



HAL
open science

New adaptive strategies for nonparametric estimation in linear mixed models

Charlotte Dion

► **To cite this version:**

Charlotte Dion. New adaptive strategies for nonparametric estimation in linear mixed models. 2013.
hal-00906379v1

HAL Id: hal-00906379

<https://hal.science/hal-00906379v1>

Preprint submitted on 19 Nov 2013 (v1), last revised 20 Mar 2014 (v2)

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.

New adaptive strategies for nonparametric estimation in linear mixed models

Charlotte Dion^{(1),(2)}

(1)*LJK, UMR CNRS 5224, Université Joseph Fourier, 51 rue des Mathématiques,
38041 Grenoble*

(2)*MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité,
45 rue des Saints Pères, 75006 Paris*

Abstract

This paper surveys new estimators of the density of a random effect in linear mixed-effects models. Data are contaminated by random noise, and we do not observe directly the random effect of interest. The density of the noise is supposed to be known, without assumption on its regularity. However it can also be estimated. We first propose an adaptive nonparametric deconvolution estimation based on a selection method set up in Goldenshluger and Lepski (2011). Then we propose an estimator based on a simpler model selection devised by contrast penalization. For both of them, non-asymptotic L_2 -risk bounds are established implying estimation rates, much better than the expected deconvolution ones. Finally the two data-driven strategies are evaluated on simulations and compared with previous proposals.

Keywords. Deconvolution, linear mixed models, model selection, nonparametric estimation.

1 Introduction

In order to analyze repeated measures in time for individuals with the same behavior, we focus on a linear mixed-effects model. The observation value at time t_j with $j \in \{0, \dots, J\}$ for individual k , $k \in \{1, \dots, N\}$, denoted $Y_{k,j}$, follows the model

$$Y_{k,j} = \alpha_k + \beta_k t_j + \epsilon_{k,j}, \quad (1)$$

where the coefficients α_k and β_k are the random effects of regression and depend on the subject k . The times of observation t_j are known and equidistant with a time step Δ : $t_j = j\Delta$. The random variables $(\epsilon_{k,j})_{1 \leq k \leq N, 0 \leq j \leq J}$ are the measurement errors of zero mean. They are supposed independent and identically distributed (i.i.d.) with common density f_ϵ . We suppose α_k i.i.d. with density f_α and β_k i.i.d. with density f_β . Moreover we assume that $(\epsilon_{k,j})_{1 \leq k \leq N, 0 \leq j \leq J}$ and $(\alpha_k, \beta_k)_{1 \leq k \leq N}$ are independent sequences.

Mixed models with random effects are often used, for example in pharmacokinetics. They describe both individual behavior and variability between individuals. The distribution of random effects is of special interest. It allows for example to describe the heterogeneity of the drug kinetics in the population of individuals. Mixed models have been widely studied, often with parametric strategies and Gaussian random effects and noise (Pinheiro and Bates, 2000, see). However it is not clear that this normality assumption of the random effects is truly satisfied in practice. The aim of this paper is to produce nonparametric estimation of the density of the random effects from the observations $Y_{k,j}$.

Several papers consider this problem. Wu and Zhu (2010) relax the Gaussian assumption proposing an orthogonality-based estimation of moments estimating the third and fourth moments of the random effects and errors. Komárek and Lesaffre (2008) suppose that the random effects are a mixture of Gaussian distributions, and estimate the weights of the mixture components using a penalized approach. Papageorgiou and Hinde (2012) propose semi-parametric density models of

random effects. But it is the estimation of the complete random effects density which should be appropriate. This is the framework of nonparametric estimation that is adopted in this paper.

Complete nonparametric estimation of f_β and f_α in model (1) has been little studied in the literature. The main reference is Comte and Samson (2012) who set up a nonparametric estimation with deconvolution ideas. Indeed, model (1) can be seen as a measurement error model. Deconvolution methods are various and have been initiated by the kernel study of Fan (1991), followed by adaptive strategies (Pensky and Vidakovic (1999), Comte *et al.* (2006), Delaigle and Gijbels (2004)) and optimality of rates (Butucea and Tsybakov, 2007). Except the penalized contrast estimator of Comte and Samson (2012), none of these previous estimators has been used in the mixed model framework. However, for all these methods, the noise density is usually assumed to be known with a certain regularity. The resulting rates are thus very specific and depend on the regularity of the noise and of the function under estimation. For example the noise is supposed to be ordinary smooth in Comte and Samson (2012).

This paper focuses on improving existing nonparametric strategies for the estimation of f_β , without any assumption on the noise density regularity. We focus on f_β and we refer to Comte *et al.* (2006) for f_α , noticing that for $t_0 = 0$ model (1) writes $Y_{k,0} = \alpha_k + \epsilon_{k,0}$ which is a standard convolution equation. We first propose an estimator which uses all the available observations. A Mean Integrated Squared Error (MISE) bound is computed and it shows that a bias-variance compromise must be performed. The particularity here is the selection model procedure: we adapt a method set up in Goldenshluger and Lepski (2011) for kernel estimators. Then we establish a non-asymptotic oracle risk bound using Talagrand's inequality. In a second time, an other estimator is proposed, built using the last and the first time of observations. If two times of observation are separated by a quite long time, enough information is available to obtain a good estimation. As in the first case, risk bound leads us to propose a model selection strategy. In that case a more classical and easy to implant strategy can be adopted, namely a penalized criterion (Massart, 2007; Birgé and Massart, 1998).

What is new is that our two strategies take advantage of both nonparametric and deconvolution ideas. Without adding any assumption on the noise regularity, the two estimators recover standard rates in density estimation for f_β (Stone, 1980; Donoho *et al.*, 1996) while deconvolution rates were expected. Precisely logarithmic speed of convergence implied by deconvolution methods is avoided. Therefore our two estimators have better rates.

This paper is organized as follows. We proceed with the construction of the estimators in Section 2.1 and 2.4. Bias-variance decompositions of the \mathbb{L}^2 -risk are proved (Propositions 1, 5) and lead us to propose adaptive strategies. The main results (Theorems 3, 7) prove that the resulting estimators are adaptive. A short Section 3 provides a comparison of our estimators with Comte and Samson (2012). Lastly, we illustrate the different methods by simulation experiments presented in Section 4. We briefly discuss therein the case of unknown noise density (see Section 4.3 and Section 6.8) which is also implemented for comparison. Proofs are gathered in Section 6.

2 Construction of two estimators, risk bounds and adaptive results

Let us introduce some notations. For two functions f and g in $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$, the scalar product is defined by $\langle f, g \rangle = \int_{\mathbb{R}} f(x)\overline{g(x)}dx$ and the associated norm is $\|f\|^2 = \int_{\mathbb{R}} |f(x)|^2 dx$. The Fourier transform of f is $f^*(x) = \int_{\mathbb{R}} e^{ixu} f(u) du$ for all $x \in \mathbb{R}$. When f is the density of a random variable X , $f^*(u) = \mathbb{E}[e^{iuX}]$ is called the characteristic function of X . Then the convolution product of f and g for all $x \in \mathbb{R}$, is $f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy$. Finally we remind the Plancherel-Parseval's formula: $\forall f \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$, $2\pi\|f\|^2 = \|f^*\|^2$.

2.1 Estimator using all the observations: \hat{f}

In this section we build an estimator of the density f_β using all the available observations. We consider the normalized variables $Z_{k,m}$ defined by the difference between two observations, for all $k \in \{1, \dots, N\}$, and for all $m \in \{1, \dots, J\}$

$$Z_{k,m} := \frac{Y_{k,m} - Y_{k,0}}{m\Delta} = \beta_k + \frac{\epsilon_{k,m} - \epsilon_{k,0}}{m\Delta} =: \beta_k + W_{k,m}. \quad (2)$$

For a given m , the variables $W_{k,m}$, $k \in \{1, \dots, N\}$, are i.i.d. with density f_{W_m} . Due to the independence between the ϵ_k 's and the β_k 's, the $Z_{k,m}$ are i.i.d. with density denoted by f_{Z_m} , and definition (2) implies

$$f_{Z_m} = f_\beta \star f_{W_m}. \quad (3)$$

The Fourier transform of (3) yields

$$f_{Z_m}^* = f_\beta^* f_{W_m}^*.$$

By independence of the $\epsilon_{k,m}$ we have for all $u \in \mathbb{R}$

$$\begin{aligned} f_{W_m}^*(u) &= \mathbb{E}[e^{iu(\epsilon_{k,m} - \epsilon_{k,0})/m\Delta}] = \mathbb{E}[e^{iu\epsilon_{k,m}/m\Delta}] \mathbb{E}[e^{-iu\epsilon_{k,0}/m\Delta}] \\ &= \left| f_\epsilon^* \left(\frac{u}{m\Delta} \right) \right|^2. \end{aligned}$$

Let us consider the following assumptions:

- (H1) $f_\beta \in \mathbb{L}_2(\mathbb{R})$ and $f_\beta^* \in \mathbb{L}_1(\mathbb{R})$
(H2) f_ϵ is known and $f_\epsilon^* \neq 0$.

(H1) is needed to compute the $\mathbb{L}_2(\mathbb{R})$ -risk and apply the Fourier inversion: $f_\beta(x) = 1/(2\pi) \int_{\mathbb{R}} e^{-iux} f_\beta^*(u) du$. Assumption (H2) is true for most densities (Gaussian, exponential, Gamma, Laplace) but excludes for example the uniform density on a symmetric interval. Under (H2), we obtain for all $u \in \mathbb{R}$

$$f_\beta^*(u) = \frac{f_{Z_m}^*(u)}{\left| f_\epsilon^* \left(\frac{u}{m\Delta} \right) \right|^2},$$

and using the Fourier inversion, we deduce the following closed formula for f_β

$$\forall x \in \mathbb{R}, f_\beta(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \frac{f_{Z_m}^*(u)}{\left| f_\epsilon^* \left(\frac{u}{m\Delta} \right) \right|^2} du. \quad (4)$$

To propose an estimator of f_β , $f_{Z_m}^*$ is replaced by its empirical estimator

$$\forall u \in \mathbb{R}, \hat{f}_{Z_m}^*(u) = \frac{1}{N} \sum_{k=1}^N e^{iuZ_{k,m}}.$$

If this estimator is directly plugged in equation (4), we face some integration problems. To avoid them we introduce a cut-off. We choose $m\Delta$ as a cut-off where m is as before the time of observation subscript m . Finally the estimator is defined by:

$$\hat{f}_{\beta,m}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{\hat{f}_{Z_m}^*(u)}{\left| f_\epsilon^* \left(\frac{u}{m\Delta} \right) \right|^2} du = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{1}{N} \sum_{k=1}^N \frac{e^{iuZ_{k,m}}}{\left| f_\epsilon^* \left(\frac{u}{m\Delta} \right) \right|^2} du. \quad (5)$$

We will see how this original choice of the cut-off will unexpected better rates for $\hat{f}_{\beta,m}$. In the following, we first compute the MISE of \hat{f} .

2.2 Risk bound for the estimator \hat{f}

We define the function $f_{\beta,m}$ by the relation

$$f_{\beta,m}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} f_\beta^*(u) du \quad (6)$$

which amounts to $\mathbb{E}[\hat{f}_{\beta,m}] = f_{\beta,m}$. Thus $\hat{f}_{\beta,m}$ is an unbiased estimator of $f_{\beta,m}$.

The bias-variance decomposition is the first step to study the MISE. Then we bound the two terms using Plancherel-Parseval's theorem, and we get the following bound.

Proposition 1. Under (H1)-(H2), denoting $F_\epsilon := \int_{-1}^1 (1/|f_\epsilon^*(v)|^4) dv$, we have

$$\mathbb{E}[\|\widehat{f}_{\beta,m} - f_\beta\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_\beta^*(u)|^2 du + F_\epsilon \frac{m\Delta}{2\pi N}. \quad (7)$$

The proofs are relegated to Section 6. The first term of the bound (7) is the integrated bias term $\|f_{\beta,m} - f_\beta\|^2$, which represents the error due to the cutting of the integral in $\widehat{f}_{\beta,m}$. This term is decreasing when m is increasing. The second term is the variance term which increases with m and comes from the estimation of $f_{\beta,m}$ by $\widehat{f}_{\beta,m}$. It has the order m/N : it is the order we would obtain in a direct density estimation context. This was unexpected here because we do not observe β . This gain is due to the choice of the cut-off linked to the observation index. Besides $F_\epsilon = \int_{-1}^1 1/|f_\epsilon^*(v)|^4 dv$ is a constant which depends on the model, and under **(H2)** it is a known numerical constant. Finally the bound (7) depends on the density of the noise only through a constant.

The next step is to select the cut-off m among the collection. It follows from Proposition 1 that we have to perform a bias-variance compromise to minimize the MISE. When f_β is in the Sobolev space with regularity parameter b , i.e. f_β belongs to the set $\mathcal{A}_b(L)$ defined by

$$\mathcal{A}_b(L) = \{f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \int_{\mathbb{R}} |f^*(x)|^2 (1+x^2)^b dx \leq L\}$$

with $b > 0$, $L > 0$, f_β is said *ordinary smooth*, that is, its decay is polynomial. In this particular case, the bias term satisfies:

$$\|f_{\beta,m} - f_\beta\|^2 = \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_\beta^*(u)|^2 du \leq \frac{L}{2\pi} (m\Delta)^{-2b}.$$

This gives that the \mathbb{L}_2 -risk of $\widehat{f}_{\beta,m}$ is bounded:

$$\mathbb{E}[\|\widehat{f}_{\beta,m} - f_\beta\|^2] \leq C_1 m^{-2b} + C_2 \frac{m}{N}$$

with $C_1 = (L/(2\pi))\Delta^{-2b} > 0$, and $C_2 = F_\epsilon \Delta / (2\pi) > 0$. Therefore, the optimal theoretical choice of m is $m_{opt} = C_b N^{\frac{1}{2b+1}}$ with $C_b = (2bC_1/C_2)^{1/(2b+1)}$. Then we can prove the following corollary of Proposition 1.

Corollary 2. If $f \in \mathcal{A}_b(L)$, and $C_b N^{\frac{1}{2b+1}} \leq J$, then under (H1)-(H2)

$$\mathbb{E}[\|\widehat{f}_{\beta,m_{opt}} - f_\beta\|^2] \leq KN^{\frac{-2b}{2b+1}}$$

for a constant K depending on b , L and F_ϵ .

The order of the risk is $N^{-2b/(2b+1)}$ for a large N , it is the nonparametric estimation rate of convergence obtained when the observation are realizations of the variable of interest. We refer to Lacour (2006) Table 1 for classical rates of convergence for nonparametric deconvolution. Comparing with these rates, we observe that here we improve the speed of convergence. This order is $N^{-b/(2b+1)}$ which is always larger than the parametric rate $1/\sqrt{N}$, but can get close to it when b increases. The condition $C_b N^{\frac{1}{2b+1}} \leq J$ requires that J is asymptotically large when N grows, but the larger b the weaker the condition. Nevertheless, this choice is theoretical because it depends on the regularity b of f_β which is unknown.

2.3 Adaptive selection of m and oracle risk bound for \widehat{f}

In this section we propose an adaptive selection of m which should outbalance the bias-variance compromise. We aim at finding a procedure that does not depend on the regularity of the density f_β , but only on the data. Suppose that m belongs to a finite set \mathcal{M} , the theoretical optimal choice for m is

$$m_{th} = \operatorname{argmin}_{m \in \mathcal{M}} \{\|f_{\beta,m} - f_\beta\|^2 + V(\widehat{f}_{\beta,m})\},$$

where $V(\widehat{f}_{\beta,m}) := \mathbb{E}[\|\widehat{f}_{\beta,m} - f_{\beta,m}\|^2] \leq F_\epsilon m \Delta / (2\pi N)$. The two terms appearing in this decomposition are unknown and thus replaced by estimators. The variance term is replaced by a penalty function proportional to the variance bound:

$$\text{pen}^{(1)}(m) = \kappa^{(1)} F_\epsilon \frac{m \Delta}{N}$$

with $m \in \mathcal{M} := \{1, \dots, N \wedge J\}$ where $N \wedge J = \min(N, J)$ and $\kappa^{(1)}$ a numerical constant. From the theoretical point of view $\kappa^{(1)} \geq 24/\pi$, but in practice, $\kappa^{(1)}$ is calibrated by intensive preliminary simulation experiments (see Section 4). This choice of \mathcal{M} implies $m \leq N$ and the penalty is bounded.

The dependence in m which holds in several aspects, prevents us from using usual selection method defined by penalized criterion. More precisely, Equation (5) depends on m in the bounds of the integral, and in the integrative term, thus estimator $\widehat{f}_{\beta,m}$ can not be written as the minimizer of a contrast (see Massart (2007)). This is why we use a more sophisticated method of selection due to a recent work of Goldenshluger and Lepski (2011). The index m is adaptively chosen as the value which minimizes the sum of two terms as follows

$$\widehat{m}^{(1)} = \underset{m \in \mathcal{M}}{\text{argmin}} \{ \widehat{\Gamma}_m + \text{pen}^{(1)}(m) \}$$

with for $m \in \mathcal{M} = \{1, \dots, N \wedge J\}$

$$\begin{aligned} \widehat{\Gamma}_m &:= \max_{1 \leq j \leq N \wedge J} \left(\|\widehat{f}_{\beta, m \wedge j} - \widehat{f}_{\beta, j}\|^2 - \text{pen}^{(1)}(j) \right)_+ \\ &= \max_{m \leq j \leq N \wedge J} \left(\|\widehat{f}_{\beta, m} - \widehat{f}_{\beta, j}\|^2 - \text{pen}^{(1)}(j) \right)_+ \end{aligned} \quad (8)$$

where $(\cdot)_+$ is the positive part. The contrast $\widehat{\Gamma}_m$ approximates the bias. The penalty term is subtracted in (8) to offset the variance effect implied by the estimation of the bias term. The final estimator $\widehat{f}_{\beta, \widehat{m}^{(1)}}$ satisfies the following oracle inequality.

Theorem 3. *Under (H1)-(H2), there exist two constants C and C' such that*

$$\mathbb{E}[\|\widehat{f}_{\beta, \widehat{m}^{(1)}} - f_\beta\|^2] \leq C \inf_{m \in \mathcal{M}} (\|f_{\beta, m} - f_\beta\|^2 + \text{pen}^{(1)}(m)) + \frac{C'}{N}. \quad (9)$$

The constant C is numerical as soon as $\kappa^{(1)}$ is fixed, but not C' .

We prove this result with the same kind of tools as Comte *et al.* (2013) for a continuous time model. Theorem 3 provides a non-asymptotic result and bound (9) shows that $\widehat{f}_{\beta, \widehat{m}^{(1)}}$ realizes automatically the bias-variance compromise. This estimator fulfills the objective: it does not require any assumption on the regularity of f_β or f_ϵ . The bound we obtain depends on the density f_ϵ only through the constant F_ϵ . This is new in the mixed model framework. Again, note that the variance is of order m/N which is better than a deconvolution order.

We establish the following asymptotic result in the particular case where f_β is in the Sobolev space with regularity b .

Corollary 4. *If $f_\beta \in \mathcal{A}_b(L)$ with $b > \frac{1}{2}$ and $J \geq \sqrt{N}$, then, under (H1)-(H2), there exists a constant $C > 0$ such that*

$$\mathbb{E}[\|\widehat{f}_{\beta, \widehat{m}^{(1)}} - f_\beta\|^2] \leq CN^{-\frac{2b}{2b+1}}.$$

Thus, if J is large enough compared to \sqrt{N} , the estimator reaches the best possible rate for a function with regularity b (Tsybakov (2004) shows that $N^{-b/(2b+1)}$ is the optimal rate of convergence), without requiring the knowledge of b . This speed is the one we obtain when we observe directly the variable of interest. It is not the classical deconvolution speed which can be logarithmic in the Gaussian case for example (Comte and Samson, 2012, see), or $N^{-b/(2b+2\gamma+1)}$ if the noise is in $\mathcal{A}_\gamma(L)$. Theorem 3 does not require $J > \sqrt{N}$ but Corollary 4 shows that it should not be too small.

2.4 Estimator built using the last observation time \widehat{f}^J

In order to simplify the previous estimator and its selection method, we propose a second estimator. This estimator is based only on the variables defined by (2) for $m = J$, $\forall k \in \{1, \dots, N\}$,

$$Z_{k,J} = \frac{Y_{k,J} - Y_{k,0}}{J\Delta} = \beta_k + \frac{\epsilon_{k,J} - \epsilon_{k,0}}{J\Delta} = \beta_k + W_{k,J}.$$

The same arguments as before lead to, for any real-value x :

$$f_\beta(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \frac{f_{Z_J}^*(u)}{|f_\epsilon^*(\frac{u}{J\Delta})|^2} du.$$

Then we replace $f_{Z_J}^*$ by its empirical estimator and we introduce for some m the cut-off $m\Delta$. The second estimator is defined by the following equation:

$$\widehat{f}_{\beta,m}^J(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{1}{N} \sum_{k=1}^N \frac{e^{iuZ_{k,J}}}{|f_\epsilon^*(\frac{u}{J\Delta})|^2} du. \quad (10)$$

We notice that m is only a cut-off here and no longer an observation time, then it does not need to be an integer. This estimator can also be seen as the argument minimum of a contrast type density. Indeed, for $t \in S_m := \{t \in \mathbb{L}_2, \text{supp}(t^*) = [-m\Delta, m\Delta]\}$, considering

$$\begin{aligned} \gamma(t) &:= \|t\|^2 - \frac{2}{N} \sum_{k=1}^N \frac{1}{2\pi} \int t^*(-u) \frac{e^{iuZ_{k,J}}}{|f_\epsilon^*(\frac{u}{J\Delta})|^2} du \\ &= \|t\|^2 - 2 \langle t, \widehat{f}_{\beta,m}^J \rangle = \|t - \widehat{f}_{\beta,m}^J\|^2 - \|\widehat{f}_{\beta,m}^J\|^2, \end{aligned}$$

We have

$$\widehat{f}_{\beta,m}^J = \underset{t \in S_m}{\operatorname{argmin}} \gamma(t).$$

This new estimator uses less observations than the estimator defined by (5): NJ observations are used in the first case, only $2N$ in the second. Therefore, we have to check that it remains competitive. Note that intermediate observations will be used to estimate $|f_\epsilon^*|$, if they are available. In order to lead a theoretical study, we compute the MISE of $\widehat{f}_{\beta,m}^J$. Definition (6) still gives that $\widehat{f}_{\beta,m}^J$ is an unbiased estimator of $f_{\beta,m}$. The bias-variance decomposition of the risk gives the following proposition.

Proposition 5. *Under (H1)-(H2),*

$$\mathbb{E}[\|\widehat{f}_{\beta,m}^J - f_\beta\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_\beta^*(u)|^2 du + \frac{1}{2\pi N} \int_{-m\Delta}^{m\Delta} \frac{du}{|f_\epsilon^*(\frac{u}{J\Delta})|^4}. \quad (11)$$

The same comments we made after Proposition 1 apply here. We notice that if the observation time $J\Delta$ is large enough, as the Fourier transform of a density evaluated in 0 is equal to 1, the variance term is under control. Under the condition $m \leq N \wedge J$ we can prove the following result.

Corollary 6. *If $m \leq N \wedge J$, and $B_\epsilon := \sup_{v \in [-1,1]} \frac{1}{|f_\epsilon^*(v)|^4}$, then under (H1)-(H2)*

$$\mathbb{E}[\|\widehat{f}_{\beta,m}^J - f_\beta\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_\beta^*(u)|^2 du + B_\epsilon \frac{m\Delta}{\pi N}. \quad (12)$$

The dependence on the noise is contained in the constant B_ϵ . The estimator $\widehat{f}_{\beta,m}^J$ is well defined when $m > J$. Nevertheless, the variance in inequality (11) can be large in this case, that is the reason why we suppose $m \leq N \wedge J$ to obtain the bound (12). This assumption yields a variance order m/N , which is the best one can expect.

In the following we propose a penalization procedure. With the assumption of Corollary 6, the cut-off m is supposed to belong to the finite set $\mathcal{M} = \{1, \dots, N \wedge J\}$ which is the same as in Section 2.3. The theoretical optimal choice of m is defined by

$$m_{th}^{(2)} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \{\|f_{\beta,m} - f_\beta\|^2 + V(\widehat{f}_{\beta,m}^J)\}.$$

However, the Plancherel-Parseval's formula implies $\|f_\beta - f_{\beta,m}\|^2 = \|f_\beta\|^2 - \|f_{\beta,m}\|^2$, thus

$$m_{th}^{(2)} = \operatorname{argmin}_{m \in \mathcal{M}} \{-\|f_{\beta,m}\|^2 + V(\widehat{f}_{\beta,m}^J)\}$$

where $V(\widehat{f}_{\beta,m}^J) := \mathbb{E}[\|\widehat{f}_{\beta,m}^J - f_{\beta,m}\|^2] \leq B_\epsilon \frac{m\Delta}{\pi N}$. The Goldenshluger and Lepski criterium Γ_m prevents us from using this simplification before, nevertheless it was always true. As f_β is unknown, so are $\|f_{\beta,m}\|^2$ and $V(\widehat{f}_{\beta,m}^J)$. Therefore the value $m_{th}^{(2)}$ must be estimated, and we consider

$$\widehat{m}^{(2)} = \operatorname{argmin}_{m \in \mathcal{M}} \{-\|\widehat{f}_{\beta,m}^J\|^2 + \operatorname{pen}^{(2)}(m)\} = \operatorname{argmin}_{m \in \mathcal{M}} \{\gamma(\widehat{f}_{\beta,m}^J) + \operatorname{pen}^{(2)}(m)\} \quad (13)$$

with $\operatorname{pen}^{(2)}$ a penalty function proportional to the variance term obtained in the bound (12):

$$\operatorname{pen}^{(2)}(m) = \kappa^{(2)} B_\epsilon \frac{\Delta m}{N}$$

where $\kappa^{(2)}$ is a numerical constant calibrated in Section 4. As before the penalty function is bounded. This method is more natural to realize the bias-variance trade off than the one set up in Section 2.1. But we can not set it up for the estimator $\widehat{f}_{\beta,m}$ in a theoretical point of view because of the dependence in m . It is easier to compute and the oracle risk bound we obtain is as good as the last one (9). In order to evaluate the order of the L^2 -risk of the final estimator $\widehat{f}_{\beta,\widehat{m}^{(2)}}^J$ we establish the following oracle inequality.

Theorem 7. *Under (H1)-(H2), there exist two constants C_1 and C_2 such that*

$$\mathbb{E}[\|\widehat{f}_{\beta,\widehat{m}^{(2)}}^J - f_\beta\|^2] \leq C_2 \inf_{m \in \mathcal{M}} (\|f_{\beta,m} - f_\beta\|^2 + \operatorname{pen}^{(2)}(m)) + \frac{C_1}{N}.$$

Similarly to Theorem 3, this oracle inequality is a non-asymptotic result and is true for all N . The choice $\widehat{m}^{(2)}$ realizes automatically the bias penalty trade-off. It is worth stressing that the penalization procedure used to compute $\widehat{f}_{\beta,\widehat{m}^{(2)}}^J$ is simpler and faster than the selection method used for $\widehat{f}_{\beta,\widehat{m}^{(1)}}$ in Section 2.3. Thus, we shall check that its practical performances remain good even if the number of observations used here is much more limited.

3 Comparisons with previous results

In this section, estimators $\widehat{f}_{\beta,m}$ and $\widehat{f}_{\beta,m}^J$ are compared with the estimator proposed in Comte and Samson (2012). Their study is led in the same context. Thus we can compare the construction, the penalty order and theoretical results. Let us introduce the estimator in Comte and Samson (2012) denoted by $\widetilde{f}_{\beta,m}$, for any real-value x

$$\widetilde{f}_{\beta,m}(x) = \frac{2}{NJ} \sum_{k=1}^N \sum_{j=1}^{J/2} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \frac{e^{iuU_{k,j}}}{|f_\epsilon^*(\frac{u}{\Delta})|^2} du \quad (14)$$

with for $k \in \{1, \dots, N\}, j \in \{1, \dots, J/2\}$ the random variables defined as

$$U_{k,j} := \frac{Y_{k,2j} - Y_{k,2j-1}}{\Delta}.$$

We first compare the construction. This estimator is computed by deconvolution method as the two estimators presented in this paper. The main difference is the average over j appearing in the definition (14). Through this manipulation, this estimator uses all the observations. The aim of the authors by averaging over j was to improve the variance. Then note that the parameter Δ appears (alone) in the denominator of the estimator. This forbids to take a Δ too small. For example if the noise is Gaussian $1/|f_\epsilon^*(\frac{u}{\Delta})|^2 = e^{\sigma_\epsilon^2 u^2 / \Delta^2}$ can become very large as soon as $\Delta < 1$. An other remark is that the cut-off πm is independent of the observations as for $\widehat{f}_{\beta,m}^J$. In our work the cut-off is $m\Delta$. We choose Δ instead of π in order to simplify the calculation.

In a second time we compare the adaptive procedure. The cut-off πm is selected as the minimizer of a penalized contrast, as for the estimator $\widehat{f}_{\beta,m}^J$. The chosen penalty is composed of two terms,

when we have only one. The first term is m/N and the second term is $1/(NJ)$ multiplied with an integrative term depending on m , for which assumptions on f_ϵ are needed to determinate the order. On the contrary, for both of our estimators, the penalty functions have the order m/N and are explicit. They are thus simpler. Finally an oracle theorem is obtained in Comte and Samson (2012) assuming the noise is ordinary smooth. This assumption is strong and needed to bound the variance term, and to obtain a MISE order. However, they obtain the optimal rate of convergence of deconvolution cases (in minimax sense Fan (1991)), which is rather slow. In the present work we follow a different strategy aiming to reduce directly the effect of the noise instead of reducing the variance order. Thanks to this approach, no assumption on the noise regularity is needed to obtain Theorems 3, and 7. Furthermore, the obtained rates of convergence are similar to those obtained in the case of density estimation with direct observations.

Nevertheless, we can define a framework of application for each estimator. Estimator $\widehat{f}_{\beta,m}$ given by equation (5) uses all the observations like the one of Comte and Samson (2012) given by (14). Note that it is possible to compute $\widehat{f}_{\beta,m}$ when $\Delta < 1$, while this does not seem reasonable with $\widetilde{f}_{\beta,m}$. Estimator $\widehat{f}_{\beta,m}^J$ given by (10) uses only the first and the last time of observation. It should benefit from a large total time of observations $J\Delta$. Thus, when J is small we should prefer using $\widehat{f}_{\beta,m}$ to $\widehat{f}_{\beta,m}^J$ if all the observations are available. When J is large and Δ is small both $\widehat{f}_{\beta,m}^J$ and $\widetilde{f}_{\beta,m}$ should be efficient. And when only the first and the last time of observation are available we use $\widehat{f}_{\beta,m}^J$. However if J is large and Δ not too small we would prefer $\widetilde{f}_{\beta,m}$.

4 Numerical illustration

4.1 Implementation

In this part we implement the three estimators $\widehat{f}_{\beta,m}$, $\widehat{f}_{\beta,m}^J$, and $\widetilde{f}_{\beta,m}$ in order to compare their performances on simulations. We try different scenarios: different designs for β, α, ϵ (ordinary smooth and super smooth) and different values of the parameters J, Δ, N , to observe the behavior of each estimator. First we consider the case of known noise density. We consider simulated data according to model (1). For β we choose the distributions:

- Gaussian $\mathcal{N}(0, 1)$
- mixed Gaussian with $0.3\mathcal{N}(-1, 1/16) + 0.7\mathcal{N}(1, 1/16)$
- Gamma $\Gamma(25, 1/25) \times 5$
- mixed Gamma $(0.3\Gamma(2, 1/2) + 0.7\Gamma(20, 1/5)) \times (1/\sqrt{3})$

for α :

- Gaussian $\mathcal{N}(0, 1)$
- Gamma $\Gamma(25, 1/25) \times 5$

for ϵ :

- Gaussian $\mathcal{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon = 1/4$ and $1/2$
- Laplace $\epsilon = \sigma_\epsilon \times Y$ with Y standard Laplace distribution: $f_Y(y) = (1/\sqrt{2})\exp(-\sqrt{2}|x|)$.

We choose the same distributions as in Comte and Samson (2012) in order to compare our results to theirs. These densities are all associated with variance one. Each estimation calls for Riemann Sum to compute the integrals. They appear in the estimators and in the quantities F_ϵ, B_ϵ . We choose a thin discretization of the interval by taking 200 points.

Then we need to calibrate the penalty constants $\kappa^{(1)}, \kappa^{(2)}$. We refer to Comte and Samson (2012) for the formula of the penalty of estimator \widetilde{f} and we denote $\widetilde{\kappa}$ their constant. For this purpose we investigate different cases of estimated function, with different parameters values, a large number of times, with different values of $\widetilde{\kappa}, \kappa^{(1)}, \kappa^{(2)}$. At first, we evaluate the visual quality of the recover of f_β (all the estimators in the collection and the final estimator, in order to see if the chosen one is really the best one among them). Secondly we compare the MISE computed over 100 estimations. Finally for $\widehat{f}_{\beta,m}$ we choose $\kappa^{(1)} = 1$, for $\widehat{f}_{\beta,m}^J$: $\kappa^{(2)} = 1$, and for $\widetilde{f}_{\beta,m}$ we choose $\widetilde{\kappa} = 6$.

In Section 4.3 we consider that the noise density is unknown, and the numerical results are presented. These new estimators are established in Section 6.8.

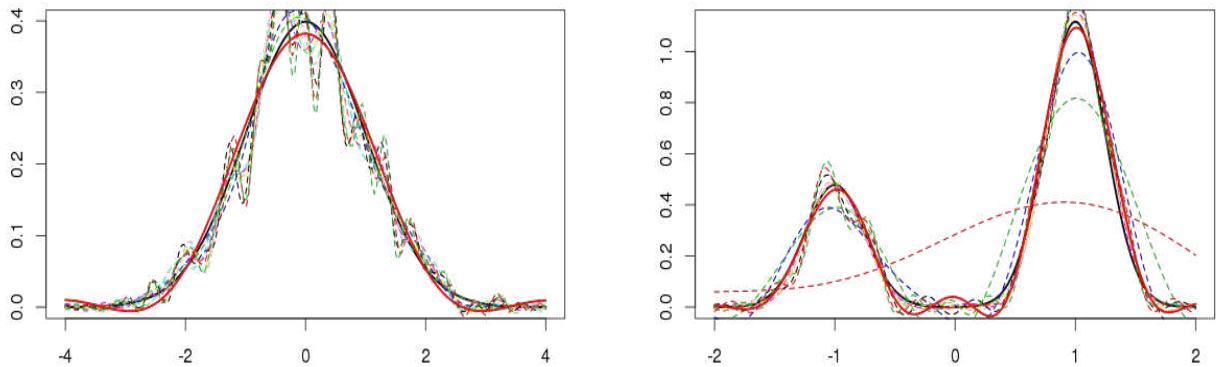


Figure 1: The black bold line is the true density of β : Gaussian on the left, mixed Gaussian on the right. The colored dashed lines are the proposed estimators $\hat{f}_{\beta,m}$, the red bold line is the selected estimator $\hat{f}_{\beta,\hat{m}^{(1)}}$, with $N = 500$, $J = 10$, $\Delta = 2$, $\sigma_\epsilon = 0.25$

4.2 Numerical simulation results

All the designs discussed in Section 4.1 produce the same kind of results, thus we only present the case β Gaussian and mixed Gaussian, with α, ϵ Gaussian.

First we investigate the results on graphs. Figures 1, 2, and 3, represent two cases: on the left β is Gaussian, on the right β is mixed Gaussian. Figure 1 compares the estimator $\hat{f}_{\beta,\hat{m}^{(1)}}$ (the bold red line), and all the estimators of the collection $\hat{f}_{\beta,m}$ (in coloured dashed lines), with the true density of β . We can see that the performance of the selection procedure is visually satisfying. Figure 2 represents 10 final estimators $\hat{f}_{\beta,\hat{m}^{(2)}}^J$ (green dotted lines) versus the true density of β . This figure ensures the stability of the estimation procedure. Indeed the beam of estimators is really narrow. Figures 3 and 4 compare the three estimators for different values of Δ . On Figure 3 we see that estimators are really close to each other and close to the true density f_β . Figure 4 illustrates the role of the parameter Δ . Indeed, due to its construction, estimator $\hat{f}_{\beta,m}$ of Comte and Samson (2012) defined by (14) is more sensitive than the others to a small Δ , which is clear on Figure 4.

During this study we also observe differences between estimators when σ_ϵ increases, especially in the case of a bimodal density f_β . This highlights the role of the noise variance. Estimator \hat{f}^J is the most stable with respect to the variance effects.

Finally we compare empirical MISE ($\times 100$) of the three estimators. Results are presented in Table 1 as a function of J, N and of the distribution of β . The estimator $\hat{f}_{\beta,m}$ has the smallest MISE mainly for $J = 6, 20$ and β Gaussian, and $\hat{f}_{\beta,m}^J$ is the winner in the other cases. These two estimators seem very competitive. We notice that the influence of N is clear: the larger N the smaller the MISE. Nevertheless, the role of J is less explicit on this table. Indeed, when β is Gaussian, the MISE are smaller when $J = 6$ than when $J = 100$. However, when J is higher estimator $\hat{f}_{\beta,\hat{m}}^J$ produces the best MISE. More precisely the results are better when $J = N$ than when $J > N$ and Table 1 leads us to conclude that the results are better if J is of order \sqrt{N} as we noticed in the theoretical study of $\hat{f}_{\beta,m}$.

4.3 Case of unknown noise density

We briefly consider the case of f_ϵ unknown. Several papers study the deconvolution method with unknown noise density from a theoretical point of view: Neumann (1997), Comte and Lacour (2011), Meister and Neumann (2010), Kappus and Mabon (2014). Here, we plug estimators of unknown quantities using a transformation of the data adapted from Comte and Samson (2012).

We generalize $\hat{f}_{\beta,m}$ and $\hat{f}_{\beta,m}^J$ to unknown f_ϵ by computing new estimators $\widehat{\widehat{f}}_{\beta,m}$ and $\widehat{\widehat{f}}_{\beta,m}^J$ (see

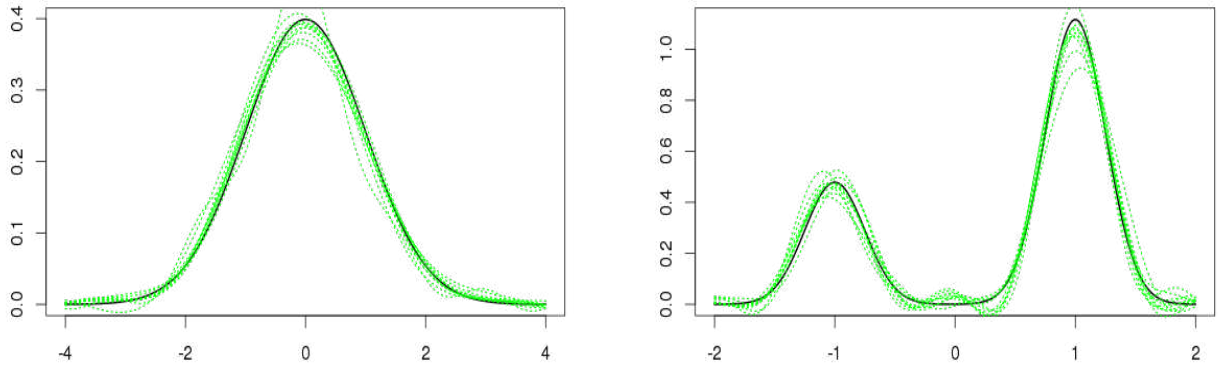


Figure 2: The black bold line is the true density of β : Gaussian on the left, mixed Gaussian on the right. The green dotted lines: 10 $\hat{f}_{\beta, \hat{m}^{(2)}}^J$ estimators with $N = 500$, $J = 6$, $\Delta = 2$, $\sigma_\epsilon = 0.25$

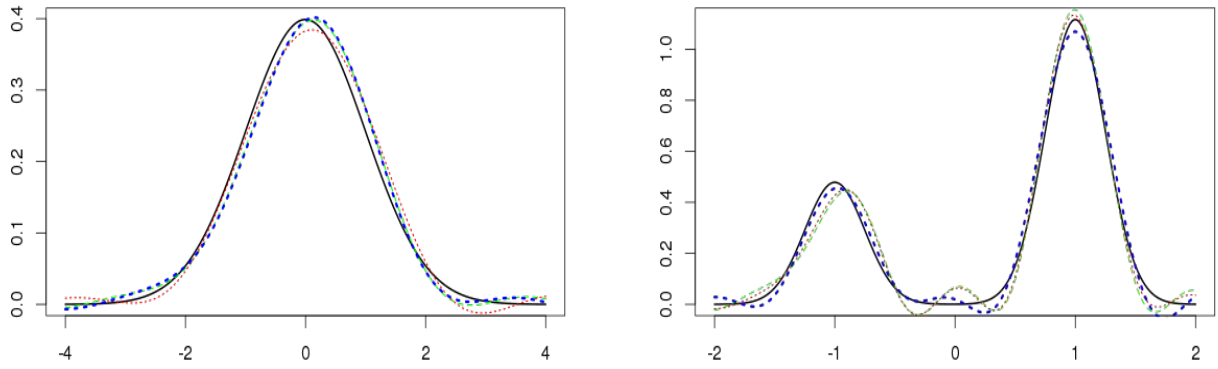


Figure 3: The black line is the true density of β : Gaussian on the left, mixed Gaussian on the right. The green dashed line is $\tilde{f}_{\beta, \tilde{m}}$, the red thin dotted line is $\hat{f}_{\beta, \hat{m}^{(1)}}^J$, the thick blue dotted line is $\hat{f}_{\beta, \hat{m}^{(2)}}^J$, $N = 500$, $J = 6$, $\Delta = 2$, $\sigma_\epsilon = 0.25$

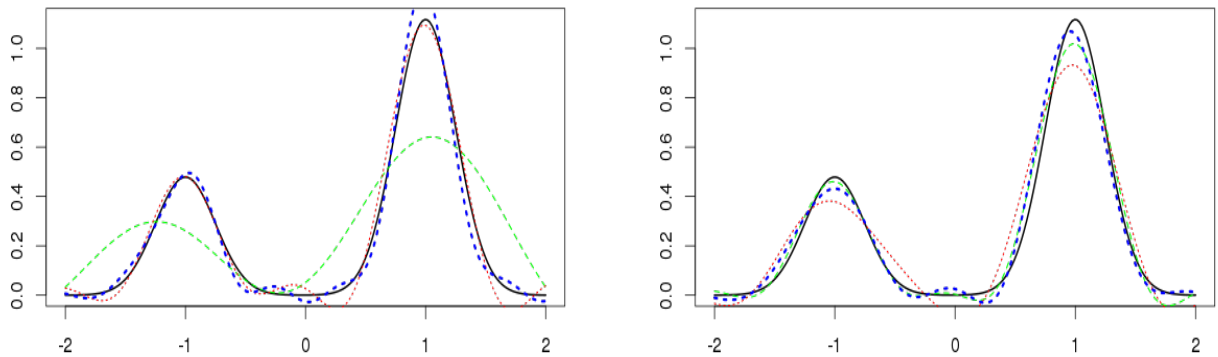


Figure 4: The black bold line is the true density of β : mixed Gaussian. The green dashed line is $\tilde{f}_{\beta, \tilde{m}}$, the red thin dotted line is $\hat{f}_{\beta, \hat{m}^{(1)}}^J$, the thick blue dotted line is $\hat{f}_{\beta, \hat{m}^{(2)}}^J$, $N = 500$, $J = 6$, $\sigma_\epsilon = 1$. On the left $\Delta = 1$, on the right $\Delta = 3$

Table 1: Empirical MISE ($\times 100$) computed from 100 simulated data sets with α, ϵ Gaussian, and $\Delta = 2$ when f_ϵ is known

Estimator	J	distribution of β	$\sigma_\epsilon = 1/4$		$\sigma_\epsilon = 1/2$		
			$N = 20$	$N = 200$	$N = 20$	$N = 200$	$N = 2000$
$\widehat{f}_{\beta,m}$	6	Gaussian	5.083	0.474	6.787	0.552	0.039
$\widetilde{f}_{\beta,m}$	6	Gaussian	1.941	0.245	4.210	0.297	0.106
$\widehat{f}_{\beta,m}^J$	6	Gaussian	2.275	0.295	2.878	0.317	0.030
$\widetilde{f}_{\beta,m}$	20	Gaussian	3.712	0.434	4.256	5.745	0.037
$\widehat{f}_{\beta,m}$	20	Gaussian	1.782	0.293	2.327	4.145	0.111
$\widehat{f}_{\beta,m}^J$	20	Gaussian	2.443	0.441	2.537	2.758	0.032
$\widetilde{f}_{\beta,m}$	6	MixedGaussian	30.02	2.497	34.83	6.790	3.037
$\widehat{f}_{\beta,m}$	6	MixedGaussian	30.57	3.508	29.95	4.411	0.372
$\widehat{f}_{\beta,m}^J$	6	MixedGaussian	9.714	2.769	10.79	2.840	2.242
Estimator	J	distribution of β	$N = 20$	$N = 100$	$N = 20$	$N = 100$	$N = 2000$
$\widetilde{f}_{\beta,m}$	100	Gaussian	3.833	0.698	12.88	0.715	0.532
$\widehat{f}_{\beta,m}$	100	Gaussian	4.618	1.093	31.29	1.245	0.509
$\widehat{f}_{\beta,m}^J$	100	Gaussian	2.734	0.509	9.968	0.633	2.221

Section 6.8 for details). The selection procedures presented in Section 2.3 and 2.4 are using the constants F_ϵ and B_ϵ . These quantities are estimated using relation (27) established in Section 6.8, and we obtain final estimators by applying the previous set up adaptive selection procedure with plugged estimated quantities. Of course, the new estimator $\widehat{f}_{\beta,m}^J$ requires that intermediate observations between the first and the last are available.

In Table 2, we compare empirical MISE of $\widetilde{f}_{\beta,m}$, $\widehat{f}_{\beta,m}^J$ and $\widehat{f}_{\beta,m}$ built in Comte and Samson (2012), as a function of J, N and the distribution of β . We notice that $\widehat{f}_{\beta,m}$ has the smallest MISE most of the time when $J = 6, 20$ and β is Gaussian, and it is $\widehat{f}_{\beta,m}^J$ when $J = 6$ with β mixed Gaussian or $J = 100$ and β Gaussian.

The curves of these estimators have the same behaviors than with a known noise density (not shown).

5 Concluding remarks

In summary, in this work we introduce two advanced estimators $\widehat{f}_{\beta,m}$ and $\widehat{f}_{\beta,m}^J$ which are completely data driven. We obtained convincing theoretical results, validated by numerical study. In particular the original selection method set up in Section 2.1 leads to an adaptive estimator $\widehat{f}_{\beta,m}$. Furthermore it satisfies an oracle inequality which shows that it realizes automatically the bias-penalty trade-off for all N and all regularity of f_β, f_ϵ . The second estimator $\widehat{f}_{\beta,m}^J$ seems simpler, nevertheless it produces really good theoretical results and also on simulations, in Section 4.2. It also has the advantage of using only the first and the last time of observation by opposition with $\widetilde{f}_{\beta,m}$.

The numerical study confirms the theoretical study: we can define a field of applications for each estimator. Precisely, when J is small and all the observations are available we will prefer $\widehat{f}_{\beta,m}$. But when $J\Delta$ is large, and/or when only the first and the last observations are available, we will prefer $\widehat{f}_{\beta,m}^J$.

We may wonder if the estimation of f_α could be improved. Moreover as no independence assumption is set between α and β , the joint distribution of the couple may be estimated over anisotropic regularity spaces. These two questions will be studied in forgoing works.

Table 2: Empirical MISE ($\times 100$) computed from 100 simulated data sets with α, ϵ Gaussian, and $\Delta = 2$, with f_ϵ unknown and estimated

Estimator	J	distribution of β	$\sigma_\epsilon = 1/4$		$\sigma_\epsilon = 1/2$		
			$N = 20$	$N = 200$	$N = 20$	$N = 200$	$N = 2000$
$\widetilde{f_{\beta,m}}$	6	Gaussian	6.367	0.660	6.368	0.541	0.0462
$\widehat{f_{\beta,m}}$	6	Gaussian	2.224	0.283	2.317	0.304	0.107
$\widehat{f_{\beta,m}^J}$	6	Gaussian	3.140	0.309	3.526	0.288	0.044
$\widetilde{f_{\beta,m}}$	20	Gaussian	4.316	0.406	4.159	0.402	0.041
$\widehat{f_{\beta,m}}$	20	Gaussian	1.934	0.248	2.094	0.260	0.109
$\widehat{f_{\beta,m}^J}$	20	Gaussian	2.573	0.304	2.120	0.263	0.037
$\widetilde{f_{\beta,m}}$	6	mixed Gaussian	25.74	2.751	35.48	6.610	3.776
$\widehat{f_{\beta,m}}$	6	mixed Gaussian	30.38	3.980	29.82	3.598	0.378
$\widehat{f_{\beta,m}^J}$	6	mixed Gaussian	10.29	1.227	10.28	1.363	0.168
Estimator	J	distribution of β	$N = 20$	$N = 100$	$N = 20$	$N = 100$	$N = 2000$
$\widetilde{f_{\beta,m}}$	100	Gaussian	3.622	0.729	3.271	0.761	0.378
$\widehat{f_{\beta,m}}$	100	Gaussian	6.621	0.994	5.462	1.398	0.729
$\widehat{f_{\beta,m}^J}$	100	Gaussian	2.136	0.515	2.314	0.532	0.280

Acknowledgements

I am really thankful to Fabienne Comte and Adeline Samson for their contribution to this work. I also thank Christophe Giraud for useful discussions, and Gwennaëlle Mabon for the read-through.

6 Proofs

6.1 Talagrand's inequality

The following result follows from the Talagrand concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998).

Proposition 8. *Consider $n \in \mathbb{N}^*$, \mathcal{F} a class at most countable of measurable functions, and $(X_i)_{i \in \{1, \dots, N\}}$ a family of real independent random variables. One defines, for all $f \in \mathcal{F}$,*

$$\nu_N(f) = \frac{1}{N} \sum_{i=1}^N (f(X_i) - \mathbb{E}[f(X_i)]).$$

Supposing there are three positive constants M, H and v such that $\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M$,

$\mathbb{E}[\sup_{f \in \mathcal{F}} |\nu_N f|] \leq H$, and $\sup_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N \text{Var}(f(X_i)) \leq v$, then for all $\alpha > 0$,

$$\mathbb{E} \left[\left(\sup_{f \in \mathcal{F}} |\nu_N(f)|^2 - 2(1 + 2\alpha)H^2 \right)_+ \right] \leq \frac{4}{K_1} \left(\frac{v}{N} \exp \left(-K_1 \alpha \frac{NH^2}{v} \right) + \frac{49M^2}{K_1 C^2(\alpha) N^2} \exp \left(-\frac{\sqrt{2} K_1 C(\alpha) \sqrt{\alpha} NH}{7} \right) \right)$$

with $C(\alpha) = (\sqrt{1 + \alpha} - 1) \wedge 1$, and $K_1 = \frac{1}{6}$.

6.2 Proof of Proposition 1

The MISE of $\widehat{f}_{\beta,m}$ is defined by $\mathbb{E}[\|\widehat{f}_{\beta,m} - f_{\beta}\|^2] = \|\mathbb{E}[\widehat{f}_{\beta,m}] - f_{\beta}\|^2 + \mathbb{E}[\|\widehat{f}_{\beta,m} - \mathbb{E}[\widehat{f}_{\beta,m}]\|^2]$. First, according to the definition of $f_{\beta,m}$, $\widehat{f}_{\beta,m}$ is an unbiased estimator of $f_{\beta,m}$, thus the squared integrated bias term is $\|\mathbb{E}[\widehat{f}_{\beta,m}] - f_{\beta}\|^2 = \|f_{\beta,m} - f_{\beta}\|^2$. But, for all $x \in \mathbb{R}$:

$$f_{\beta}(x) - \mathbb{E}(\widehat{f}_{\beta,m}(x)) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} (f_{\beta}^*(u) - f_{\beta,m}^*(u)) du = \frac{1}{2\pi} \int_{|u| \geq m\Delta} e^{-iux} f_{\beta}^*(u) du .$$

We obtain with Plancherel-Parserval's theorem

$$\|f_{\beta} - f_{\beta,m}\|^2 = \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_{\beta}^*(u)|^2 du .$$

We study now the term of variance $\text{Var}(\widehat{f}_{\beta,m}) = \mathbb{E}[\|\widehat{f}_{\beta,m} - \mathbb{E}(\widehat{f}_{\beta,m})\|^2] = \mathbb{E}[\|\widehat{f}_{\beta,m} - f_{\beta,m}\|^2]$:

$$\begin{aligned} \|\widehat{f}_{\beta,m} - f_{\beta,m}\|^2 &= \left\| \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{\widehat{f}_{Z_m}^*(u) - f_{Z_m}^*(u)}{|f_{\epsilon}^*(\frac{u}{m\Delta})|^2} du \right\|^2 = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \left| \frac{\widehat{f}_{Z_m}^*(u) - f_{Z_m}^*(u)}{|f_{\epsilon}^*(\frac{u}{m\Delta})|^2} \right|^2 du \\ &= \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \left| \frac{1}{N} \sum_{k=1}^N \frac{e^{iuZ_{k,m}} - \mathbb{E}[e^{iuZ_{k,m}}]}{|f_{\epsilon}^*(\frac{u}{m\Delta})|^2} \right|^2 du . \end{aligned}$$

By noticing that the random variables in the last sum are i.i.d, we have

$$\begin{aligned} \mathbb{E}[\|\widehat{f}_{\beta,m} - f_{\beta,m}\|^2] &= \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \text{Var} \left(\frac{1}{N} \sum_{k=1}^N \frac{e^{iuZ_{k,m}}}{|f_{\epsilon}^*(\frac{u}{m\Delta})|^2} \right) du = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \frac{1}{N} \frac{\text{Var}(e^{iuZ_{k,m}})}{|f_{\epsilon}^*(\frac{u}{m\Delta})|^4} du \\ &\leq \frac{1}{2\pi N} \int_{-m\Delta}^{m\Delta} \frac{du}{|f_{\epsilon}^*(\frac{u}{m\Delta})|^4} = F_{\epsilon} \frac{m\Delta}{2\pi N} . \quad \square \end{aligned}$$

6.3 Notations

We define

$$\begin{aligned} S_m &= \{t \in \mathbb{L}_2, \text{supp}(t^*) = [-m\Delta, m\Delta]\} \\ B_{m,m'} &= \{t \in S_m \vee m', \|t\| = 1\} \end{aligned}$$

and for $t \in S_m$:

$$\phi_t^j(x) = \frac{1}{2\pi} \int t^*(-u) \frac{e^{iux}}{|f_{\epsilon}^*(\frac{u}{j\Delta})|^2} du. \quad (15)$$

Then we define the empirical contrast for an estimator \widehat{f} of f , for $t \in S_m$

$$\nu(t) = \frac{1}{2\pi} \langle t^*, (\widehat{f}_m - f_m)^* \rangle . \quad (16)$$

In order to make the reading easier we suppress the subscript β , and the exponents (1) in the following.

6.4 Proof of Theorem 3

Let us study the bias term $\|\widehat{f}_{\widehat{m}} - f\|^2$. For all $m \in \mathcal{M}$,

$$\begin{aligned} \|\widehat{f}_{\widehat{m}} - f\|^2 &\leq \left(\|\widehat{f}_{\widehat{m}} - \widehat{f}_{\widehat{m} \wedge m}\| + \|\widehat{f}_{\widehat{m} \wedge m} - \widehat{f}_m\| + \|\widehat{f}_m - f\| \right)^2 \\ &\leq 3 \left(\|\widehat{f}_{\widehat{m}} - \widehat{f}_{\widehat{m} \wedge m}\|^2 + \|\widehat{f}_{\widehat{m} \wedge m} - \widehat{f}_m\|^2 + \|\widehat{f}_m - f\|^2 \right) . \end{aligned}$$

We are going to examine the three terms involved separately. For the first:

$$\begin{aligned} \|\widehat{f}_{\widehat{m}} - \widehat{f}_{\widehat{m} \wedge m}\|^2 &= \|\widehat{f}_{\widehat{m}} - \widehat{f}_m\|^2 \mathbf{1}_{\{\widehat{m} \geq m\}} \\ &= (\|\widehat{f}_{\widehat{m}} - \widehat{f}_m\|^2 - \text{pen}(\widehat{m})) \mathbf{1}_{\{\widehat{m} \geq m\}} + \text{pen}(\widehat{m}) \mathbf{1}_{\{\widehat{m} \geq m\}} \\ &\leq \widehat{\Gamma}_m + \text{pen}(\widehat{m}) . \end{aligned}$$

After,

$$\begin{aligned}
\|\widehat{f}_{\widehat{m} \wedge m} - \widehat{f}_m\|^2 &= \|\widehat{f}_{\widehat{m}} - \widehat{f}_m\|^2 \mathbf{1}_{\{\widehat{m} \leq m\}} \\
&= (\|\widehat{f}_{\widehat{m}} - \widehat{f}_m\|^2 - \text{pen}(m)) \mathbf{1}_{\{\widehat{m} \leq m\}} + \text{pen}(m) \mathbf{1}_{\{\widehat{m} \leq m\}} \\
&\leq \widehat{\Gamma}_{\widehat{m}} + \text{pen}(m),
\end{aligned}$$

what gives

$$\begin{aligned}
\|\widehat{f}_{\widehat{m}} - f\|^2 &\leq 3 \left(\widehat{\Gamma}_{\widehat{m}} + \text{pen}(m) + \widehat{\Gamma}_m + \text{pen}(\widehat{m}) + \|\widehat{f}_m - f\|^2 \right) \\
&\leq 6(\text{pen}(m) + \widehat{\Gamma}_m) + 3\|\widehat{f}_m - f\|^2.
\end{aligned} \tag{17}$$

Now we study $\widehat{\Gamma}_m$. For $j \geq m$, we have

$$\|\widehat{f}_j - \widehat{f}_m\|^2 \leq 3 \left(\|\widehat{f}_j - f_j\|^2 + \|f_j - f_m\|^2 + \|f_m - \widehat{f}_m\|^2 \right).$$

According to Plancherel-Parseval's theorem

$$\begin{aligned}
\|f_j - f_m\|^2 &= \frac{1}{2\pi} \|f_j^* - f_m^*\|^2 = \frac{1}{2\pi} \int_{m\Delta \leq |u| \leq j\Delta} |f^*(u)|^2 du \\
&\leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f^*(u)|^2 du = \|f - f_m\|^2.
\end{aligned}$$

Besides

$$\begin{aligned}
\|\widehat{f}_m - f\|^2 &= \frac{1}{2\pi} \|\widehat{f}_m^* - f^*\|^2 = \frac{1}{2\pi} \|(\widehat{f}_m^* - f_m^*) + (f_m^* - f^*)\|^2 \\
&= \|\widehat{f}_m - f_m\|^2 + \|f_m - f\|^2.
\end{aligned}$$

Thus

$$\|\widehat{f}_j - \widehat{f}_m\|^2 \leq 3(\|\widehat{f}_j - f_j\|^2 + \|\widehat{f}_m - f\|^2).$$

Finally,

$$\begin{aligned}
\widehat{\Gamma}_m &= \max_{m \leq j \leq N \wedge J} \left(\|\widehat{f}_m - \widehat{f}_j\|^2 - \text{pen}(j) \right)_+ \\
&\leq \max_{m \leq j \leq N \wedge J} \left(3\|\widehat{f}_j - f_j\|^2 + 3\|\widehat{f}_m - f\|^2 - \text{pen}(j) \right)_+ \\
&\leq 3 \max_{m \leq j \leq N \wedge J} \left(\|\widehat{f}_j - f_j\|^2 - \frac{1}{3} \text{pen}(j) \right)_+ + 3\|\widehat{f}_m - f\|^2
\end{aligned} \tag{18}$$

Finally, gathering (17) and (18) we obtain

$$\begin{aligned}
\|\widehat{f}_{\widehat{m}} - f\|^2 &\leq 6 \left(3 \max_{m \leq j \leq N \wedge J} \left(\|\widehat{f}_j - f_j\|^2 - \frac{1}{3} \text{pen}(j) \right)_+ + 3\|\widehat{f}_m - f\|^2 + \text{pen}(m) \right) + 3\|\widehat{f}_m - f\|^2 \\
&\leq 18 \max_{m \leq j \leq N \wedge J} \left(\|\widehat{f}_j - f_j\|^2 - \frac{1}{3} \text{pen}(j) \right)_+ + 21\|\widehat{f}_m - f\|^2 + 6\text{pen}(m).
\end{aligned}$$

The Proposition 1 implies for all $m \in \{1, \dots, N \wedge J\}$,

$$\begin{aligned}
\mathbb{E}[\|\widehat{f}_{\widehat{m}} - f\|^2] &\leq 18 \sum_{j=m}^J \mathbb{E} \left[\left(\|\widehat{f}_j - f_j\|^2 - \text{pen}(j)/3 \right)_+ \right] + 21\|f_m - f\|^2 + 6 \text{pen}(m) + \frac{21\Delta}{2\pi} F_\epsilon \frac{m}{N} \\
&\leq 18 \sum_{j=m}^J \mathbb{E} \left[\left(\|\widehat{f}_j - f_j\|^2 - \text{pen}(j)/3 \right)_+ \right] + 21\|f_m - f\|^2 + \left(6 + \frac{21}{2\pi\kappa} \right) \text{pen}(m)
\end{aligned} \tag{19}$$

It remains to bound from above the term $\sum_{j=m}^{N \wedge J} \mathbb{E} \left[\left(\|\widehat{f}_j - f_j\|^2 - \text{pen}(j)/3 \right)_+ \right]$, this is the aim of Lemma 9.

Lemma 9. *There exists a constant C'' such that for all $m \in \mathcal{M}$,*

$$\sum_{j=m}^{N \wedge J} \mathbb{E} \left[\left(\|\widehat{f}_j - f_j\|^2 - \text{pen}(j)/3 \right)_+ \right] \leq \frac{C''}{N}.$$

Lemma 9 and inequality (19) imply that, for all $m \in \mathcal{M}$

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leq 21\|f_m - f\|^2 + \left(6 + \frac{21}{2\pi\kappa}\right) \text{pen}(m) + 18\frac{C''}{N}.$$

Thus, in particular

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leq C \inf_{m \in \mathcal{M}} (\|f_m - f\|^2 + \text{pen}(m)) + \frac{C'}{N}$$

with C and C' two constants. \square

6.5 Proof of Lemma 9

Consider $j \in \mathcal{M}$, and $t \in S_j$, define the empirical contrast:

$$\begin{aligned} \nu(t) &= \frac{1}{2\pi} \langle t^*, (\widehat{f}_j - f_j)^* \rangle \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} t^*(-u) \left(\frac{1}{N} \sum_{k=1}^N \frac{e^{iuZ_{k,j}}}{|f_\epsilon^*(\frac{u}{j\Delta})|^2} \mathbf{1}_{[-j\Delta, j\Delta]}(u) - f_j^*(u) \right) du \\ &= \frac{1}{2\pi} \frac{1}{N} \sum_{k=1}^N \int_{-j\Delta}^{j\Delta} t^*(-u) \frac{e^{iuZ_{k,j}} - \mathbb{E}[e^{iuZ_{k,j}}]}{|f_\epsilon^*(\frac{u}{j\Delta})|^2} du \\ &= \frac{1}{N} \sum_{k=1}^N \phi_t^j(Z_{k,j}) - \mathbb{E}[\phi_t^j(Z_{k,j})]. \end{aligned}$$

Let us define $A_j = \{t \in S_j, \|t\| = 1\}$, then $\|\widehat{f}_j - f_j\|^2 = \sup_{t \in A_j} |\nu(t)|^2$. We use Talagrand's inequality (see Proposition 8) to prove the lemma. We study first the upper bound:

$$\begin{aligned} \sup_{t \in A_j} \|\phi^j\|_\infty &\leq \sup_{t \in A_j} \frac{1}{2\pi} \int_{\mathbb{R}} |t^*(u)| \frac{\mathbf{1}_{[-j\Delta, j\Delta]}(u)}{|f_\epsilon^*(\frac{u}{j\Delta})|^2} du \\ &\leq \sup_{t \in A_j} \frac{1}{2\pi} \left(\int_{\mathbb{R}} |t^*(u)|^2 du \right)^{1/2} \left(\int_{-j\Delta}^{j\Delta} \frac{1}{|f_\epsilon^*(\frac{u}{j\Delta})|^4} du \right)^{1/2} \\ &\leq \sup_{t \in A_j} \frac{1}{\sqrt{2\pi}} \|t\| \left(\int_{-j\Delta}^{j\Delta} \frac{du}{|f_\epsilon^*(\frac{u}{j\Delta})|^4} \right)^{1/2} \\ &\leq \frac{1}{\sqrt{2\pi}} \left(\int_{-j\Delta}^{j\Delta} \frac{du}{|f_\epsilon^*(\frac{u}{j\Delta})|^4} \right)^{1/2} = \sqrt{\frac{\Delta}{2\pi}} \sqrt{jF_\epsilon} := M. \end{aligned} \tag{20}$$

Then according to Proposition 1, for all $j \in \mathcal{M}$

$$\mathbb{E}[\|\widehat{f}_j - f_j\|^2] \leq F_\epsilon \frac{j\Delta}{2\pi N},$$

thus

$$\mathbb{E}[\sup_{t \in A_j} |\nu(t)|^2] = \mathbb{E}[\|\widehat{f}_j - f_j\|^2] \leq F_\epsilon \frac{j\Delta}{2\pi N} := H^2. \tag{21}$$

Finally we examine the term:

$$\begin{aligned}
4\pi^2 \text{Var} \phi_t^j(Z_{k,j}) &\leq 4\pi^2 \mathbb{E} \left(\phi_t^j(Z_{k,j}) \overline{\phi_t^j(Z_{k,j})} \right) \\
&\leq \mathbb{E} \left[\left(\int_{-j\Delta}^{j\Delta} t^*(-u) \frac{e^{iuZ_{k,j}}}{|f_\epsilon^*(\frac{u}{j\Delta})|^2} du \right) \overline{\left(\int_{-j\Delta}^{j\Delta} t^*(-v) \frac{e^{ivZ_{k,j}}}{|f_\epsilon^*(\frac{v}{j\Delta})|^2} dv \right)} \right] \\
&= \mathbb{E} \left[\left(\int_{-j\Delta}^{j\Delta} t^*(-u) \frac{e^{iuZ_{k,j}}}{|f_\epsilon^*(\frac{u}{j\Delta})|^2} du \right) \left(\int_{-j\Delta}^{j\Delta} t^*(v) \frac{e^{-ivZ_{k,j}}}{|f_\epsilon^*(\frac{v}{j\Delta})|^2} dv \right) \right] \\
&= \mathbb{E} \left[\int_{-j\Delta}^{j\Delta} \int_{-j\Delta}^{j\Delta} t^*(-u) t^*(v) \frac{e^{i(u-v)Z_{k,j}}}{|f_\epsilon^*(\frac{u}{j\Delta})|^2 |f_\epsilon^*(\frac{v}{j\Delta})|^2} dudv \right].
\end{aligned}$$

Yet $\mathbb{E}(e^{i(u-v)Z_{k,j}}) = f_Z^*(u-v) = f^*(u-v) |f_\epsilon^*(\frac{u-v}{j\Delta})|^2$, it implies

$$4\pi^2 \text{Var}(\phi^j(Z_{k,j})) \leq \int_{-j\Delta}^{j\Delta} \int_{-j\Delta}^{j\Delta} t^*(-u) t^*(v) \frac{|f_\epsilon^*(\frac{u-v}{j\Delta})|^2}{|f_\epsilon^*(\frac{u}{j\Delta})|^2 |f_\epsilon^*(\frac{v}{j\Delta})|^2} f^*(u-v) dudv.$$

Under assumption **(H2)** we obtain:

$$\begin{aligned}
\sup_{u,v \in [-j\Delta, j\Delta]} \left| \frac{f_\epsilon^*(\frac{u-v}{j\Delta})}{f_\epsilon^*(\frac{u}{j\Delta}) f_\epsilon^*(\frac{v}{j\Delta})} \right|^2 &= \sup_{x,y \in [-1,1]} \left| \frac{f_\epsilon^*(x-y)}{f_\epsilon^*(x) f_\epsilon^*(y)} \right|^2 \leq \sup_{x,y \in [-1,1]} \left| \frac{1}{f_\epsilon^*(x) f_\epsilon^*(y)} \right|^2 \\
&\leq \sup_{x \in [-1,1]} \left| \frac{1}{f_\epsilon^*(x)} \right|^4 := C^2 < +\infty.
\end{aligned}$$

thus

$$\begin{aligned}
4\pi^2 \sup_{t \in A_j} \frac{1}{N} \sum_{k=1}^N \text{Var}(\phi^j(Z_{j,k})) &\leq \sup_{t \in A_j} \int_{-j\Delta}^{j\Delta} \int_{-j\Delta}^{j\Delta} C^2 |t^*(-u) t^*(v) f^*(u-v)| dudv \\
&\stackrel{C.S.}{\leq} \sup_{t \in A_j} 2\pi C^2 \|t\|^2 \left(\int_{-j\Delta}^{j\Delta} \int_{-j\Delta}^{j\Delta} |f^*(u-v)|^2 dudv \right)^{1/2} \\
&\leq 2\pi C^2 \left(\int_{-2j\Delta}^{2j\Delta} \int_{-j\Delta}^{j\Delta} |f^*(z)|^2 dudz \right)^{1/2} \\
&\leq C^2 4\pi \sqrt{\pi \Delta} \|f\| \sqrt{j} := 4\pi^2 v,
\end{aligned}$$

with

$$v := \frac{C^2 \|f\|}{\sqrt{\pi}} \sqrt{j\Delta}. \quad (22)$$

We achieve with the Talagrand's theorem, gathering the bounds (20), (21), (22), for $\alpha = 1/2$,

$$\begin{aligned}
\mathbb{E} \left[\left(\|\widehat{f}_j - f_j\|^2 - 4H^2 \right)_+ \right] &\leq \frac{4}{K_1} \left(\frac{C^2 \|f\| \sqrt{j\Delta}}{\sqrt{\pi} N} \exp \left(-\frac{K_1}{4} \frac{F_\epsilon}{\sqrt{\pi} C^2 \|f\|} \sqrt{j\Delta} \right) \right. \\
&\quad \left. + \frac{49\Delta F_\epsilon}{2\pi K_1 C^2(1/2)} \frac{j}{N^2} \exp \left(-\frac{K_1 C(1/2)}{7\sqrt{2}} \sqrt{N} \right) \right)
\end{aligned}$$

and

$$\mathbb{E} \left[\left(\|\widehat{f}_j - f_j\|^2 - 4H^2 \right)_+ \right] \leq C_1 C^2 \|f\| \sqrt{\Delta} \frac{\sqrt{j}}{N} \exp \left(-C_2 \frac{F_\epsilon \sqrt{\Delta}}{C^2 \|f\|} \sqrt{j} \right) + C_3 F_\epsilon \Delta \frac{j}{N^2} \exp \left(-C_4 \sqrt{N} \right)$$

where C_i are positive numerical constants. Then we sum this inequality if $\kappa \geq 24/\pi$

$$\begin{aligned}
\sum_{j=m}^J \mathbb{E} \left(\|\widehat{f}_j - f_j\|^2 - \text{pen}(j)/3 \right)_+ &\leq \sum_{j=m}^J \left(C_1 C^2 \|f\| \sqrt{\Delta} \frac{\sqrt{j}}{N} \exp \left(-C_2 \frac{F_\epsilon \sqrt{\Delta}}{C^2 \|f\|} \sqrt{j} \right) \right. \\
&\quad \left. + C_3 F_\epsilon \Delta \frac{j}{N^2} \exp \left(-C_4 \sqrt{N} \right) \right)
\end{aligned}$$

We remark that $1 \leq j \leq N \wedge J$ implies:

$$\begin{aligned} \sum_{j=m}^{N \wedge J} \mathbb{E} \left[\left(\|\widehat{f}_j - f_j\|^2 - \text{pen}(j)/3 \right)_+ \right] &\leq \frac{C_1 C^2 \|f\| \sqrt{\Delta}}{N} \sum_{j=1}^{+\infty} \sqrt{j} e^{-\left(C_2 \frac{F_\epsilon \sqrt{\Delta}}{\sigma^2 \|f\|}\right) \sqrt{j}} \\ &+ \frac{C_3 F_\epsilon \Delta}{N^2} \sum_{j=1}^{+\infty} j e^{-C_4 \sqrt{j}}. \end{aligned}$$

The sum $\sum_{j=0}^{+\infty} j e^{-C \sqrt{j}}$ converges $\forall C > 0$, there exists a constant C'' such as

$$18 \sum_{j=m}^{N \wedge J} \mathbb{E} \left(\|\widehat{f}_j - f_j\|^2 - \text{pen}(j)/3 \right)_+ \leq \frac{C''}{N}. \quad \square$$

6.6 Proof of Theorem 7

We also omit the subscript β , and the exponents (2) in the following.

For $t \in S_m$, and $m \in \mathcal{M}$, we consider the contrast γ (independent on m)

$$\begin{aligned} \gamma(t) &= \|t\|^2 - \frac{2}{N} \sum_{k=1}^N \frac{1}{2\pi} \int t^*(-u) \frac{e^{iuZ_{k,J}}}{|f_\epsilon^*(\frac{u}{J\Delta})|^2} du \\ &= \|t\|^2 - \frac{2}{N} \sum_{k=1}^N \phi_t^J(Z_{k,J}) \\ &= \|t\|^2 - 2 \langle t, \widehat{f}_m^J \rangle = \|t - \widehat{f}_m^J\|^2 - \|\widehat{f}_m^J\|^2. \end{aligned}$$

Obviously $\gamma(\widehat{f}_m^J) = -\|\widehat{f}_m^J\|^2$. We define

$$\nu(t) := \frac{1}{2\pi} \langle t^*, (\widehat{f}_m^J - f_m^J)^* \rangle = \frac{1}{N} \sum_{k=1}^N \phi_t^J(Z_{k,J}) - \mathbb{E}[\phi_t^J(Z_{k,J})]$$

then

$$(\gamma(t) - \|t - f\|^2) - (\gamma(s) - \|s - f\|^2) = -2\nu(t - s) \quad (23)$$

Thanks to the definition (13) of \widehat{m} : $\gamma(\widehat{f}_{\widehat{m}}^J) + \text{pen}(\widehat{m}) \leq \gamma(f_m) + \text{pen}(m)$. Using the equation (23) we obtain

$$\|\widehat{f}_{\widehat{m}}^J - f\|^2 \leq \|f_m - f\|^2 + 2\nu(\widehat{f}_{\widehat{m}}^J - f_m) + \text{pen}(m) - \text{pen}(\widehat{m}).$$

Let us remark that

$$\nu(\widehat{f}_{\widehat{m}}^J - f_m) = \|\widehat{f}_{\beta, \widehat{m}}^J - f_m\| \nu \left(\frac{\widehat{f}_{\widehat{m}}^J - f_m}{\|\widehat{f}_{\widehat{m}}^J - f_m\|} \right).$$

This leads us to

$$\begin{aligned} \|\widehat{f}_{\widehat{m}}^J - f\|^2 &\leq \|f_m - f\|^2 + 2\|\widehat{f}_{\widehat{m}}^J - f_m\| \nu \left(\frac{\widehat{f}_{\widehat{m}}^J - f_m}{\|\widehat{f}_{\widehat{m}}^J - f_m\|} \right) + \text{pen}(m) - \text{pen}(\widehat{m}) \\ &\leq \|f_m - f\|^2 + 2\|\widehat{f}_{\beta, \widehat{m}}^J - f_m\| \sup_{t \in B_{m, \widehat{m}}} \nu(t) + \text{pen}(m) - \text{pen}(\widehat{m}) \\ &\leq \|f_m - f\|^2 + \frac{1}{4} \|\widehat{f}_{\widehat{m}}^J - f_m\|^2 + 4 \sup_{t \in B_{m, \widehat{m}}} \nu^2(t) + \text{pen}(m) - \text{pen}(\widehat{m}) \\ &\leq 3\|f_m - f\|^2 + 8 \sup_{t \in B_{m, \widehat{m}}} \nu^2(t) + 2\text{pen}(m) - 2\text{pen}(\widehat{m}) \\ &\leq 3\|f_m - f\|^2 + 4\text{pen}(m) + 8 \left[\sup_{t \in B_{m, \widehat{m}}} \nu^2(t) - p(m, \widehat{m}) \right] \\ &\leq 3\|f_m - f\|^2 + 4\text{pen}(m) + 8 \sum_{m' \in \mathcal{M}} \left(\sup_{t \in B_{m, m'}} \nu^2(t) - p(m, m') \right)_+ \end{aligned}$$

with a function p such that $\forall m, m', 4p(m, m') \leq \text{pen}(m) + \text{pen}(m')$.

Lemma 10. *There exists a constant $C > 0$ such that*

$$\sum_{m' \in \mathcal{M}} \mathbb{E} \left[\left(\sup_{t \in B_{m,m'}} \nu^2(t) - p(m, m') \right)_+ \right] \leq \frac{C}{N}.$$

We conclude that there exist two constants $C_1, C_2 > 0$ such that

$$\mathbb{E}[\|\widehat{f_{\widehat{m}}} - f\|^2] \leq C_1 \inf_{m \in \mathcal{M}} \{\|f_m - f_\beta\|^2 + \text{pen}(m)\} + \frac{C_2}{N}.$$

6.7 Proof of the Lemma 10

We remind that $B_\epsilon = \sup_{v \in [-1,1]} \frac{1}{|f_\epsilon^*(v)|^4}$. For $m \in \mathcal{M}$ we consider $t \in S_m$. We use Talagrand's inequality to demonstrate Lemma 10 similarly to Lemma 9. We get:

$$M := \sqrt{\frac{\Delta B_\epsilon m^*}{\pi}}, \quad (24)$$

$$H := \sqrt{\frac{B_\epsilon \Delta m^*}{\pi N}}, \quad (25)$$

and

$$v = \frac{B_\epsilon^2 \|f\|}{\sqrt{\pi}} \sqrt{m^* \Delta}. \quad (26)$$

According to Talagrand's inequality, for $\alpha = 1/2$ there exist four numerical constants A_1, A_2, A_3, A_4 such that:

$$\begin{aligned} \mathbb{E} \left[\left(\sup_{t \in B_{m,m'}} \nu^2(t) - 4H^2 \right)_+ \right] &\leq A_1 B_\epsilon^2 \sqrt{\Delta} \frac{\sqrt{m^*}}{N} \exp \left(-A_2 \frac{\sqrt{\Delta}}{2B_\epsilon \|f\|} \sqrt{m^*} \right) \\ &\quad + A_3 \Delta B_\epsilon \frac{m^*}{N^2} \exp(-A_4 \sqrt{N}). \end{aligned}$$

Then we use that $1 \leq m^* \leq N \wedge J$, thus for $\kappa \geq 4/\pi$ we obtain there exists a constant $C > 0$ such that

$$\begin{aligned} \sum_{m' \in \mathcal{M}} \mathbb{E} \left[\left(\sup_{t \in B_{m,m'}} \nu^2(t) - p(m, m') \right)_+ \right] &\leq \sum_{m=1}^{+\infty} A_1 B_\epsilon^2 \frac{\sqrt{m\Delta}}{N} \exp \left(-A_2 \frac{\sqrt{m\Delta}}{2B_\epsilon \|f\|} \right) \\ &\quad + \sum_{m=1}^{+\infty} A_3 B_\epsilon \frac{m\Delta}{N^2} \exp(-A_4 \sqrt{m}) \\ &\leq \frac{C}{N} \end{aligned}$$

with $p(m, m') = 4\sqrt{\frac{B_\epsilon \Delta m^*}{\pi N}}$. \square

6.8 Estimator with unknown noise density

We compute an estimator of the density f_β when f_ϵ is unknown, corresponding to (5). We use the same idea as Comte and Samson (2012). However it is $1/|f_\epsilon^*(\frac{u}{m\Delta})|^2$ which has to be estimated and not $1/(f_\epsilon^*(\frac{u}{\Delta}))^2$. This quantity is more reachable and the noise does not need to be symmetric for the estimation. In this part we stand assumption **(H2)** for: **(H2')** f_ϵ is unknown and $f_\epsilon^* \neq 0$.

We consider the random variables

$$A_k := \Delta(3Z_{k,3} - 2Z_{k,2} - Z_{k,1}) = \epsilon_{k,3} - \epsilon_{k,2} - \epsilon_{k,1} + \epsilon_{k,0}.$$

These new variables are i.i.d. with a density f_A . Then we have

$$\forall u \in \mathbb{R}, f_A^*(u) = \mathbb{E}[e^{iAu}] = |f_\epsilon^*(u)|^4 = \mathbb{E}[\cos(Au)].$$

This leads to the unbiased empirical estimator of $|f_\epsilon^*|^4$:

$$\widehat{|f_\epsilon^*|^4}(u) = \frac{1}{N} \sum_{k=1}^N \cos(A_k u).$$

Nevertheless, this function can reach zero and has to be inverted. We define a truncated estimator following Comte and Lacour (2011) and we propose the estimator of $|1/f_\epsilon^*|^2$ following:

$$\frac{1}{\widehat{|f_\epsilon^*|^2}(u)} = \frac{\mathbf{1}_{\widehat{|f_\epsilon^*|^4}(u) \geq N^{-1/2}}}{[\widehat{|f_\epsilon^*|^4}(u)]^{1/2}}. \quad (27)$$

The new estimator of f_β is

$$\forall x \in \mathbb{R}, \widehat{f_{\beta,m}}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{1}{N} \sum_{k=1}^N \frac{e^{iuZ_{k,m}}}{\widehat{|f_\epsilon^*|^2}\left(\frac{u}{m\Delta}\right)} du. \quad (28)$$

Proposition 11. *If $m > 3$, and denoting $K_\epsilon := 5F_\epsilon + 68G_\epsilon$, where $G_\epsilon := \int_{-1}^1 \frac{du}{|f_\epsilon^*(u)|^8}$, under (H1)-(H2') we have*

$$\mathbb{E}[\|\widehat{f_{\beta,m}} - f_\beta\|^2] \leq \|f_{\beta,m} - f_\beta\|^2 + K_\epsilon \frac{m\Delta}{\pi N}.$$

Proof

We denote

$$R(u) := \frac{1}{\widehat{|f_\epsilon^*|^2}(u)} - \frac{1}{|f_\epsilon^*(u)|^2}$$

which is the subject of the following lemma. This lemma is an extension of the Neumann's Lemma (Neumann (1997)), demonstrated in Comte and Samson (2012) with a symmetric noise. The proof can be straightforwardly adapted and thus is omitted.

Lemma 12.

$$\mathbb{E} \left[\left(\frac{1}{\widehat{|f_\epsilon^*|^2}(u)} - \frac{1}{|f_\epsilon^*(u)|^2} \right)^2 \right] \leq \frac{2}{|f_\epsilon^*(u)|^4} \wedge \frac{3N^{-1/2}}{|f_\epsilon^*(u)|^8} \wedge \frac{34N^{-1}}{|f_\epsilon^*(u)|^{12}}.$$

We notice that $R(u)$ and $\widehat{f_{Z_m}^*} = 1/N \sum_{k=1}^N e^{iuZ_{k,m}}$ are independent as soon as $m > 3$. Let us decompose the norm: $\|\widehat{f_{\beta,m}} - f_\beta\|^2$.

$$\|\widehat{f_{\beta,m}} - f_\beta\|^2 = \|\widehat{f_{\beta,m}} - f_{\beta,m}\|^2 + \|f_{\beta,m} - f_\beta\|^2 \leq 2\|\widehat{f_{\beta,m}} - \widehat{f_{\beta,m}}\|^2 + 2\|\widehat{f_{\beta,m}} - f_{\beta,m}\|^2 + \|f_{\beta,m} - f_\beta\|^2.$$

The first member of the right hand side is simplified by Plancherel-Parseval's Theorem

$$\begin{aligned} \|\widehat{f_{\beta,m}} - \widehat{f_{\beta,m}}\|^2 &= \left\| \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \widehat{f_{Z_m}^*}(u) R\left(\frac{u}{m\Delta}\right) du \right\|^2 \\ &= \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} |\widehat{f_{Z_m}^*}(u) R\left(\frac{u}{m\Delta}\right)|^2 du \\ &\leq \frac{1}{\pi} \int_{-m\Delta}^{m\Delta} |\widehat{f_{Z_m}^*}(u) - f_{Z_m}^*(u)|^2 |R\left(\frac{u}{m\Delta}\right)|^2 du \\ &+ \frac{1}{\pi} \int_{-m\Delta}^{m\Delta} |f_{Z_m}^*(u)|^2 |f_\epsilon^*\left(\frac{u}{m\Delta}\right)|^4 |R\left(\frac{u}{m\Delta}\right)|^2 du \end{aligned}$$

note that

$$\mathbb{E} \left[|\widehat{f_{Z_m}^*}(u) - f_{Z_m}^*(u)|^2 \right] = \text{Var} \left(\widehat{f_{Z_m}^*}(u) - f_{Z_m}^*(u) \right) = \text{Var} \left(\frac{1}{N} \sum_{k=1}^N e^{iuZ_{k,m}} \right) \leq \frac{1}{N}.$$

Then, using Lemma 12, we obtain

$$\begin{aligned}
\mathbb{E}[\|\widehat{\widehat{f_{\beta,m}}} - \widehat{f_{\beta,m}}\|^2] &\leq \frac{1}{\pi} \int_{-m\Delta}^{m\Delta} \frac{1}{N} \mathbb{E}[|R\left(\frac{u}{m\Delta}\right)|^2] du + \frac{1}{\pi} \int_{-m\Delta}^{m\Delta} |f_{\beta}^*(u)|^2 \frac{3N^{-1/2}}{|f_{\epsilon}^*\left(\frac{u}{m\Delta}\right)|^4} \wedge \frac{34N^{-1}}{|f_{\epsilon}^*\left(\frac{u}{m\Delta}\right)|^8} du \\
&\leq \frac{2}{\pi N} \int_{-m\Delta}^{m\Delta} \frac{1}{|f_{\epsilon}^*\left(\frac{u}{m\Delta}\right)|^4} du + \frac{1}{\pi} \left(\frac{3}{\sqrt{N}} \int_{-m\Delta}^{m\Delta} \frac{|f_{\beta}^*(u)|^2}{|f_{\epsilon}^*\left(\frac{u}{m\Delta}\right)|^4} du \right) \wedge \left(\frac{34}{N} \int_{-m\Delta}^{m\Delta} \frac{|f_{\beta}^*(u)|^2}{|f_{\epsilon}^*\left(\frac{u}{m\Delta}\right)|^8} du \right) \\
&\leq \frac{2m\Delta}{\pi N} F_{\epsilon} + 34 \frac{m\Delta}{\pi N} G_{\epsilon}.
\end{aligned}$$

Finally, using Proposition 1, it yields

$$\mathbb{E}[\|\widehat{\widehat{f_{\beta,m}}} - f_{\beta}\|^2] \leq \|f_{\beta,m} - f_{\beta}\|^2 + \frac{m\Delta}{\pi N} F_{\epsilon} + 4 \frac{m\Delta}{\pi N} F_{\epsilon} + 68 \frac{m\Delta}{\pi N} G_{\epsilon}. \square$$

The same construction can be used to propose a version of $\widehat{f_{\beta,m}^J}$ with estimated noise density. Adopting the same arguments we consider then the new estimator of f_{β}

$$\forall x \in \mathbb{R}, \widehat{\widehat{f_{\beta,m}^J}}(x) = \frac{1}{2\pi} \int_{-\Delta m}^{\Delta m} e^{-iux} \frac{1}{N} \sum_{k=1}^N \frac{e^{iuZ_{k,J}}}{\widehat{|f_{\epsilon}^*|^2}\left(\frac{u}{J\Delta}\right)} du. \quad (29)$$

Using Corollary 6 and Lemma 12 we obtain the following risk bound.

Proposition 13. *If $3 < m \leq N \wedge J$, and denoting $D_{\epsilon} := 4B_{\epsilon} + 136C_{\epsilon}$, where $C_{\epsilon} := \sup_{u \in [-1,1]} \frac{1}{|f_{\epsilon}^*(u)|^8}$, under (H1)-(H2') we have*

$$\mathbb{E}[\|\widehat{\widehat{f_{\beta,m}}} - f_{\beta}\|^2] \leq \|f_{\beta,m} - f_{\beta}\|^2 + D_{\epsilon} \frac{m\Delta}{\pi N}.$$

References

- Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329–375.
- Butucea, C. and Tsybakov, A. (2007). Sharp optimality in density deconvolution with dominating bias. ii. *Teor. Veroyatnost. i Primenen.* **52**, 336–349.
- Comte, F., Genon-Catalot, V. and Samson, A. (2013). Nonparametric estimation for stochastic differential equation with random effects. *Stochastic Process. Appl.* **7**, 2522–2551.
- Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. Roy. Stat. Soc. B* **73**, 601–627.
- Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.* **34**, 431–452.
- Comte, F. and Samson, A. (2012). Nonparametric estimation of random-effects densities in linear mixed-effects model. *J. Nonparametr. Statist.* **24**, 951–975.
- Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.* **56**, 19–47.
- Donoho, D. L., Johnstone, I. M., Kerkycharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 183–218.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.
- Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608–1632.

- Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution (work in progress).
- Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33**, 1060–1077.
- Komárek, A. and Lesaffre, E. (2008). Generalized linear mixed model with a penalized gaussian mixture as a random effects distribution. *Comput. Statist. Data Anal* **52**, 3441–3458.
- Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342**, 877–882.
- Massart, P. (2007). Concentration inequalities and model selection. In *Lectures on Probability Theory and Statistics, Ecole d'Été de Probabilité de St-Flour 2003*. Springer-Verlag, Berlin.
- Meister, A. and Neumann, M. H. (2010). Deconvolution from non-standard error densities under replicated measurements. *Statist. Sinica* **20**, 1609–1636.
- Neumann, M. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.* **7**, 307–330.
- Papageorgiou, G. and Hinde, J. (2012). Multivariate generalized linear mixed models with semi-nonparametric and smooth nonparametric random effects densities. *Stat. Comput.* **22**, 79–92.
- Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**, 2033–2053.
- Pinheiro, J. and Bates, D. (2000). *Mixed-effect models in S and Splus*. Springer-Verlag, New York.
- Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348–1360.
- Tsybakov, A. (2004). *Introduction à l'estimation non-paramétrique*. Springer, Berlin.
- Wu, P. and Zhu, L. (2010). An orthogonality-based estimation of moments for linear mixed models. *Ann. Statist.* **37**, 253–263.