$\begin{aligned}\|f_\epsilon - f^-\|^2 &\leqslant \int_{x=0}^1\int_{y\in\mathbb{R}}\left[f_\epsilon\left(y - (b - b^*)(x)\right) - f_\epsilon(y)\right]^2 1_{U(x)} f_X(x)\,\mathrm{d}y\,\mathrm{d}x \\ &\leqslant \int_{x=0}^1 2\,\mathrm{Lip}(f_\epsilon)^2 (b - b^*)(x)^2 f_X(x)\,\mathrm{d}x.\end{aligned}$$

Thus $\|f_\epsilon - f^-\|^2 \leqslant \mathrm{Lip}(f_\epsilon)^2 \|b - b^*\|^2_{f_X}$. By integrating over Z^-, and applying Proposition 4.1, we get:

$$\mathbb{E}\big(\|f_\epsilon - f^-\|^2\big) \leqslant C_1 \,\mathrm{Lip}(f_\epsilon)^2 \inf_{m \in \mathcal{M}_n} \Big[\inf_{t \in m} \|b - t\|^2_{f_X} + \mathrm{pen}_1(m)\Big] + \frac{C_2}{n}. \tag{11}$$

With the same arguments over the support of f^-, we get:

$$\|f^-\|^2 = \int_{y\in\mathbb{R}} \Bigg[\int_{x=0}^{1} f_\epsilon\big(y - (b - b^*)(x)\big) f_X(x)\,\mathrm{d}x\Bigg]^2 \mathrm{d}y \leqslant \int_{x=0}^{1} \int_{y\in\mathbb{R}} \big[f_\epsilon\big(y - (b - b^*)(x)\big)\big]^2 \,\mathrm{d}y f_X(x)\,\mathrm{d}x \tag{12}$$

i.e. $\|f^-\|^2 \leqslant \|f_\epsilon\|^2_{f_X}$. Gathering the inequalities (10)–(12) the result follows. □

References

[1] Y. Baraud, Model selection for regression on a random design, ESAIM Probab. Statist. 6 (2002) 127–146.

[2] Y. Baraud, Model selection for regression on a fixed design, Probab. Theory Related Fields 117 (2000) 467–493.

[3] R.A. DeVore, G.G. Lorentz, Constructive Approximation, Springer-Verlag, 1993.

[4] S. Efromovich, Estimation of the density of regression errors, Annals of Statistics 33 (2005) 2194–2227.

[5] W. Härdle, G. Kerbyacharian, D. Picard, A. Tsybakov, Wavelets, Approximation, and Statistical Applications, Lecture Notes in Statistics, vol. 129, 1998.

[6] C. Lacour, Adaptative estimation of the transition density of a particular hidden Markov chain, J. Multivariate Anal. 2006, in press, available online; arXiv: math/0611681v1 [math.ST], hal-00115612, version 1.

[7] P. Massart, Concentration inequalities and model selection, in: Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, in: Lecture Notes in Mathematics, vol. 1896, Springer, Berlin, 2007.