ADAPTIVE DENSITY DECONVOLUTION FOR CIRCULAR DATA

F. COMTE[1] AND M.-L. TAUPIN[2]

Abstract. We consider the problem of density deconvolution in the context of circular random variable. We aim at estimating the density of a random variable X from a sample $Z_1, \cdots, Z_n$ in the convolution model where $Z_i = X_i + \varepsilon_i$, $i = 1, \ldots, n$ and ε is a noise independent of X, all the variables being defined on $\mathbb{R}$-modulo 2π. In this context, we propose an adaptive estimator of the density of X by using a model selection procedure allowing to find non-asymptotic bounds for the integrated quadratic risk. These bounds hold in the independent case as well as in the dependent case.

Keywords and phrases. Absolutely regular random variables. Adaptive estimation. Circular variables. Density deconvolution. Model selection. Penalized contrast estimator.

September 2003

1. Introduction

In this paper, we consider the problem of density deconvolution with circular random variables. Consider the following convolution model

$$Z = X + \varepsilon \tag{1}$$

where the variables Z, X and ε are circular random variables defined on $\mathbb{R}$-modulo 2π and ε is a noise independent of X. We denote by μ_π the Lebesgue measure on $\mathbb{R}$-modulo 2π, by $\mathbb{L}_2(\mu_\pi)$ the set of square integrable functions with respect to μ_π, and by $< ., . >$ the associated scalar product. We assume that the errors ε are circular random variables having known density f_ε with respect to μ_π, and are independent from the X_i's. We also assume that the circular random variables X_i's are identically distributed and admit a density, denoted by g, with respect to the measure μ_π. In this model we are interested in the problem of the nonparametric estimation of the density g of the X_i's from the sample $Z_1, \cdots, Z_n$ of the circular random variable Z. Due to the independence between X and ε, the density of Z is the convolution product (on $\mathbb{R}$-modulo 2π) of g, the density of X, and the one of ε, f_ε (see section 2.1). This motivates the term of deconvolution even in the context of circular random variables.

Let us describe previous known results on density deconvolution, starting with those for real random variables. The problem of density deconvolution for random variables on the real line appears frequently in the literature. In particular, it has been widely studied

[1] Université Paris V, MAP5, FRE CNRS 2428. email: comte@biomedicale.univ-paris5.fr.

[2] IUT de Paris V et Université d'Orsay, Address of the author to whom the proofs should be sent: M.L. Taupin, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8628,
Université Paris-Sud, Bâtiment 425, 91405 Orsay Cedex, France.
email: Marie-Luce.Taupin@math.u-psud.fr.

using kernel estimators and Fourier transform properties. One can for instance cite Carroll and Hall (1988), Devroye (1989), Fan (1991a, b), Liu and Taylor (1989), Masry (1991, 1993a, b), Stefansky (1990), Stefansky and Carroll (1990), Taylor and Zhang (1990) and Zhang (1990), Cator (2001), Youndje and Wells (2002). Those estimators were studied from many points of view: pointwise and global asymptotic optimality, asymptotic normality, case of dependent ε_i's...

In the same model but using a different estimator, Koo (1999) considers the problem of logspline density deconvolution when the log-density belongs to a Besov space and the errors are ordinary or super smooth and obtains usual rates of convergence.

It is well known that two factors determine the estimation accuracy in the standard density deconvolution problem : first the smoothness of the density to be estimated and second the smoothness of the error density. The smoother the error density, the slower the optimal rate of convergence: logarithmic rates of convergence appear when the error density is super smooth, in the standard context of ordinary smooth density g. All these smoothness properties are described by the rate of decay of the Fourier transforms: polynomial decay for what is called ordinary smoothness (or Sobolev-type smoothness) and exponential decay for super smooth functions. Most previous results concern density to be estimated with Sobolev-type smoothness and ordinary or super smooth error density.

In most of the previous papers, the smoothness parameters of the unknown density are supposed to be known and thus those papers deal with non adaptive estimation. Let us now give more details on recent results about adaptive estimation.

In the context of pointwise density deconvolution in the real Gaussian white noise model, Goldenshluger (1999) proposes an adaptive estimator when the errors are ordinary smooth or super smooth and the density g has Sobolev-type smoothness. Its pointwise quadratic risk converges with a rate logarithmically slower than the minimax rate, when the error density is ordinary smooth and with the minimax rate when the error density is super smooth.

Following the development of wavelet methods, Pensky and Vidakovic (1999), Pensky (2002) and Fan and Koo (2002) study some wavelet thresholding estimators to build adaptive estimators in the context of density deconvolution for real variables. In particular, Pensky and Vidakovic (1999) study the case of super smooth functions in presence of ordinary or super smooth errors which allows to recover much faster rates.

Again, for real variables, the case of super smooth functions is also studied by Butucea (2003) in case of ordinary smooth error density and by Butucea and Tsybakov (2003) when both the density g and the error density f_ε are super smooth. In both papers, exact upper and lower bounds for pointwise and integrated quadratic risk are given, using kernel estimator. One consequence of Butucea and Tsybakov's (2003) paper is that Pensky and Vidakovic's (1999) results are sub-optimal when both g and f_ε are super smooth. This sub-optimality comes from the use of wavelets. In Comte and Taupin (2003) a penalized contrast and adaptive estimator is proposed. Its construction is based on a development of the function to be estimated in an orthonormal basis generated by the function $\sin(\pi x)/(\pi x)$. This procedure allows to construct an adaptive and optimal or nearly optimal estimator of the density by using model selection.

Let us now come to density deconvolution for circular data. Efromovich (1997) proposes an estimator constructed as a truncated development of the density to be estimated in the trigonometric basis where the theoretical coefficients are replaced by empirical estimators. Let us be more precise. Denote by u^*, the Fourier transform of the function u in $\mathbb{L}^2([0,2\pi])$ defined as $u^*(j)=\int_0^{2\pi}\exp\{ijt\}u(t)dt$ for j an integer and $i^2=-1$. If the Fourier transform g^* of the density of X is integrable, then g can be written as

$$g(x)=\frac{1}{2\pi}\sum_{j\in\mathbb{Z}}g^*(j)\exp\{-ijx\},\ 0\le x\le 2\pi.$$

In Model (1), if h denotes the density of Z, then (see section 2.1) $h^*(j)=g^*(j)f_\varepsilon^*(j)$, for $j\in\mathbb{N}$. Therefore as soon as f_ε^* does not vanish, he proposes to estimate $g(x)$ by

$$\widehat{g}(x)=\frac{1}{2\pi}\sum_{|j|\le J_n}\frac{1}{n}\sum_{k=1}^{n}\frac{\exp\{ijZ_k\}}{f_\varepsilon^*(j)}\exp\{-ijx\}, \tag{2}$$

for some J_n to be well chosen. For the pointwise quadratic risk and the integrated quadratic risk, he establishes sharp minimax results for the estimation of g and its derivatives when the errors are super smooth and the density g is ordinary smooth. He also gives results when the errors density f_ε is unknown, by using additional observations. The rates of convergence are the same as those obtained for random variables on the real line.

In a slightly different density deconvolution model, Goldenshluger (2001) considers the problem of the estimation of the density f of a random variable θ in the model

$$Z_t=\exp\{i\theta_t\}+\varepsilon_t,\ t=1,\cdots,n,$$

where the errors ε have known density defined on the complex plane $\mathbb{C}=\mathbb{R}^2$. In this context the behavior of the minimax risk does not depend on the error density.

In the present work, we follow the idea of Efromovich (1997) based on the development of the unknown density function in the trigonometric basis by noting that the estimator defined by (2) can be seen as a minimum contrast estimator. Indeed, if we denote by $\varphi_j(x)=\exp\{-ijx\}/\sqrt{2\pi}$, this estimator is the projection estimator on the space

$$\mathrm{Vect}\{\varphi_j,|j|\le J_n\},$$

associated to the contrast

$$\|\ t\ \|^2_{\mathbb{L}_2([0,\pi])}-2\sum_{|j|\le J_n}\frac{1}{n}\sum_{i=1}^{n}\frac{\langle t,\varphi_j\rangle\,\varphi_j(Z_i)}{f_\varepsilon^*(-j)}.$$

As proved by Efromovich (1997), when the density g is ordinary smooth and the errors are super smooth, this estimator is optimal in a sharp minimax sense. According to Butucea's (2003), and Butucea and Tsybakov's (2003) results for variables on the real line, it seems reasonable to infer that it is also optimal when g is super smooth. Nevertheless the choice of the optimal J_n depends on the regularity of the unknown density. We deduce from these facts, by using a model selection procedure, a penalized contrast estimator which is adaptive and optimal or nearly optimal. More precisely, we establish a non-asymptotic bound for the integrated quadratic risk of the penalized contrast estimator that ensures an automatic trade-off between the squared bias of (2) and a penalty term having the same order as the variance of (2), or the same order as the variance up to some logarithmic factor. This possible loss in the penalty only appears in cases where such a logarithmic term is negligible with respect to the rates. In particular, our estimator is adaptive and optimal in a minimax sense when the density g is ordinary smooth, that is, in the cases described by Efromovich (1997), and is probably optimal (or nearly optimal) when g is super smooth.

Let us go further in the comparison between the rates of convergence for adaptive estimation obtained for random variables on the real line and those obtained for circular random variables. According to Comte and Taupin's (2003)result, it appears that the results obtained for circular data are very close to the former ones except in one case where the rate for the adaptive estimator for circular data is better than the one obtained for variables on the real line. This a case where a logarithmic loss appearing for real line random variables does not appear for circular variables: for instance, the logarithmic loss is avoided for Gaussian errors (see also Remark 3.3).

The problem of a logarithmic loss in adaptation is already known in different models but it generally appears only for the pointwise estimation. For instance, in Goldenshluger (1999) the adaptive estimator has a pointwise risk converging with a logarithmic loss with respect to the minimax rate when both the errors density and g are ordinary smooth. But in Comte and Taupin (2003), in the same case, the adaptive penalized contrast estimator has an optimal integrated quadratic risk. Nevertheless, such a logarithmic loss in adaptation, for $\mathbb{L}_2$ estimation has been underscored by Tsybakov (2000) in case of inverse problems for real random variables. In particular he exhibits some example where a logarithmic loss in adaptation is not avoidable. His example for real line variables can be compared to some particular case of our results for circular variables, where our estimator has its quadratic risk with the same logarithmic loss. This logarithmic loss in adaptation seems thus non avoidable at least in one case (see Remark 3.2).

The methodology following the line of model selection strategies as presented in Barron et al (1999) allows, as in Comte and Taupin (2003) straightforward extensions to non independent variables (β-mixing random variables).

In Section 2, we present some generalities about circular data, the model, the assumptions, the estimator and the aim of the study. Section 3 states the results while the proofs are gathered in Section 4.

2. Description of the problem

2.1. **About circular data.** We briefly describe what is called circular random variables and refer for instance to Mardia (1972) for more details on the subject.

We say that a random variable Y is circular or wrapped (around the circumference of the unit circle) if its density is such that $f_{Y_w} \geq 0$ with $\int f_{Y_w}(x)\mu_\pi(dx) = 1$. This implies in particular that f_{Y_w} is a 2π-periodic function on $\mathbb{R}$ with $\int_0^{2\pi} f_{Y_w}(y)dy = 1$. Moreover, if Y is a circular random variable and ψ a function defined on $\mathbb{R}$-modulo 2π, then $\mathbb{E}(\psi(Y)) = \int \psi(y) f_{Y_w}(y)\mu_\pi(dy)$.

If $Y_w = Y(\mod 2\pi)$ is a wrapped random variable coming from a random variable on the real line Y with density f_Y, then the density f_{Y_w}, of Y_w is given by

$$f_{Y_w}(x) = \sum_{k\in\mathbb{Z}} f_Y(x + 2k\pi).$$

The characteristic function of such a random variable Y, is defined as $f_Y^*(t) = \mathbb{E}(\exp\{itY\}) = \int_{\mathbb{R}} \exp\{ity\} f_Y(y)dy$ and satisfies $f_Y^*(t) = f_{Y_w}^*(t)$ for any integer value of t. In fact the theory of Fourier series for 2π-periodic functions shows that, in order to define f_{Y_w}, it is sufficient to take t as an integer in the definition of $f^*_{Y_w}$ (see Mardia (1972), Section 3.2 and Rudin (1962), Section 1.2 for further details).

Subsequently, for f_1 and f_2 two functions defined on $\mathbb{R}$-modulo 2π, belonging to $\mathbb{L}_2(\mu_\pi)$, we denote by

$$\langle f_1, f_2\rangle = \int f_1(x)\overline{f_2}(x)\mu_\pi(dx), \quad \text{and} \quad \| f_1 \|^2 = \| f_1 \|^2_{\mathbb{L}_2(\mu_\pi)} = \int |f_1(x)|^2 \mu_\pi(dx),$$

where $z\overline{z} = |z|^2$. If the density f_{Y_w} defined on $\mathbb{R}$-modulo 2π, belongs to $\mathbb{L}_2(\mu_\pi)$, then it can be developed in the trigonometric basis as follows

$$f_{Y_w}(x) = \sum_{k\in\mathbb{Z}} a_j(f_{Y_w})\varphi_j(x),$$

with

$$a_j(f_{Y_w}) = \langle f_{Y_w}, \varphi_j\rangle\,, \text{ and } \varphi_j(x) = \exp\{-ijx\}/\sqrt{2\pi}. \tag{3}$$

We have moreover that the Fourier coefficient $a_j(f_{Y_w})$ of the density of Y_w is such that, for j an integer,

$$a_j(f_{Y_w}) = f^*_{Y_w}(j)/\sqrt{2\pi} = f^*_Y(j)/\sqrt{2\pi}.$$

Let us come to Model (1). For j an integer, by using the independence between X and ε, we have $\mathbb{E}[\exp\{ijZ\}] = \mathbb{E}[\exp\{ijX\}]\mathbb{E}[\exp\{ij\varepsilon\}]$, and consequently,

$$h^*(j) = g^*(j)f^*_\varepsilon(j), \tag{4}$$

Furthermore we have $h = g \star f_\varepsilon$, where, for two funtions f_1 and f_2 defined on $\mathbb{R}$-modulo 2π and belonging to $\mathbb{L}_2(\mu_\pi)$, we define the convolution product of f_1 and f_2 by

$$f_1 \star f_2(z) = \int f_1(x)\overline{f_2}(z-x)\mu_\pi(dx).$$

We refer to Mardia (1972), Fisher (1993), and Efromovich (1997) for further details as examples, models, results or references on circular variables.

2.2. **Model and Assumptions.** Consider Model (1) and the following assumptions.

- **A1** The X_i's and the ε_i's are identically distributed random variables.
- **A2** The sequences $(X_i)_{i\in\mathbb{N}}$ and $(\varepsilon_i)_{i\in\mathbb{N}}$ are independent.
- **A3** The ε_i's and the X_i's are independent random variables.
- **A3'** The ε_i's and the X_i's are both absolutely regular (or β-mixing, see Doukhan (1994), pp.4-5).

Under **A2** and **A3**, the ε_i's and the X_i's are both independent and identically distributed random variables and therefore so is the sequence of the Z_i's.

Whereas, under **A2** and **A3'**, the sequence of the (X_i, ε_i)'s is also absolutely regular and therefore so is the sequence of the Z_i's. We denote by $(\beta_k)_{k\in\mathbb{N}}$ the mixing coefficients of this last sequence.

The following assumption will be required for g.

A4 The density function g of the Z_i's, defined on $\mathbb{R}$-modulo 2π, belongs to $\mathbb{L}_2(\mu_\pi)$.

Our aim is to estimate g from a sample $(Z_1, \dots, Z_n)$.

As it is mentioned in the introduction, the rate of convergence for estimating g is strongly related to smoothness of the error density f_ε, described by the rate of decrease of the Fourier Transform $f^*_\varepsilon(j)$ as j goes to infinity. More precisely, the smoother f_ε, the slower is the rate of convergence for estimating g. Nevertheless, this rate of convergence can be improved by assuming some additional regularity conditions on g. These regularity conditions are described by considering functions of the space $\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)$ defined by

$$\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha) = \left\{ f \text{ density } : \sum_{j\in\mathbb{Z}} |f^*(j)|^2 |j|^{2\alpha} \exp\{2\rho|j|^\nu\} \le A_\alpha \right\}, \tag{5}$$

where α, ρ, ν are nonnegative real numbers. When $\nu = 0$, this corresponds to some Sobolev spaces of order α densities with respect to μ_π. When $\nu > 0$, this corresponds to analytic functions, often called "super-smooth" functions. For the sake of simplicity, we shall set the constant $\rho = 0$ when $\nu = 0$ and assume that $\rho > 0$ as soon as $\nu > 0$.
At the same time we assume some slightly different conditions on f_ε described as follows.

A5 The density f_ε^* does not vanish and belongs to $\mathbb{L}_2(\mu_\pi)$.
A6 There exist nonnegative real numbers γ, B and δ such that, for all j in $\mathbb{Z}$,

$$|f_\varepsilon^*(j)| \ge A_0(j^2+1)^{-\gamma/2} \exp\{-B|j|^\delta\}.$$

Assumption **A7** when $\delta = 0$ amounts to consider what we also call "ordinary smooth" errors, and "super smooth" errors when $\delta > 0$. For the sake of simplicity, we set $B = 0$ when $\delta = 0$ and we assume that $B > 0$ when $\delta > 0$. Assumption **A7** holds for many practically important wrapped distributions on the real line. When $\delta > 0$ it includes Gaussian or Cauchy wrapped distributions and when $\delta = 0$ it includes for instance the double exponential.

It is noteworthy that, again for sake of simplicity, the terms "ordinary smooth" and "super smooth" can be as well as used for the density g and for the errors density with a slight difference in the definition.

2.3. **The projection spaces and the estimators.** Let us consider a collection of spaces $(S_m)_{m\in\mathcal{M}_n}$, where a space (or a model) S_m is a finite dimensional linear space of the form

$$S_m = \text{Vect}\{\varphi_j,\ |j| \le D_m\}, \text{ with } \dim(S_m) = 2D_m + 1, \tag{6}$$

and the basis functions φ_j are defined by (3). In our context, $D_m = m$ or $D_m = 2^m$ for m a nonnegative integer; therefore we consider that, for two nonnegative integers m and m', $D_{m\vee m'} = D_m \vee D_{m'}$, the models are embedded models and that the cardinality of $\mathcal{M}_n$ equals m_n or $\log_2(m_n)$.

Denoting by g_m the orthogonal projection of g on the space S_m, g_m is given by

$$g_m = \sum_{|j|\le D_m} a_j(g)\varphi_j \text{ with } a_j(g) = \langle g, \varphi_j\rangle.$$

Associate to this collection of models the following contrast function, for t belonging to some model S_m of the collection $(S_m)_{m\in\mathcal{M}_n}$

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^n \left[\|t\|^2 - 2V_t(Z_i)\right], \quad \text{with} \quad V_t(x) = \sum_{|j|\le D_m} \frac{\langle t, \varphi_j\rangle}{f_\varepsilon^*(-j)}\varphi_j(x).$$

According to (4),

$$\begin{aligned}\mathbb{E}\left[V_t(Z_i)\right] &= \sum_{|j|\le D_m} \frac{\langle t,\varphi_j\rangle}{f_\varepsilon^*(-j)}\mathbb{E}(\varphi_j(Z_1)) = \frac{1}{\sqrt{2\pi}}\sum_{|j|\le D_m}\frac{\langle t,\varphi_j\rangle}{f_\varepsilon^*(-j)} f_\varepsilon^*(-j)g^*(-j)\\ &= \sum_{|j|\le D_m}\langle t,\varphi_j\rangle\overline{\langle g,\varphi_j\rangle} = \langle t, g\rangle,\end{aligned}$$

and we find that $\mathbb{E}(\gamma_n(t)) = \|t-g\|^2 - \|g\|^2$, which is all the smaller that t is nearer of g. This illustrates that $\gamma_n(t)$ is the relevant choice for the empirical version of the $\mathbb{L}_2(\mu_\pi)$ distance between t and g.

Associated to the collection of models , the collection of estimators $\hat g_m$ of g is defined by

$$\hat g_m = \text{ Argmin }_{t\in S_m}\gamma_n(t), \tag{7}$$

where by using that, $t\mapsto V_t$ is linear, and that $(\varphi_j)_{|j|\le D_m}$ is an orthonormal basis of S_m, we have

$$\hat g_m = \sum_{|j|\le D_m}\hat a_j(g)\varphi_j \quad\text{where}\quad \hat a_j(g) = \frac{1}{nf_\varepsilon^*(j)}\sum_{k=1}^n\overline{\varphi_j(Z_k)} = \frac{1}{n\sqrt{2\pi}f_\varepsilon^*(j)}\sum_{k=1}^n\exp\{ijZ_k\},$$

with $\mathbb{E}(\hat a_j(g)) = < g,\varphi_j >= a_j(g)$.

2.4. **The aim of the study.** In order to motivate our approach let us study the rate of convergence of one estimator $\hat g_m$. According to (7), for any given m belonging to $\mathcal{M}_n$, $\hat g_m$ satisfies,

$$\gamma_n(\hat g_m) - \gamma_n(g_m) \le 0. \tag{8}$$

Denoting by $\nu_n(t)$ the centered empirical process

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^n\left[V_t(Z_i) - \langle t,g\rangle\right], \tag{9}$$

we have that

$$\gamma_n(t)-\gamma_n(s) = \|t-g\|^2 - \|s-g\|^2 - 2\nu_n(t-s) \tag{10}$$

and therefore, by using (8), we deduce that

$$\|g-\hat g_m\|^2 \le \|g-g_m\|^2 + 2\nu_n(\hat g_m - g_m).$$

Since $\hat a_j(g) - a_j(g) = \nu_n(\varphi_j)$, we get that

$$\nu_n(\hat g_m - g_m) = \sum_{|j|\le D_m}(\hat a_j(g)-a_j(g))\nu_n(\varphi_j) = \sum_{|j|\le D_m}[\nu_n(\varphi_j)]^2, \tag{11}$$

and consequently

$$\mathbb{E}\|g-\hat g_m\|^2 \le \|g-g_m\|^2 + 2\sum_{|j|\le D_m}\text{Var}[\nu_n(\varphi_j)]. \tag{12}$$

The rate of convergence of $\hat g_m$ is obtained by selecting the space S_m that makes $\|g - g_m\|^2 + 2\sum_{|j|\le D_m}\text{Var}[\nu_n(\varphi_j)]$ as small as possible. Let us study the order of the bias term $\|g-g_m\|^2 = \sum_{|j|>D_m}|g^*(j)|^2/(2\pi)$ which depends on the smoothness of the function g.

Consider that g belongs to $\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)$ defined by (5). Then, the bias term $\|g-g_m\|^2$ is bounded by

$$(2\pi)^{-1}(D_m+1)^{2\alpha}\exp\{-2\rho(D_m+1)^\nu\}\sum_{|j|\geq D_m+1}|g^*(j)|^2|j|^{2\alpha}\exp\{2\rho|j|^\nu\}$$

and consequently

$$\|g-g_m\|^2 \;\leq\; (2\pi)^{-1}A_\alpha(D_m+1)^{2\alpha}\exp\{-2\rho(D_m+1)^\nu\}. \tag{13}$$

Let us come to the variance term $\sum_{|j|\leq D_m}\mathrm{Var}[\nu_n(\varphi_j)]$. If the Z_i's are independent random variables, then

$$\mathrm{Var}[\nu_n(\varphi_j)] \;=\; \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\left[V_{\varphi_j}(Z_i)\right]=\frac{1}{n}\mathrm{Var}\left[V_{\varphi_j}(Z_1)\right]$$

which is bounded by

$$\frac{1}{n}\mathbb{E}\left|\frac{\varphi_j(Z_1)}{f_\varepsilon^*(-j)}\right|^2\leq\frac{1}{2\pi n|f_\varepsilon^*(j)|^2}.$$

Consequently, we have the following bound for the variance

$$\sum_{|j|\leq D_m}\mathrm{Var}[\nu_n(\varphi_j)]\leq\Delta_1(m)/(2\pi n) \tag{14}$$

with $\Delta_1(m)$ defined by

$$\Delta_1(m)=\sum_{|j|\leq D_m}|f_\varepsilon^*(j)|^{-2}. \tag{15}$$

Under **A6**, we get the bound

$$\Delta_1(m)\leq\frac{2}{A_0^2}\int_0^{D_m+1}(1+x^2)^\gamma\exp\{2Bx^\delta\}dx,$$

and therefore, arguing as in Comte and Taupin (2003), we infer that

$$\sum_{|j|\leq D_m}\mathrm{Var}[\nu_n(\varphi_j)]\leq\lambda_1\frac{(D_m+1)^{(2\gamma+1-\delta)}\exp\left\{2B(D_m+1)^\delta\right\}}{n} \tag{16}$$

with $\lambda_1=\lambda_1(\gamma,A_0,B,\delta)$ defined by

$$\lambda_1=\frac{2^\gamma R(B,\delta)}{\pi A_0^2}, \tag{17}$$

and $R(B,\delta)$ given by

$$R(B,\delta)=\mathbf{1}_{\{\delta=0\}}+\frac{1}{2B(\delta\wedge 1)}\mathbf{1}_{\{\delta>0\}}=\begin{cases}1 & \text{if } \delta=0\\ 1/(2B\delta) & \text{if } 0<\delta\leq 1\\ 1/(2B) & \text{if } \delta>1.\end{cases} \tag{18}$$

Consequently, combining (12) and (16) we get that

$$\mathbb{E}(\|g-\hat{g}_m\|^2)\leq\|g-g_m\|^2+2\lambda_1\frac{(D_m+1)^{(2\gamma+1-\delta)}\exp\left\{2B(D_m+1)^\delta\right\}}{n}. \tag{19}$$

Now, when the Z_i's are absolutely regular variables, we apply (38) (See Theorem 4.1 in Section 4.2), which is the Delyon's (1990) covariance Inequality, successfully exploited by Viennet (1997). Hence Inequality (19) becomes

$$\mathbb{E}(\|g-\hat{g}_m\|^2) \leq \|g-g_m\|^2 + 8(\sum_k \beta_k)\lambda_1 \frac{(D_m+1)^{(2\gamma+1-\delta)}\exp\{2B(D_m+1)^\delta\}}{n}. \tag{20}$$

Let us now study this risk $\mathbb{E}(\|g-\hat{g}_m\|^2)$, starting by the study of a special case. Assume for a moment that g belongs to $\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)$ with $\nu=0$ and that f_ε satisfies Assumption **A6** with $\delta=0$. Then according (16), $\sum_{|j|\leq D_m}\mathrm{Var}[\nu_n(\varphi_j)]$ has the order $D_m^{1+2\gamma}/n$, and according to (13), $\|g-g_m\|^2$ has the order $D_m^{-2\alpha}$. Those bounds lead to choose the space $S_{\breve{m}}$ with dimension $D_{\breve{m}}=n^{1/(2\alpha+2\gamma+1)}$. Therefore the estimator $g_{\breve{m}}$ attains the rate $n^{-2\alpha/(2\alpha+2\gamma+1)}$, which is known to be the optimal rate for real variables (see Fan (1991a)). One can see that this choice of $D_{\breve{m}}$ depends on the unknown smoothness of g, α. This motivates us to complete the procedure by an automatic selection of the space via a penalization of the contrast. Note that bounds (13) and (16), respectively for the square bias term and for the variance are the same (up to constants) as the analogue for real random variables in Comte and Taupin (2003), Butucea (2003) and Butucea and Tsybakov (2003), where the trade-off has been studied and shown to provide optimal rates. Combining those facts, with the results in Efromovich (1997) we may infer that those rates are also optimal for circular density deconvolution. Table 1 details the optimal choices of D_m and the associated rates in function of the smoothness of the errors ε and of the function g, the first one being known and not the second one.

g \ f_ε	$\delta=0$	$\delta>0$
$\nu=0$	$D_{\breve{m}}+1=n^{1/(2\alpha+2\gamma+1)}$ rate $n^{-2\alpha/(2\alpha+2\gamma+1)}$	$D_{\breve{m}}+1=[\ln(n)/(2B+1)]^{1/\delta}$ rate $(\ln(n))^{-2\alpha/\delta}$
$\nu>0$	$D_{\breve{m}}+1=[\ln(n)/2\rho]^{1/\nu}$ rate $\dfrac{\ln(n)^{(2\gamma+1)/\nu}}{n}$	$D_{\breve{m}}$ and the rate depend on the integer k such that $\frac{\nu}{\delta}\wedge\frac{\delta}{\nu}\in[\frac{k}{k+1},\frac{k+1}{k+2}[$.

Table 1. Optimal choice of the dimension ($D_{\breve{m}}$) and obtained rates.

Remark 2.1. When $\delta>0$ and $\nu>0$, then $D_{\breve{m}}$ has not an explicit expression but is solution of Equation (15) in Comte and Taupin (2003). In fact it depends on the integer k such that ν/δ or δ/ν (depending on which one is less than one) belongs to $[k/(k+1),(k+1)/(k+2)[$. The main point to keep in mind here is that the rate is always faster than any powers of $\ln(n)$ and even reaches the order $\ln(n)^b n^{-\rho/(B+\rho)}$ with $b=[-2\alpha B+(2\gamma-\nu+1)\rho]/[\nu(B+\rho)]$ when $\nu=\delta$.

3. The results

The model selection procedure is required when the optimal choice of the space S_m (here the choice of D_m) depends on the unknown function g. We aim at finding the best model $\hat{m}$ in $\mathcal{M}_n$, based on the data and not on prior information on g, such that the risk of the resulting estimator is almost as good as the risk of the best estimator in the family. The model selection is performed in an automatic way, using the following penalized criteria

$$\tilde{g} = \hat{g}_{\hat{m}} \text{ with } \hat{m} = \text{ Argmin }_{m\in\mathcal{M}_n} \left[\gamma_n(\hat{g}_m) + \text{ pen}(m)\right]. \tag{21}$$

The quantity pen is a penalty function, based on the observations, that we would like to find such that

$$\mathbb{E} \parallel \tilde{g} - g \parallel^2 \leq \inf_{m\in\mathcal{M}_n} \left[\parallel g - g_m \parallel^2 + 2\lambda_1 \frac{(D_m+1)^{2\gamma+1-\delta}\exp\{2B(D_m+1)^\delta\}}{n}\right]. \tag{22}$$

in the independent case or such that
(23)
$$\mathbb{E} \parallel \tilde{g} - g \parallel^2 \leq \inf_{m\in\mathcal{M}_n} \left[\parallel g - g_m \parallel^2 + 8\lambda_1 (\sum_k \beta_k) \frac{(D_m+1)^{2\gamma+1-\delta}\exp\{2B(D_m+1)^\delta\}}{n}\right],$$

in the β-mixing case, where $\lambda_1 = \lambda_1(\gamma, A_0, B, \delta)$ is given by (17).

We do not reach (22) in all cases but state the following theorem.

Theorem 3.1. *Consider the model (1) under Assumptions* **A1-A3**, **A4–A6** *and the collection of estimators $\hat{g}_m$ defined by (7) for $1 \leq m \leq m_n \leq n^{1/(2\gamma+1)}$.*
1) *Let $\delta > 1$ and $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (21) with*

$$\text{pen}(m) = \kappa\lambda_1 \frac{(D_m+1)^{2\gamma-\delta+1}\exp\{2B(D_m+1)^\delta\}}{n},$$

where κ is some universal numerical constant, and $\lambda_1 = \lambda_1(\gamma, A_0, B, \delta)$ and $R(B, \delta)$ are defined by (17) and (18).
2) *Let $0 \leq \delta \leq 1/3$, and $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (21) with*

$$\text{pen}(m) = \kappa \left(\lambda_1 + \frac{BR^{1/2}(2B,\delta)\|f_\varepsilon\|}{A_0^2}\right) \frac{(D_m+1)^{2\gamma-\delta+1}\exp\{2B(D_m+1)^\delta\}}{n},$$

for some universal numerical constant κ.
3) *Let $1/3 < \delta \leq 1$ and $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (21) with*

$$\text{pen}(m) = \kappa \left(\lambda_1 + \frac{BR^{1/2}(2B,\delta)\|f_\varepsilon\|}{A_0^2}\right) \frac{(D_m+1)^{2\gamma+1/2+\delta/2}\exp\{2B(D_m+1)^\delta\}}{n}$$

where κ is some universal numerical constant.
Then in these three cases, $\tilde{g}$ satisfies

$$\mathbb{E}(\|g - \tilde{g}\|^2) \leq K \inf_{m\in\{1,\ldots,m_n\}} [\|g - g_m\|^2 + \text{pen}(m)] + \frac{c}{n}, \tag{24}$$

where K and c are constants depending on f_ε, B, δ, γ.

Remark 3.1. The construction of the estimator $\tilde{g}$ does not require any knowledge on the unknwon density g and its rate of convergence is nearly the rate of the best estimator over the family $(\hat{g}_m)_{m\in\mathcal{M}_n}$. Furthermore, its rate is easily deduced from (24) as soon as g belongs to some space $\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)$ defined by (5), using Inequality (13) with the important advantage that the procedure reach automatically the rate without any prior information on α, ν and ρ.

Remark 3.2. In the first two cases, $\delta > 1$, and $0 \leq \delta \leq 1/3$, Inequality (22) holds up to constants, since the penalty function $\mathrm{pen}(m)$ is of the variance order. It follows that in both cases, the resulting rates are optimal. In the last case, namely when $1/3 < \delta \leq 1$, the penalty function $\mathrm{pen}(m)$ is not exactly of the variance order, but of order $D_m^{2\gamma+\delta/2+1/2}\exp(2B(D_m+1)^\delta)/n$, with a loss of order $D_m^{(3\delta-1)/2}$. This loss in the variance term has no consequence on the rate when the rate is determined by the squared bias term. It follows that the rate remains optimal if the bias $\|g-g_m\|^2$ is the dominating term in the trade-off between $\|g-g_m\|^2$ and $\mathrm{pen}(m)$, which happens for instance, when $\nu = 0$. When $\mathrm{pen}(m)$ is the dominating term in the trade-off between $\|g-g_m\|^2$ and $\mathrm{pen}(m)$, there is a loss of order $D_m^{(3\delta-1)/2}$. According to Remark 2.1, this happens in cases where D_m is of logarithmic order and consequently the loss is logarithmic, when the rate is faster than logarithmic: therefore the loss happens only in cases when it can be seen as negligible.

Is this logarithmic loss avoidable? Tsybakov (2000) partially answers the question. Indeed, in the particular case $\delta = \nu = 1$, he shows that, for general inverse problems with real variables, a logarithmic loss in adaptivity of order $\ln(n)^{B/(\rho+B)}$ appears and is not avoidable. And, when $\nu = \delta = 1$, the optimal $D_{\breve{m}} + 1$ equals $\ln(n)^{B/(\rho+B)}$ and therefore the loss in adaptivity of our estimator is of order $\ln(n)^{B/(\rho+B)}$, which is the same as the loss exhibited by Tsybakov (2000) for real variables. It follows that, at least in this case, the logarithmic loss seems not avoidable and consequently, the rate of our estimator seems optimal among adaptive estimators .

Consequently, the adaptive procedure remains a good strategy, even if in that case, it implies a small loss in the rate as a price to pay for ignoring how smooth the unknown function g is. This strategy is all the more relevant that it solves almost optimally the problem in cases where the optimal $D_{\breve{m}}$ and thus the rate are difficult to compute.

Remark 3.3. The case $\delta > 1$ requires a special comment. In the case of real random variables, Comte and Taupin (2003) exhibit an adaptive penalized constrast estimator having a nearly optimal rate of convergence when $\delta > 1$. This fact comes from a logarithmic loss in the penalty function compared to the expected variance order. Whereas, in the case of circular random variables, the adaptive estimator reaches the optimal rate even when $\delta > 1$. This fact is even more noteworthy that the case $\delta > 1$ contains the case of Gaussian errors ($\delta = 2$).

In the mixing case, we establish the following Corollary.

Corollary 3.1. *Consider the model (1) under* **A1**, **A2**, **A3'**, **A4**–**A6**, *and the collection of estimators $\hat{g}_m$ defined by (7) for $1 \leq m \leq m_n$ where m_n is such that* $\mathrm{pen}(m_n)$ *is bounded by some constant. Assume moreover that the Z_i's are arithmetically β-mixing, that is $\beta_k \leq Ck^{-(1+\theta)}$, for all $k \in \mathbb{N}$, with $\theta > 3$.*

1) *Let* $0 \leq \delta \leq 1/3$ *or* $\delta > 1$. *Let* $\tilde{g} = \hat{g}_{\hat{m}}$ *be defined by (21) with*

$$\mathrm{pen}(m) = C_1(\sum_{k\in\mathbb{N}}\beta_k)\frac{(D_m+1)^{2\gamma+1-\delta}\exp\{2B(D_m+1)^\delta\}}{n},$$

for some constant C_1 *depending on* $\gamma, A_0, B, \delta, \sum_k \beta_k$, $\sum_k k\beta_k$ *and* $\|h\|_\infty$.
2) *Let* $1/3 < \delta \leq 1$ *and let* $\tilde{g} = \hat{g}_{\hat{m}}$ *be defined by (21) with*

$$\mathrm{pen}(m) = C_2(\sum_{k\in\mathbb{N}}\beta_k)\frac{(D_m+1)^{2\gamma+1/2+\delta/2}\exp\{2B(D_m+1)^\delta\}}{n}$$

for some constant C_2 *depending on* $\gamma, A_0, B, \delta, \sum_k \beta_k$, $\sum_k k\beta_k$ *and* $\|h\|_\infty$.
Then in these two cases, $\tilde{g}$ *satisfies*

$$\mathbb{E}(\|g-\tilde{g}\|^2) \leq K \inf_{m\in\{1,\ldots,m_n\}}\left[\|g-g_m\|^2 + \mathrm{pen}(m)\right] + \frac{c}{n}, \tag{25}$$

where K *and* c *are constants depending on* f_ε *and on the mixing coefficients.*

Obviously, the previous result holds in the geometrical β-mixing case, that is when $\beta_k \leq C\exp\{-\theta k\}$, for all $k \in \mathbb{N}$ with no condition on the rate θ.

The result given in Corollary 3.1 is analogous to the results obtained in the independent case, up to the constants. Nevertheless, the above corollary is mainly a result of robustness of the estimators since the penalty found in the mixing framework contains unknown coefficients, namely $\|h\|_\infty$, $\sum_k \beta_k$ and $\sum_k k\beta_k$.

4. Proofs

4.1. **Proof of Theorem 3.1 : the i.i.d. case.**

The following decompositions illustrate that the deconvolution problem can be treated using a classical model selection scheme. By definition, $\tilde{g}$ satisfies that for all $m \in \mathcal{M}_n$,

$$\gamma_n(\tilde{g}) + \mathrm{pen}(\hat{m}) \leq \gamma_n(g_m) + \mathrm{pen}(m).$$

Therefore, by applying (10), we get

$$\| \tilde{g} - g \|^2 \quad \leq \quad \|g-g_m\|^2 + 2\nu_n(\tilde{g}-g_m) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}). \tag{26}$$

Next, we use that if $t = t_1 + t_2$ with t_1 in S_m and t_2 in $S_{m'}$, then t belongs to $S_{m\vee m'}$. Denoting by $B_{m,m'}(0,1)$ the set

$$B_{m,m'}(0,1) = \{t \in S_{m\vee m'} \;/\; \|t\| = 1\},$$

we get that

$$|\nu_n(\tilde{g}-g_m)| \leq \|\tilde{g}-g_m\| \sup_{t\in B_{m,\hat{m}}(0,1)} |\nu_n(t)|.$$

Consequently, by using that $2ab \leq x^{-1}a^2 + xb^2$, we find

$$\frac{1}{2}\|\tilde{g}-g\|^2 \quad \leq \quad \frac{3}{2}\|g-g_m\|^2 + \mathrm{pen}(m) + 4\sup_{t\in B_{m,\hat{m}}(0,1)}\nu_n^2(t) - \mathrm{pen}(\hat{m}).$$

For some positive function $p(m,m')$ such that $4p(m,m') \leq \mathrm{pen}(m)+\mathrm{pen}(m')$, we get that

$$\frac{1}{2} \| \tilde{g} - g \|^2 \quad \leq \quad \frac{3}{2} \| g - g_m \|^2 + 2\mathrm{pen}(m) + 4\sum_{m'\in\mathcal{M}_n} W_n(m') \tag{27}$$

where $W_n(m')$ is defined as

$$W_n(m') := \left[\left(\sup_{t\in B_{m,m'}(0,1)} |\nu_n(t)|\right)^2 - p(m,m')\right]_+ . \tag{28}$$

The main point of the proof lies in finding $p(m,m')$ such that

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_n(m')) \leq C/n, \tag{29}$$

where C is a constant. Combining (27) and (29) we infer that, for all m in $\mathcal{M}_n$,

$$\mathbb{E}\|g-\tilde{g}\|^2 \leq 3\|g-g_m\|^2 + \frac{12C}{n} + 4\text{pen}(m),$$

which can also be written

$$\mathbb{E}\|g-\tilde{g}\|^2 \leq K \inf_{m\in\mathcal{M}_n} \left[\|g-g_m\|^2 + \text{pen}(m)\right] + \frac{12C}{n}, \tag{30}$$

where $K = 4$ suits. If $\text{pen}(m)$ has the same order as the variance order $\Delta_1(m)/n$, then equation (30) guarantees an automatic trade-off between the squared bias term $\|g-g_m\|^2$ and the variance term, up to some multiplicative constant and (22) follows. It remains thus to find $p(m,m')$ such that (29) holds. In the i.i.d. case, (29) follows by applying the following version of Talagrand's Inequality given below.

Lemma 4.1. *Let $Y_1,\dots,Y_n$ be i.i.d. random variables and $\bar{\nu_n}(f)$ be defined by $\bar{\nu_n}(f) = (1/n)\sum_{i=1}^n [f(Y_i) - \mathbb{E}(f(Y_i))]$ for f belonging to a countable class $\mathcal{F}$ of uniformly bounded measurable functions. Then*

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} |\bar{\nu_n}(f)|^2 - 2(4+\xi^2)H^2\right]_+ \leq \frac{6}{K_1}\left(\frac{v}{n}\exp\left\{-K_1\xi^2\frac{nH^2}{v}\right\} + \frac{4M_1^2}{K_1 n^2}\exp\left\{-\frac{K_1\xi}{\sqrt{2}}\frac{nH}{M_1}\right\}\right),$$

where K_1 is a universal constant,

$$\sup_{f\in\mathcal{F}}\|f\|_\infty \leq M_1, \quad \mathbb{E}\left(\sup_{f\in\mathcal{F}}|\bar{\nu_n}(f)|\right) \leq H, \quad \textit{and} \quad \sup_{f\in\mathcal{F}}\text{Var}(f(Y_1)) \leq v.$$

We apply Talagrand's Inequality (see Lemma 4.1) to the process $\bar{\nu_n}(V_t) = \nu_n(t)$ with $\mathcal{F} = B_{m,m'}(0,1)$. Indeed, by usual density arguments we can conclude that this result can be applied to this class of functions. Let us denote by $m^* = m\vee m'$ and by $H = H(m^*)$ a judicious bound for $\mathbb{E}\left(\sup_{t\in B_{m,m'}(0,1)}|\nu_n(t)|\right)$. From Definition (28) of $W_n(m')$, by taking $p(m,m') = 2(4+\xi^2)H^2$, we get that

$$\mathbb{E}(W_n(m') \leq \mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)} |\nu_n(t)|^2 - 2(4+\xi^2)H^2\right]_+ .$$

By Applying Lemma 4.1 to the above right hand side , we get the global bound

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_n(m') \leq K \sum_{m'\in\mathcal{M}_n} [I(m^*) + II(m^*)],$$

where $I(m^*)$ and $II(m^*)$ are defined by

$$I(m^*) = \frac{v}{n}\exp\left\{-K_1\xi^2\frac{nH^2}{v}\right\} \text{ and } II(m^*) = \frac{M_1^2}{n}\exp\left\{-\frac{K_1\xi}{\sqrt{2}}\frac{nH}{M_1}\right\},$$

with M_1, H and v such that

$$\sup_{t\in B_{m,m'}(0,1)}\|V_t\|_\infty \leq M_1,\ \mathbb{E}\left(\sup_{t\in B_{m,m'}(0,1)}|\nu_n(t)|\right) \leq H \text{ and } \sup_{t\in B_{m,m'}(0,1)}\mathrm{Var}(V_t(Z_1)) \leq v.$$

It remains thus to suitably choose M_1, H, v and also $\xi^2 = \xi^2(m,m')$ such that

$$\sum_{m'\in\mathcal{M}_n}[I(m^*)+II(m^*)] \leq C/n,$$

in the way that (29) holds, with $p(m,m') = 2(4+\xi^2)H^2(m^*)$ and $\mathrm{pen}(m) = 4p(m,m)$.

For t belonging to $B_{m,m'}(0,1)$, write $t = \sum_{|j|\leq D_{m^*}} a_j(t)\varphi_j$ with $\sum_{|j|\leq D_{m^*}} a_j(t)^2 = 1$. It follows, by applying (14), that

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)}|\nu_n(t)|\right] &\leq \mathbb{E}\left[\left(\sum_{|j|\leq D_{m^*}}(\nu_n(\varphi_j))^2\right)^{1/2}\right] \leq \left(\sum_{|j|\leq D_{m^*}}\mathrm{Var}(\nu_n(\varphi_j))\right)^{1/2} \\ &\leq \left(\frac{\Delta_1(m^*)}{2\pi n}\right)^{1/2},\end{aligned}$$

with $\Delta_1(m)$ defined by (15). Consequently, according to (16), we take

$$H^2 = H^2(m^*) = \lambda_1\frac{(D_{m^*}+1)^{2\gamma+1-\delta}\exp\{2B(D_{m^*}+1)^\delta\}}{n},$$

with $\lambda_1 = \lambda_1(\gamma, A_0, B, \delta)$ defined by (17). In the same manner,

$$\begin{aligned}\|V_t\|_\infty &\leq \left(\sum_{|j|\leq D_{m^*}} a_j(t)^2 \sum_{|j|\leq D_{m^*}}\left|\frac{\varphi_j}{f_\varepsilon^*(j)}\right|^2\right)^{1/2} \\ &\leq \left(\frac{1}{2\pi}\sum_{|j|\leq D_{m^*}}\left|\frac{1}{f_\varepsilon^*(j)}\right|^2\right)^{1/2}\end{aligned}$$

and therefore

$$\|V_t\|_\infty \leq \left(\frac{\Delta_1(m^*)}{2\pi}\right)^{1/2}. \tag{31}$$

with $\Delta_1(m)$ defined by (15). Consequently we take $M_1 = \sqrt{nH^2}$. Lastly, since $\mathrm{Var}(V_t(Z_1)) \leq \mathbb{E}|V_t(Z_1)|^2$ with

$$\mathbb{E}|V_t(Z_1)|^2 = \sum_{|j|\leq D_{m^*},|k|\leq D_{m^*}}\frac{a_j(t)\bar{a}_k(t)f_\varepsilon^*(j-k)g^*(j-k)}{2\pi f_\varepsilon^*(j)f_\varepsilon^*(-k)},$$

we get, by using that $\|g^*\|_\infty \leq 1$ and by applying Cauchy-Schwarz Inequality, that

$$\begin{aligned}
[\mathrm{Var}(V_t(Z_1))]^2 &\leq \frac{1}{(2\pi)^2} \sum_{|j|\leq D_m, |k|\leq D_m} \left| \frac{f_\varepsilon^*(j-k)}{f_\varepsilon^*(j) f_\varepsilon^*(-k)} \right|^2 \\
&\leq \frac{1}{(2\pi)^2} \left(\sum_{|j|\leq D_m, |k|\leq D_m} \frac{|f_\varepsilon^*(j-k)|^2}{|f_\varepsilon^*(j)|^4} \sum_{|j|\leq D_m, |k|\leq D_m} \frac{|f_\varepsilon^*(j-k)|^2}{|f_\varepsilon^*(-k)|^4} \right)^{1/2}.
\end{aligned}$$

Since $\sum_{k\in\mathbb{Z}} |f_\varepsilon^*(k)|^2 = 2\pi \|f_\varepsilon\|^2$, we find $[\mathrm{Var}(V_t(Z_1))]^2 \leq \|f_\varepsilon\|^2 \Delta_2(m^*)/(2\pi)$, with

$$\Delta_2(m) = \sum_{|j|\leq D_m} |f_\varepsilon^*(j)|^{-4}. \tag{32}$$

Clearly, the resulting choices of v are given by the bounds found for $\Delta_2(m)$. If $\delta > 1$, $\Delta_2(m)$ is bounded in the following way

$$\begin{aligned}
\Delta_2(m) &\leq 2A_0^{-4}(D_m+1)(D_m^2+1)^{2\gamma} \exp\{4BD_m^\delta\} \\
&\leq 2A_0^{-4}(D_m+1)^{4\gamma+1} \exp\{4BD_m^\delta\}.
\end{aligned}$$

It follows that, if $\delta > 1$ we take

$$v = \frac{\| f_\varepsilon \|}{A_0^2\sqrt{\pi}} (D_m+1)^{2\gamma+1/2} \exp\{2BD_m^\delta\}. \tag{33}$$

Now, if $\delta \leq 1$, then write that

$$\begin{aligned}
\Delta_2(m) &\leq \frac{2}{A_0^4} \int_0^{D_m+1} (x^2+1)^{4\gamma} \exp(4B|x|^\delta) dx \\
&\leq \frac{2R(2B,\delta)}{A_0^4} (D_m+1)^{4\gamma+1-\delta} \exp\{4B(D_m+1)^\delta\}
\end{aligned}$$

Consequently, if $\delta \leq 1$ we take

$$v = \frac{R^{1/2}(2B,\delta) \| f_\varepsilon \|}{A_0^2\sqrt{\pi}} (D_m+1)^{2\gamma+1/2-\delta/2} \exp\{2B(D_m+1)^\delta\}. \tag{34}$$

Let us study $\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_n(m'))$

1) Case $\delta > 1$. In that case, v is given by (33), $I(m^*)$ equals

$$\frac{\|f_\varepsilon\|}{\sqrt{\pi}A_0^2} \frac{(D_{m^*}+1)^{2\gamma+1/2}}{n} \exp\left[2BD_{m^*}^\delta - \frac{\sqrt{\pi}A_0^2\lambda_1}{\| f_\varepsilon \|} \xi^2 D_m^{*\,1/2-\delta} \exp\{2B[(D_{m^*}+1)^\delta - D_{m^*}^\delta]\} \right]$$

and therefore, for some constant K',

$$I(m^*) \leq K' \frac{D_{m^*}^{2\gamma+1/2}}{n} \exp\left\{ 2BD_{m^*}^\delta - \frac{\sqrt{\pi}A_0^2\lambda_1}{\| f_\varepsilon \|} \xi^2 \exp\{2^\delta B\delta D_{m^*}^{\delta-1}\} \right\}.$$

Choose $\xi^2 = 1$. For any m and m' in $\mathcal{M}_n$ such that D_{m^*} is great enough, the following inequality always holds

$$2BD_{m^*}^\delta - \frac{\sqrt{\pi}A_0^2\lambda_1}{\| f_\varepsilon \|} \exp\left(2^\delta B\delta D_{m^*}^{\delta-1}\right) < -(D_{m^*})^{1/2}.$$

It follows that there exists some constant c such that

$$\sum_{m'\in\mathcal{M}_n} I(m^*) \le c/n \text{ and } \sum_{m'\in\mathcal{M}_n} II(m^*) \le c/n.$$

This choice for ξ^2 ensures that $\sum_{m'\in\mathcal{M}_n}\mathbb{E}[W_n(m')] \le C/n$, for some constant C. Consequently we take

$$p(m,m') = 10\lambda_1 \frac{(D_m^*+1)^{2\gamma+1-\delta}\exp\{2B(D_m^*+1)^\delta\}}{n},$$

and

$$\mathrm{pen}(m) = 40\lambda_1\frac{(D_m+1)^{2\gamma+1-\delta}\exp\{2B(D_m+1)^\delta\}}{n}.$$

2) Case $0\le\delta\le 1/3$**.** Then v is given by (34), $\delta \le (1/2-\delta/2)$ and $I(m^*)$ equals

$$\frac{R^{1/2}(2B,\delta)\|f_\varepsilon\|}{A_0^2\sqrt{\pi}}\frac{(D_m^*+1)^{2\gamma+1/2-\delta/2}\exp\{2B(D_m^*+1)^\delta\}\exp\{-K''\xi^2 D_{m^*}^{(1/2-\delta/2)}\}}{n} \tag{35}$$

where $K''=(K_1\lambda_1A_0^2\sqrt{\pi})/[R^{1/2}(2B,\delta)\|f_\varepsilon\|]$. The choice $\xi^2=(4B)/(K'')$ ensures the convergence of $\sum_{m'\in\mathcal{M}_n} I(m^*)$. Indeed, if we denote by $\psi = 2\gamma+(1/2-\delta/2)$, by $\omega = (1/2-\delta/2)$, then for $a,b\ge 1$, we infer that

$$\begin{aligned}(a\vee b)^\psi&\exp\{2B(a\vee b)^\delta\}\exp\{-K''\xi^2(a\vee b)^\omega\}\\ &\le (a^\psi\exp\{2Ba^\delta\}+b^\psi\exp\{2Bb^\delta\})\exp\{-(K''\xi^2/2)(a^\omega+b^\omega)\}\\ &\le a^\psi\exp\{2Ba^\delta\}\exp\{-(K''\xi^2/2)a^\omega\}\exp\{-(K''\xi^2/2)b^\omega\}\\ &\qquad + b^\psi\exp\{2Bb^\delta\}\exp\{-(K''\xi^2/2)b^\omega)\}.\end{aligned} \tag{36}$$

When $0\le\delta\le 1/3$, $(1/2-\delta/2)\ge\delta$ and the function $a\mapsto a^\psi\exp\{2Ba^\delta\}\exp\{-(K''\xi^2/2)a^\omega\}$ is bounded on $\mathbb{R}^+$ by a constant, κ, only depending on γ, δ and K''. It follows that there exists some constant C, such that $\sum_{m'\in\mathcal{M}_n} I(m^*)\le C/n$. The same holds for $\sum_{m'\in\mathcal{M}_n} II(m^*)$ and consequently, (29) holds. In that case we choose

$$p(m,m') = 2(4+\xi^2)\lambda_1\frac{(D_m^*+1)^{2\gamma+1-\delta}\exp\{2B(D_m^*+1)^\delta\}}{n},$$

and

$$\mathrm{pen}(m) = 8(4+\xi^2)\lambda_1\frac{(D_m+1)^{2\gamma+1-\delta}\exp\{2B(D_m+1)^\delta\}}{n}.$$

3) Case $1/3<\delta\le 1$**.** In that case, $\delta > (1/2-\delta/2)$, v is given by (34) and $I(m^*)$ by (35). Bearing in mind Inequality (36) we choose $\xi^2=\xi^2(m,m')$ such that

$$2B(D_{m^*}+1)^\delta - \frac{\pi^{1/2}K_1A_0^2\lambda_1}{R^{1/2}(2B,\delta)\parallel f_\varepsilon\parallel}\xi^2(D_{m^*}+1)^\omega = -2B(D_{m^*}+1)^\delta$$

that is

$$\xi^2=\xi^2(m,m') = \frac{4BR^{1/2}(2B,\delta)\|f_\varepsilon\|(D_{m^*}+1)^{\delta-\omega}}{K_1A_0^2\lambda_1\sqrt{\pi}} = \frac{4BR^{1/2}(2B,\delta)\|f_\varepsilon\|(D_{m^*}+1)^{(3\delta-1)/2}}{K_1A_0^2\lambda_1\sqrt{\pi}}.$$

This choice for ξ^2 ensures that, for some constant c, $\sum_{m'\in\mathcal{M}_n} I(m^*)\le c/n$, that $\sum_{m'\in\mathcal{M}_n} II(m^*)\le c/n$ and consequently (29) is fulfilled. The result follows by taking

$$p(m,m') = 2(4+\xi^2(m,m'))\lambda_1(D_{m^*}+1)^{2\gamma+1-\delta}\exp(2B(D_{m^*}+1)^\delta)/n,$$

and

$$\mathrm{pen}(m) = 8(4+\xi^2(m,m))\lambda_1(D_m+1)^{2\gamma+1-\delta}\exp(2B(D_m+1)^{\delta})/n.$$

4.2. **Proof of Corollary 3.1 : the absolutely regular case.** We use below Delyon's (1990) covariance Inequality, successfully exploited by Viennet (1997) for partial sums of strictly stationary processes, that we start by recalling.

Theorem 4.1. *(Delyon (1990), Viennet (1997)) Let P be the distribution of Z_0 on a probability space $\mathcal{X}$, $\int f dP = \mathbb{E}_P(f)$ for any function f P-integrable. For $r \geq 2$, let $\mathcal{L}(r,\beta,P)$ be the set of functions $b : \mathcal{X} \to \mathbb{R}^+$ such that*

$$b = \sum_{l\geq 0}(l+1)^{r-2}b_l \text{ with } 0 \leq b_l \leq 1 \text{ and } \mathbb{E}_P(b_l) \leq \beta_l,$$

We define B_r as $B_r = \sum_{l\geq 0}(l+1)^{r-2}\beta_l$. Then for $1 \leq p < \infty$ and any function b in $\mathcal{L}(2,\beta,P)$,

$$\mathbb{E}_P(b^p) \leq pB_{p+1}, \tag{37}$$

as soon as $B_{p+1} < \infty$. The following result holds for a strictly stationary absolutely regular sequence, $(Z_i)_{i\in\mathbb{Z}}$, with β-mixing coefficients $(\beta_k)_{k\geq 0}$: if $B_2 < +\infty$, there exists $b \in \mathcal{L}(2,\beta,\infty)$ such that for any positive integer n and any measurable function $f \in \mathbb{L}_2(P)$, we have

$$\mathrm{Var}\left(\sum_{i=1}^{n} f(Z_i)\right) \leq 4n\mathbb{E}_P(bf^2) = 4n\int b(x)f^2(x)dP(x). \tag{38}$$

By applying Inequality (38), we state (20).

We also use Berbee's coupling Lemma extended to sequences (see Bryc's (1982) construction), to build approximating variables for the Z_i's. More precisely, we build variables $Z_i^\star$ such that if $n = 2p_nq_n + r_n$, $0 \leq r_n < q_n$, and $\ell = 0, \cdots, p_n - 1$

$$A_\ell^\star = (Z^\star_{2\ell q_n+1}, ..., Z^\star_{(2\ell+1)q_n}),\ B_\ell^\star = (Z^\star_{(2\ell+1)q_n+1}, ..., Z^\star_{(2\ell+2)q_n}),$$

and analogous definition without stars, then

- $A_\ell^\star$ and A_ℓ have the same law,
- $\mathbb{P}(A_\ell \neq A_\ell^\star) \leq \beta_{q_n}$,
- $A_\ell^\star$ and $(A_0, A_1, ..., A_{\ell-1}, A_0^\star, A_1^\star, \cdots, A_{\ell-1}^\star)$ are independent.

The blocks $B_\ell^\star$ are built in the same way. Without loss of generality and for sake of simplicity we assume that $r_n = 0$.

Starting from (26), and denoting by $\nu_n^\star$ the empirical contrast computed on the $Z_i^\star$, we write

$$\|\tilde g - g\|^2 \leq 2|\nu_n(\tilde g - g_m) - \nu_n^\star(\tilde g - g_m)| + 2|\nu_n^\star(\tilde g - g_m)| + \|g - g_m\|^2 + \mathrm{pen}(m) - \mathrm{pen}(\hat m)$$

and therefore arguing as for (27) in the independent case we get that

$$\begin{aligned}\frac{1}{2}\|\tilde g - g\|^2 &\leq \frac{3}{2}\|g - g_m\|^2 + 2\mathrm{pen}(m) + 4\left(\sup_{t\in B_{m,\hat m}(0,1)}|\nu_n^\star(t)|^2 - p(m,\hat m)\right)\\ &\quad +2|\nu_n(\tilde g - g_m) - \nu_n^\star(\tilde g - g_m)|,\end{aligned} \tag{39}$$

where $p(m,m')$ is a function such that $4p(m,m') \le \mathrm{pen}(m)+\mathrm{pen}(m')$ as previously. Now write that

$$\begin{aligned} 2|\nu_n(\tilde g - g_m) - \nu_n^\star(\tilde g - g_m)| &= \frac{2}{n}\left|\sum_{i=1}^n [V_{(\tilde g - g_m)}(Z_i) - V_{(\tilde g - g_m)}(Z_i^\star)]\right| \\ &\le \frac{4}{n}\|\tilde g - g_m\| \sup_{t\in B_{m_n}(0,1)} \|V_t\|_\infty \sum_{i=1}^n \mathbf{1}_{Z_i\neq Z_i^\star}, \end{aligned}$$

and apply (31) to get the Bound

$$2|\nu_n(\tilde g - g_m) - \nu_n^\star(\tilde g - g_m)| \le 4\|\tilde g - g_m\| \left(\frac{\Delta_1(m_n)}{2\pi}\right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{Z_i\neq Z_i^\star}\right).$$

It follows, again by using that $2ab \le x^{-1}a^2 + xb^2$, that

$$\begin{aligned} 2|\nu_n(\tilde g - g_m) - \nu_n^\star(\tilde g - g_m)| &\le \frac{1}{4}\|\tilde g - g_m\|^2 + \frac{16\Delta_1(m_n)}{2\pi}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{Z_i\neq Z_i^\star}\right)^2 \\ &\le \frac{3}{8}\|\tilde g - g\|^2 + \frac{3}{4}\|g - g_m\|^2 + \frac{16\Delta_1(m_n)}{2\pi}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{Z_i\neq Z_i^\star}\right)^2. \qquad (40) \end{aligned}$$

Let $W_n^\star(m')$ be defined as in (28) with $Z_i^\star$ replacing Z_i. By gathering (39) and (40) we get that

$$\frac{1}{8}\|\tilde g - g\|^2 \le \frac{9}{4}\|g - g_m\|^2 + 4\sum_{m'\in\mathcal{M}_n} W_n^\star(m') + 2\mathrm{pen}(m) + \frac{16\Delta_1(m_n)}{2\pi}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{Z_i\neq Z_i^\star}\right)^2.$$

Consequently

$$\|\tilde g - g\|^2 \le 18\|g - g_m\|^2 + 16\mathrm{pen}(m) + 32\sum_{m'\in\mathcal{M}_n} W_n^\star(m') + \frac{64\Delta_1(m_n)}{\pi}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{Z_i\neq Z_i^\star}\right)^2.$$

Therefore by using that $\mathbb{E}\left(n^{-1}\sum_{i=1}^n \mathbf{1}_{Z_i\neq Z_i^\star}\right)^2 \le \beta_{q_n}$, we infer that, for all $m\in\mathcal{M}_n$,

$$\mathbb{E}\|g - \tilde g\|^2 \le 18\|g - g_m\|^2 + 16\mathrm{pen}(m) + \frac{C_1 + 32C_2}{n},$$

provided that

$$\frac{64\Delta_1(m_n)}{\pi}\beta_{q_n} \le \frac{C_1}{n} \qquad (41)$$

and

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_n^\star(m')) \le \frac{C_2}{n}. \qquad (42)$$

In order to ensure that our estimators converge, we only consider models such that the penalty is bounded, that is, by using (16), such that $(D_{m_n}+1)^{2\gamma+1-\delta}\exp\{2B(D_{m_n}+1)^\delta\} \le n$. This implies that (41) is fulfilled as soon as

$$\beta_{q_n} \le C_1'/n^2. \qquad (43)$$

When the mixing is geometrical, (43) holds with no condition on θ by choosing $q_n = \ln(n)/(4\theta)$. When the mixing is arithmetical, we take $q_n = [n^c]$ with $c \in]0, 1/2[$ and (43) becomes $n^{-c(1+\theta)} \le \kappa n^{-2}$ which holds as soon as $\theta > 3$. We now come to the proof of (42), and study $W_n^\star(m')$ which must be split into two terms $\frac{1}{2}(W_{,1}^\star(m') + W_{n,2}^\star(m'))$ involving respectively the odd and even blocks and which are of the same type. More precisely $W_{n,1}^\star(m')$ and $W_{n,2}^\star(m')$ are given, for $k = 1, 2$ by

$$W_{n,k}^\star(m') \;=\; \left[\left(\sup_{t\in B_{m,m'}(0,1)} \left|\frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\left(V_t(Z^\star_{(2\ell+k-1)q_n+i}) - \langle t, g\rangle\right)\right|\right)^2 - p_1(m, m')\right]_+$$

We only study $W_{n,1}^\star(m')$ and conclude for $W_{n,2}^\star(m')$ by using analogous arguments. Let us denote by $m^* = m \vee m'$. We apply Lemma 4.1 to the process $\nu_{n,1}^\star(t)$ defined by,

$$\nu_{n,1}^\star(t) \;=\; \frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\left(V_t(Z^\star_{2\ell q_n+i}) - \langle t, g\rangle\right) = \frac{1}{p_n}\sum_{\ell=1}^{p_n}\nu_{q_n,\ell}^\star(t), \tag{44}$$

considered as the sum of the p_n independent random variables $\nu_{q_n,\ell}^\star(t)$ defined as

$$\nu_{q_n,\ell}^\star(t) = (1/q_n)\sum_{j=1}^{q_n} V_t(Z^\star_{2\ell q_n+j}) - \langle t, g\rangle. \tag{45}$$

If we denote by $H^\star = H^\star(m')$ a bound for $\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}^\star(t)|\right]$, to be suitably chosen, and if we take $p_1(m, m') = 2(4 + \xi^2)(H^\star)^2$, we get that

$$\mathbb{E}(W_{n,1}^\star(m')) \le \mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}^\star(t)|^2 - 2(4 + \xi^2)(H^\star)^2\right]_+ .$$

Apply Lemma 4.1 to the above right hand side to obtain the global bound

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_{n,1}^\star(m')) \le K \sum_{m'\in\mathcal{M}_n} [I^\star(m^*) + II^\star(m^*)],$$

with $I^\star(m^*)$ and $II^\star(m^*)$ defined by

$$I^\star(m^*) = \frac{v^\star}{p_n}\exp\left\{-K_1\xi^2\frac{p_n(H^\star)^2}{v^\star}\right\} \text{ and } II(m^*) = \frac{(M_1^\star)^2}{p_n^2}\exp\left\{-\frac{K_1\xi}{\sqrt{2}}\frac{p_nH^\star}{M_1^\star}\right\},$$

with $M_1^\star$, $H^\star$ and $v^\star$ such that

$$\sup_{t\in B_{m,m'}(0,1)} \| \nu_{q_n,\ell}^\star(t) \|_\infty \le M_1^\star, \quad \mathbb{E}(\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}^\star(t)|) \le H^\star, \quad \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu_{q_n,\ell}^\star(t)) \le v^\star.$$

As in the independent case, it remains thus to suitably choose $M_1^\star$, $H^\star$, $v^\star$ and also $\xi^2 = \xi^2(m, m')$ such that

$$\sum_{m'\in\mathcal{M}_n} [I^\star(m^*) + II^\star(m^*)] \le C/n.$$

in the way that (42) holds, with $p(m, m') = 2(4 + \xi^2)(H^\star)^2(m^*)$ and $\mathrm{pen}(m) = 4p(m, m)$.

Clearly, $M_1^\star = 2M_1$. By applying Inequalities (37) and (38) for $p=1$, we propose to take $(H^\star)^2$ given by

$$(H^\star)^2 = (H^\star(m^*))^2 = \frac{4\lambda_1(\sum_k \beta_k)(D_{m^*}+1)^{2\gamma+1-\delta}\exp\{2B(D_{m^*}+1)^\delta\}}{n},$$

where $\lambda_1 = \lambda_1(\gamma, A_0, B, \delta)$ is still given by (17). By using again Inequality (38) and arguing as in the independent case,

$$\begin{aligned}\left[\sup_t \mathrm{Var}(\nu^\star_{q_n,\ell}(t))\right]^2 &\le \left(\frac{2}{\pi q_n}\right)^2 \sum_{|j|\le D_{m^*}, |k|\le D_{m^*}} \left|\frac{(bh)^*(j-k)}{f^*(j)f^*(-k)}\right|^2 \\ &\le \left(\frac{2}{\pi q_n}\right)^2 \sum_{|k|\le D_{m^*}} |(bh)^*(k)|^2 \Delta_2(m^*),\end{aligned}$$

where $\Delta_2(m)$ is defined by (32). Then, write

$$\sum_{k\in\mathbb{Z}} |(bh)^*(k)|^2 = 2\pi\|bh\|^2 \le 2\pi\|h\|_\infty \int b^2(x)h(x)d\mu_\pi(x) \le 4\pi\|h\|_\infty \sum_{k\in\mathbb{N}}(1+k)\beta_k.$$

It follows that for $\delta \le 1$, we take

$$v^\star = \frac{4\|h\|_\infty^{1/2} R^{1/2}(2B,\delta)\left(\sum_{k\in\mathbb{N}}(1+k)\beta_k\right)^{1/2}}{A_0^2\sqrt{\pi}} \frac{(D_{m^*}+1)^{2\gamma+1-\delta}\exp\{2B(D_{m^*}+1)^\delta\}}{q_n}, \tag{46}$$

and for $\delta > 1$ we take

$$v^\star = \frac{4\|h\|_\infty^{1/2}\left(\sum_{k\in\mathbb{N}}(1+k)\beta_k\right)^{1/2}}{A_0^2\sqrt{\pi}} \frac{(D_{m^*}+1)^{2\gamma+1/2}\exp\{2BD_{m^*}^\delta\}}{q_n}. \tag{47}$$

Now, by taking $p_1(m,m') = 2(4+\xi^2)(H^\star)^2$ we get the global bound

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W^\star_{n,1}(m')) \le K \sum_{m'\in\mathcal{M}_n} [I^\star(m^*) + II^\star(m^*)],$$

with $m^* = m \vee m'$, where $I^\star(m^*)$ has the same order as the analogous term $I(m^*)$ in the independent case and is bounded in the same manner, and where $II^\star(m^*)$ is defined by

$$II^\star(m^*) = \frac{\lambda_1^2 q_n^2 (D_{m^*}+1)^{4\gamma+2-2\delta}\exp\{4B(D_{m^*}+1)^\delta\}}{n^2} \exp\left\{-\frac{K_1\xi\left(\sum_k\beta_k\right)^{1/2}}{\sqrt{2}}\frac{\sqrt{n}}{q_n}\right\}.$$

The main difference lies in the study of the second term $II^\star(m^*)$. When the mixing is geometrical then $q_n = \ln(n)/(4\theta)$ ensures that

$$\frac{q_n^2}{n^2}\sum_{m'\in\mathcal{M}_n}(D_{m^*}+1)^{2\gamma+1-\delta}\exp\{2B(D_{m^*}+1)^\delta\}\exp\left\{-\frac{K_1\xi(\sum_k\beta_k)^{1/2}}{\sqrt{2}}\frac{\sqrt{n}}{q_n}\right\} \le C/n. \tag{48}$$

When the mixing is arithmetical, then $q_n = [n^c]$, with c in $]0,1/2[$ ensures that (48) holds and consequently

$$\sum_{m'\in\mathcal{M}_n} II^\star(m^*) \le C/n.$$

Finally $\sum_{m'\in\mathcal{M}_n}\mathbb{E}[W_n^\star(m')]\leq 2\sum_{m'\in\mathcal{M}_n}\mathbb{E}[W_{n,1}^\star(m')+W_{n,2}^\star(m')]\leq C/n$, if

$$p_1(m,m')=p_2(m,m')=8\lambda_1(4+\xi^2(m,m'))(\sum_k\beta_k)D_{m^*}^{2\gamma+1-\delta}\exp\{2B(\pi D_{m^*})^\delta\}/n,$$

with $\xi^2(m,m')=1$ if $\delta>1$, $\xi^2(m,m')=4B/K^{(3)}$, and

$$K^{(3)}=\frac{\sqrt{\pi}K_1A_0^2\sum_k\beta_k}{4\|h\|_\infty^{1/2}R^{1/2}(2B,\delta)\left(\sum_k(1+k)\beta_k\right)^{1/2}}$$

if $0\leq\delta\leq 1/3$ and $\xi^2(m,m')=(4B/K^{(3)})(D_{m^*}+1)^{(3\delta-1)/2}$ if $1/3<\delta\leq 1$. The result follows by choosing $p(m,m')=2p_1(m,m')+2p_2(m,m')$ and

$$\text{pen}(m)=128\lambda_1(4+\xi^2(m,m))(\sum_k\beta_k)D_m^{2\gamma+1-\delta}\exp\{2B(\pi D_m)^\delta\}/n.$$

References

[1] Barron, A.R., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields* **113**, 301-413.

[2] Berbee, H.C.P (1979). *Random walks with stationary increments and renewal theory.* Cent. Math. Tracts, Amsterdam.

[3] Bryc, W. (1982). On the approximation theorem of Berkes and Philipp. *Demonstratio Mathematica* **15**, 807-815.

[4] Butucea, C. (2003). Deconvolution of supersmooth densities with smooth noise. Submitted paper.

[5] Butucea, C. and Tsybakov, A.B. (2003). Fast asymptotics in density deconvolution. Manuscript.

[6] Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184-1186.

[7] Cator, E. A. (2001). Deconvolution with arbitrarily smooth kernels. *Stat. Probab. Lett.* **54**, 205-214.

[8] Comte, F. and Taupin, M.-L.(2003). Penalized contrast estimator for density deconvolution with mixing variables. Preprint 2003-2 MAP5, http://www.math-info.univ-paris5.fr/map5/publis/titres03.html.

[9] Delyon, B. (1990). Limit theorem for mixing processes. *Tech. Report IRISA* Rennes **1**, 546.

[10] Devroye, L.(1989). Consistent deconvolution in density estimation. *Canad. J. Statist.* **17**, 235-239.

[11] Efromovich, S. (1997). Density estimation in the case of supersmooth measurement errors. *J. Amer. Statist. Assoc.* **92**, 526-535.

[12] Fan, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.* **19**, 1257-1272.

[13] Fan, J. (1991b). Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1**, 541-551.

[14] Fan, J. and Koo J.-Y. (2002). Wavelet deconvolution. *IEEE TransacT. on Information Theory* **48**, 734-747.

[15] Goldenshluger, A. (1999) On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5** 907-926.

[16] Goldenshluger, A. (2002). Density Deconvolution in the Circular Structural Model. *J. Multivariate Anal.* **81** 360-375.

[17] Koo, J. Y. (1999). Logspline deconvolution in Besov spaces. *Scand. J. of Statistics* **26** 73-86.

[18] Liu, M.C. and Taylor, R.L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* **17**, 427-438.

[19] Mardia K. V. (1972). *Statistics of directional Data*, New-York: Academic Press.

[20] Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory* **37**, 1105-1115.

[21] Masry, E. (1993a). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Process. Appl.* **47**, 53-74.

[22] Masry, E. (1993b). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *J. Multivariate Anal.* **44**, 47-68.

[23] Pensky, M. (2002). Density deconvolution based on wavelets with bounded supports. *Stat. Probab. Lett.* **56** 261-269.
[24] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**, 6, 2033-2053.
[25] Rudin, W. (1962). *Fourier Analysis on groups* Interscience Publishers, New-York, Wiley and Sons.
[26] Stefansky, L. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Letters* **9**, 229-235.
[27] Stefansky, L. and Carroll, R.J. (1990). Deconvolution kernel density estimators. *Statistics* **21**, 169-184.
[28] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126**, 505-563.
[29] Taylor, R.L. and Zhang, H.M. (1990). On strongly consistent non-parametric density estimator for deconvolution problem. *Comm. Statist. Theory Methods* **19**, 3325-3342.
[30] Tsybakov,A.B. (2000) On the best rate of adaptive estimation in some inverse problems. *C.R. Acad. Sci. Paris, Série 1*, **330**, 835-840.
[31] Viennet, G.(1997). Inequalities for absolutely regular sequences: application to density estimation. *Probab. Th. Rel. Fields* **107**, 467-492.
[32] Youndjé, E. and Wells, M. T. (2002). Least squares cross-validation for the kernel deconvolution density estimator. *C. R. Acad. Sci. Paris, Ser. I* **334**,509-513.
[33] Zhang, C.H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806-831.