DECONVOLUTION ESTIMATION OF ONSET OF PREGNANCY WITH REPLICATE OBSERVATIONS

FABIENNE COMTE[1], ADELINE SAMSON[1], AND JULIEN J STIRNEMANN[1,2]

[1] *MAP5, UMR CNRS 8145, Université Paris Descartes*
[2] *Obstetrics and Maternal - Fetal Medicine, GHU Necker-Enfants Malades, Université ParisDescartes.*

Abstract. In general, the precise date of onset of pregnancy is unknown and may only be estimated from ultrasound biometric measurements of the embryo. We want to estimate the density of the random variables corresponding to the interval between last menstrual period and true onset of pregnancy. The observations correspond to the variables of interest up to an additive noise. We suggest an estimation procedure based on deconvolution. It requires the knowledge of the density of the noise which is not available. But we have at our disposal another specific sample with replicate observations for twin pregnancies. This allows both to estimate the noise density and to improve the deconvolution step. Convergence rates of the final estimator are studied and compared to other settings. Our estimator involves a cut-off parameter for which we propose a cross-validation type procedure. Lastly, we estimate the target density in spontaneous pregnancies with an estimation of the noise obtained from replicate observations in twin pregnancies.

Keywords. Deconvolution; Density estimation; Nonparametric methods; Dating of pregnancy; Mean square risk; Replicate observations

1. Introduction

In spontaneously conceived pregnancies, the date of pregnancy is unknown. Although pregnancies occur at around 14 days following last menstrual period (LMP), the fertile window of a woman may vary widely based upon hormonal studies (Wilcox et al. [2000]). These studies, however, provide day-specific probabilities of a fertile window within a female cycle in non-pregnant women and not the probability density of onset of pregnancy in pregnant women. Since the exact date of pregnancy is never precisely known in women conceiving spontaneously, the probability distribution function of onset of pregnancy within female cycles is unknown in the general population. This density, however, may have important implications both for clinical practice and physiology knowledge.

Ultrasound is the most widely used method for dating pregnancies in clinical practice. First trimester biometric measurements such as the crown-rump length (CRL) have been proven to perform better than LMP for dating pregnancies. Several formulas, derived from simple regression analysis have been developed for dating pregnancies (Sladkevicius et al. [2005]), the most widely used being the formula initially suggested by Robinson [1973]. Denoting by X the interval between LMP and true onset of pregnancy, and by Y the interval between LMP and ultrasound estimate, the purpose of this study is to estimate the density f of X. However, only the noisy observations

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n \tag{1}$$

are available. Here, the X_j and the ε_j for $j = 1, \dots, n$ are independent identically distributed and the sequences $(X_j)_{1\le j\le n}$ and $(\varepsilon_j)_{1\le j\le n}$ are independent. Moreover, in this setting the density f_ε of ε is unknown and the noise cannot be directly estimated from a preliminary sample of ε_j.

Since X is measured with an unknown error, the estimation of f may be seen as a deconvolution problem. Regarding the assumptions on the distribution of the error, several approaches have been studied in the literature. Numerous works have addressed this problem under the assumption of a known density for the error. These works comprise kernel methods (see Fan [1991], Liu and Taylor [1989], Stefanski and Carroll [1990], Hesse [1999], Delaigle and Gijbels [2004]) as well as wavelet methods (see Fan and Koo [2002], Pensky and Vidakovic [1999]). Minimax optimality of convergence rates have been studied by Fan [1991], Butucea [2004], Butucea and Tsybakov [2008a]. When a sample of the error is given, density estimation has been addressed by Diggle and Hall [1993] and Neumann [1997]. The latter considers the case of ordinary smooth densities for both densities of the error and X, and provides minimax rates of convergence. More contributions by Johannes [2009] and Comte and Lacour [2011] propose different approaches with regard to bandwidth selection. A full scheme of estimation in this setting with data-driven bandwidth selection may be found in Comte and Lacour [2011].

In this article, we consider yet a different setting in which neither a known density nor a sample of noise are available. Rather, we consider the situation of replicate and noisy observations of the random variable X. Consider we have a sample of pregnancies with two replicate measurements of X:

$$Y_{n+j,1} = X_{n+j} + \varepsilon_{n+j,1}, \quad Y_{n+j,2} = X_{n+j} + \varepsilon_{n+j,2}, \quad j = 1, \dots, M \tag{2}$$

with X_{n+j}, $\varepsilon_{n+j,1}$ and $\varepsilon_{n+j,2}$, for $j = 1, \dots, M$, independent and identically distributed. The sequences $(X_{n+j})_{1\le j\le M}$, $(\varepsilon_{n+j,1})_{1\le j\le M}$ and $(\varepsilon_{n+j,2})_{1\le j\le M}$ are independent. These noisy observations could be replicate measurements of CRL of the same embryo or measurements of CRL in twin pregnancies. Therefore, we consider that two independent samples are available: the first, of size M, containing replicate observations and the second, of size n containing non-replicate observations. Density estimation by deconvolution with replicate observations has been studied by Delaigle et al. [2008], Li and Vuong [1998] and Meister and Neumann [2009]. Our approach suggests an estimator that is related to the truncated estimator of Neumann [1997]. The second sample with replicate observations allows both to estimate the noise density and to improve the deconvolution step. Our estimator involves a cut-off parameter for which we propose a cross-validation type procedure. We also provide the first step towards the theoretical justification of an adaptive procedure.

The outline of this article is as follows. In Section 2 we define our estimator. We then majorate the $\mathbb{L}^2$ risk based upon a new version of the fundamental lemma of Neumann [1997]. Convergence rates are compared to the settings of known noise density and observed noise samples. We discuss the relationship between M, n and the resulting convergence rates and show that the convergence rate is the same as that found with an assumed known noise density in several cases. In Section 4 we discuss model selection and the choice of an appropriate penalty. Simulations are conducted to illustrate the performance of our estimator together with a comparison with existing results in Section 5. Finally, in Section 6, we apply our method to real data and estimate the distribution f of onset of pregnancy within a female cycle using ultrasound measurements in twin pregnancies as replicate noisy observations.

2. Model and estimator

We denote f_Y, f and f_ε the densities of Y, X and ε. We denote by $g^*(x) = \int e^{itx} g(x)dx$, the Fourier transform of any integrable function g. The characteristic functions of each of the variables Y, X and ε are therefore denoted f_Y^*, f^* and f_ε^* respectively. For a function $g : \mathbb{R} \mapsto \mathbb{R}$, we denote by $\|g\|^2 = \int_{\mathbb{R}} g^2(x)dx$ the $\mathbb{L}^2$ norm. For two real numbers a and b, we denote $a \wedge b = \min(a, b)$. As a rule in this paper, unless otherwise specified, C and C' will denote universal constants that may change from line to line.

In the following, we consider the model described by the two independent samples (1) and (2). The convolution problem may be written as $f_Y(x) = f \star f_\varepsilon(x) = \int f(x-y)f_\varepsilon(y)dy$ where $\star$ denotes the convolution operator. Using the characteristic functions, we have $f_Y^*(u) = f^*(u)f_\varepsilon^*(u)$. Fourier inversion of $f^* = f_Y^*/f_\varepsilon^*$ can then be used to propose an estimator of f.

In the case of a known density f_ε of the noise, Comte et al. [2006] propose an estimate of f based on this idea and using a πm cut-off for integrability purpose. More precisely, they consider the following estimator:

$$\frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} du, \tag{3}$$

where $\hat{f}_Y^*$ is the empirical characteristic function based on the observations of Y. The risk bound obtained in this case is considered as a benchmark of the best reachable bound.

In our setting we consider f_ε unknown, therefore we must estimate it or at least the square of its characteristic function f_ε^* as we will see shortly. We suggest that the estimation of f_ε^* relies upon replicate observations given in (2). Because of replications, we can only estimate $(f_\varepsilon^*)^2$ and not directly f_ε^* as when a noise sample is available.

The following preliminary assumption regarding the behavior of f_ε will be considered fulfilled throughout the article.

Assumption(A1) We assume ε is symmetric and that its characteristic function never vanishes.

This assumption seems very acceptable in our context where errors have no reason to be rather positive than negative. We emphasize that it implies simplification of the problem. For non necessarily symmetric errors, the reader is referred to Li and Vuong [1998] where a solution in the general context is proposed and may be adapted to our setting.

Assuming ε symmetric is equivalent to assuming f_ε^* real-valued. Therefore, Assumption **(A1)** implies that

$$\forall u \in \mathbb{R}, f_\varepsilon^*(u) \in \mathbb{R}_+^*.$$

Under this reasonable assumption, we have:

$$\mathbb{E}\big(e^{iu(\varepsilon_{n+j,1}-\varepsilon_{n+j,2})}\big) = |\mathbb{E}\big(e^{iu\varepsilon_{n+j,1}}\big)|^2 \overset{\text{(A1)}}{=} \Big(\mathbb{E}\big(e^{iu\varepsilon_{n+j,1}}\big)\Big)^2 = \big(f_\varepsilon^*(u)\big)^2.$$

Therefore, given that $\mathbb{E}\big(e^{iu(\varepsilon_{n+j,1}-\varepsilon_{n+j,2})}\big) = \mathbb{E}\big(e^{iu(Y_{n+j,1}-Y_{n+j,2})}\big)$ and under the hypothesis **(A1)**, we have the following estimation of $(f_\varepsilon^*)^2$:

$$\widehat{(f_\varepsilon^*)^2}(u) = \frac{1}{M}\sum_{j=1}^{M} \cos\big(u(Y_{n+j,1} - Y_{n+j,2})\big). \tag{4}$$

Our definition does not involve absolute value nor positive part since our estimator of f_ε^* is used over a domain for which $\widehat{(f_\varepsilon^*)^2}$ is positive. Indeed, we define a truncated estimate of $1/f_\varepsilon^*$:

$$\frac{1}{\tilde{f}_\varepsilon^*(u)} = \frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)\geq M^{-1/2}}}{\sqrt{\widehat{(f_\varepsilon^*)^2}(u)}}. \tag{5}$$

Fourier inversion of $f^* = f_Y^*/f_\varepsilon^*$ yields the following estimator of f, when still using a πm cut-off for integrability purpose :

(6)

$$\hat{f}_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-ixu}\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)}du, \text{ where } \hat{f}_Y^*(u) = \frac{1}{n+M}\left(\sum_{j=1}^{n} e^{iuY_j} + \sum_{j=1}^{M} e^{iuY_{n+j,1}}\right).$$

This estimator can also be viewed as a deconvolution kernel estimator with the sinc kernel and the bandwidth $1/(\pi m)$.

Remark 1. *Data $(Y_{n+j,2})_{1\leq j\leq M}$ are not used in $\hat{f}_Y^*(u)$. Their use would slightly decrease the main variance term (see Appendix 8.1). But this would induce several technicalities in the proofs due to the dependency between $Y_{n+j,1}$ and $Y_{n+j,2}$.*

The level of truncation required in (5) differs from the case where a noise sample is available. As already said, because of replications, we can only estimate $(f_\varepsilon^*)^2$ and not directly f_ε^*. Therefore the truncation is in $M^{-1/4}$ in (5), instead of $M^{-1/2}$.

The estimator $\hat{f}_m(x)$ differs from the estimator proposed by Delaigle et al. [2008] in several ways. A cut-off πm is used instead of a ridge parameter. This cut-off allows to consider super-smooth densities f_ε and f, which is not the case with the ridge parameter. The cut-off use yields to restrict to the sinc kernel in order to optimize the bias of the estimator. We only use the estimation of f_ε^* where $\widehat{(f_\varepsilon^*)^2}$ is non-negative so the absolute value used by Delaigle et al. [2008] is not required. This substantially simplifies the theoretical study.

3. Upper bound of the $\mathbb{L}^2$ risk

Let us define f_m such that $f_m^* = f^*\mathbf{1}_{[-\pi m,\pi m]}(.)$. The function f_m is the function which is in fact estimated by $\hat{f}_m$. Therefore, this implies a nonparametric bias measured by the distance between f and f_m.We wish to bound the mean integrated squared error (MISE) defined as $\mathbb{E}(\|f-\hat{f}_m\|^2)$. We will first generalize Neumann's Lemma (Neumann [1997]) to the case of replicate measurements and use this result to deduce a risk bound.

3.1. **General MISE bound.** The extension of Neumann's lemma for replicate measurements is

Lemma 1. *Assume that* **(A1)** *holds, and let p be an integer, $p \geq 1$. There exists a constant C_p such that*

$$\mathbb{E}\left(\left|\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)}\right|^{2p}\right) \leq C_p\left(\frac{2}{(f_\varepsilon^*)^2(u)} \wedge \frac{3M^{-1/2}}{(f_\varepsilon^*)^4(u)} \wedge \frac{6M^{-1}}{(f_\varepsilon^*)^6(u)}\right)^p,$$

with $C_1 = 1$.

Proof is given in appendix.

Let us define

$$\Delta_2(m) = \int_{-\pi m}^{\pi m} \frac{du}{(f_\varepsilon^*)^2(u)} \text{ and, for } k=2,4,\ \Delta_k^{(f)}(m) = \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{(f_\varepsilon^*)^k(u)} du.$$

Using the previous lemma we may deduce the following bound for the MISE:

Proposition 1. *Assume that* **(A1)** *holds and $\hat{f}_m$ is defined by (6). Then there exists a constant C such that*

$$\mathbb{E}(\|f-\hat{f}_m\|^2) \leq \|f-f_m\|^2 + C\left(\frac{\Delta_2(m)}{n+M} + \frac{\Delta_2^{(f)}(m)}{\sqrt{M}} \wedge \frac{\Delta_4^{(f)}(m)}{M}\right). \quad (7)$$

Proof is given in appendix.

This decomposition is non asymptotic. It underlines the different terms involved in the bound of the integrated risk. We recognize in inequality (7) the bias $\|f-f_m\|^2$ and variance denoted $\mathrm{Var}(m) := Q_1(m) + Q_2(m)$ with

$$Q_1(m) := \Delta_2(m)/(n+M)$$
$$Q_2(m) := \Delta_2^{(f)}(m)/\sqrt{M} \wedge \Delta_4^{(f)}(m)/M.$$

We can also recognize $Q_1(m)$ as the variance term that arises alone when f_ε^* is assumed as known with a Y-sample size $n+M$. The following term $Q_2(m)$ is specific to our setting involving replicate observations and shows the loss in the resulting rates. In the case of observed noise, the upper bound is

$$\|f-f_m\|^2 + C\Delta_2(m)/n + (C+2)\Delta_2^{(f)}(m)/M,$$

when the Y-sample has size n and the ε-sample has size M. Moreover, for $M \geq n$, the variance term is bounded by $C\Delta_2(m)/n$, which corresponds to known noise density for sample size equal to n. Therefore, compared to the case with observed noise or known density, the term $Q_2(m)$ resulting from the estimation of $(f_\varepsilon^*)^2$ with replicate observations is a substantial step in complexity.

3.2. **Resulting rates.** Let us consider the following classical assumptions regarding the behavior of f_ε^*:

Assumption (A2) There exist $\alpha \geq 0, \beta > 0$, $\gamma \in \mathbb{R}$ ($\gamma > 0$ if $\alpha = 0$) and $k_0, k_1 > 0$ such that, $\forall u \in \mathbb{R}$,

$$k_0(u^2+1)^{-\alpha/2}\exp(-\beta|u|^\gamma) \leq |f_\varepsilon^*(u)| \leq k_1(u^2+1)^{-\alpha/2}\exp(-\beta|u|^\gamma)$$

The noise distribution is called ordinary smooth if $\gamma = 0$ and super smooth otherwise. A Gaussian noise is super smooth with $\gamma = 2$ and a Laplace noise is ordinary smooth with $\gamma = 0$ and $\alpha = 2$.

Now, we know that, under **(A2)**, the dominating variance term has the following order:

$$Q_1(m) = \frac{\Delta_2(m)}{n+M} \leq C\frac{m^{2\alpha+1-\gamma}\exp(2\beta(\pi m)^\gamma)}{n+M}.$$

Such orders are non standard for variance terms in nonparametric estimation and in particular larger than orders $m/(n+M)$ which are obtained for standard problems (e.g. density estimation without noise, corresponding to $\alpha = \gamma = 0$, $M = 0$).

If we want to give precise examples of the rates that can be obtained in the deconvolution context, we must also make assumptions on the rate of decrease of f^*. Classically, we consider the following smoothness spaces for density f on $\mathbb{R}$:

Assumption (A3) $f \in \mathcal{A}_{a,b,c}(l) = \{f \in \mathbb{L}^1 \cap \mathbb{L}^2,\ \int |f^*(u)|^2(u^2+1)^a\exp(2b|u|^c)du \leq l\}$ with $c \geq 0, b > 0$, $a \in \mathbb{R}$ ($a > 1/2$ if $c = 0$), $l > 0$.

As previously, when $c > 0$, the function f is known as super smooth, and as ordinary smooth otherwise. The spaces of ordinary smooth functions correspond to classical Sobolev classes, while super smooth functions are infinitely differentiable (analytic function), and we have necessarily $c \leq 2$. It includes for example Gaussian ($c = 2$) and Cauchy ($c = 1$) densities.

Then, under **(A3)**, we have the following bias order:

$$\|f - f_m\|^2 \leq Cm^{-2a}\exp(-2b(\pi m)^c). \tag{8}$$

Moreover, under both **(A3)** and **(A2)**, we have

$$Q_2(m) \leq C\left(\frac{m^{2(\alpha-a)_+}e^{(2\beta(\pi m)^\gamma)-2b(\pi m)^c)_+}}{\sqrt{M}}\right) \wedge \left(\frac{m^{2(2\alpha-a)_+}e^{(4\beta(\pi m)^\gamma)-2b(\pi m)^c)_+}}{M}\right).$$

Therefore, we have the following results when both f and f_ε are ordinary smooth.

Proposition 2. *We consider assumptions* **(A1)**, **(A2)**, **(A3)** *and the ordinary smooth case for both f and f_ε with $c = \gamma = 0$. The bound (7) then becomes*

$$\mathbb{E}(\|f - \hat{f}_m\|^2) \leq C\left(m^{-2a} + \frac{m^{2\alpha+1}}{n+M} + \frac{m^{2(\alpha-a)_+}}{\sqrt{M}} \wedge \frac{m^{2(2\alpha-a)_+}}{M}\right).$$

If moreover $M \geq n$ and $a \geq \alpha - 1/2$, then we have for $m_{opt} = M^{1/(2a+2\alpha+1)}$,

$$\mathbb{E}(\|f - \hat{f}_{m_{opt}}\|^2) \leq CM^{-2a/(2a+2\alpha+1)} \tag{9}$$

If $M = n^\omega$ with $\omega < 1$, $m_{opt} = n^{1/(2a+2\alpha+1)}$, and if

$$2\alpha \leq a \leq \frac{\omega}{1-\omega}(\alpha + \frac{1}{2}), \tag{10}$$

then

$$\mathbb{E}(\|f - \hat{f}_{m_{opt}}\|^2) \leq Cn^{-2a/(2a+2\alpha+1)}. \tag{11}$$

The first inequality is a consequence of (7), (8) and elementary bounds for the variance. Inequalities (9) and (11) allow us to recover the deconvolution rates as if the noise had known density for a sample size of observations of order M and n respectively. This rate is known to be the optimal one when the noise density is assumed to be known (see Fan [1991], Butucea [2004], Butucea and Tsybakov [2008a,b]). The first case was already mentioned in Delaigle et al. [2008]. The case $M < n$ is new and interesting as it corresponds to our real data setting. Note that condition (10) reduces to $a \geq 2\alpha$ when ω tends to 1. Another case is generally considered, where still $c = 0$ but $\gamma > 0$:

Proposition 3. *We consider assumptions* **(A1)**, **(A2)**, **(A3)** *and the ordinary smooth case for f while f_ε is super smooth: $c = 0$, $\gamma > 0$. Then if $M \geq n$ and $m = m_{opt} = \pi^{-1}(\log(M)/(8\beta))^{1/\gamma}$, we have*

$$\mathbb{E}(\|f - \hat{f}_{m_{opt}}\|^2) \leq C[\log(M)]^{-2a/\gamma}. \tag{12}$$

The rate in (12) is the optimal rate in this context. Unfortunately, it is logarithmic, but practical experiments show that, nevertheless, the procedure works well. It happens that, when the noise density is known, the choice of the optimal cutoff m_{opt} is known also. In our case, an adaptive procedure would be required in this case as well, since β, γ are unknown. This enhances the interest of adaptive procedures in general.

Proposition 3 is often summarized as: "when the noise is super-smooth, the rate of deconvolution is logarithmic". This is not true since the regularity parameters of f are involved in the computation of the rate. Let us give two counterexamples.

Example 1. Consider $M = n$, $a = \alpha = 0$ and $2b\pi = 2\beta\pi = 1$, $c = \gamma = 1$. We have

$$\mathbb{E}(\|f - \hat{f}_m\|^2) \leq C\left(e^{-m} + \frac{e^m}{M} + \frac{1}{\sqrt{M}} \wedge \frac{e^m}{M}\right).$$

Then the choice $m_{opt} = \log(M)/2$ yields

$$\mathbb{E}(\|f - \hat{f}_{m_{opt}}\|^2) \leq CM^{-1/2}$$

which is also still an optimal deconvolution rate in this case, and is obviously much better than logarithmic.

Example 2: Gaussian-Gaussian case. Consider $M = n$ and $a = \alpha = 0$ and $2b = 2\beta = 1$, $c = \gamma = 2$. In that case, the risk bound can be written

$$\mathbb{E}(\|f - \hat{f}_m\|^2) \leq C \left(e^{-(\pi m)^2} + \frac{m^{-1} e^{(\pi m)^2}}{M} + \frac{1}{\sqrt{M}} \wedge \frac{e^{(\pi m)^2}}{M} \right).$$

Then, choose

$$\pi m_{opt} = \left(\frac{1}{2} \log(M) + \frac{1}{4} \log(\log(M)) \right)^{1/2},$$

so that the risk is bounded by

$$\mathbb{E}(\|f - \hat{f}_{m_{opt}}\|^2) \leq C (\log(M))^{-1/4} M^{-1/2}$$

which is still a rate better than logarithmic for a sample size of order M.

We can give a more general result.

Proposition 4. *Assume that assumptions* **(A1)**, **(A2)**, **(A3)** *are fulfilled with $c > \gamma > 0$. If $M \geq n$, then there exists a constant C such that*

$$\mathbb{E}(\|f - \hat{f}_m\|^2) \leq C \left(m^{-2a} \exp(-2b(\pi m)^c) + \frac{m^{2\alpha+1-\gamma} e^{2\beta(\pi m)^\gamma}}{M} (1 + \frac{m^{2\alpha} e^{2\beta(\pi m)^\gamma}}{\sqrt{M}}) \right).$$

Consequently, for any ϵ, $0 < \epsilon < 1/2$, the choice $\pi m_0 = (\epsilon \log(M)/(2\beta))^{1/\gamma}$ yields

$$\mathbb{E}(\|f - \hat{f}_{m_0}\|^2) \leq C [\log(M)]^{(2\alpha+1-\gamma)/\gamma} M^{-1+\epsilon}.$$

The last inequality also presents a general context where a polynomial rate can be reached in the super smooth case.

4. Cross-validation cut-off selection

4.1. **General principle.**

Let us give the general principle for the automatic selection of the cutoff m. First, assume that m belongs to a set of admissible values

$$\mathcal{M} = \{m \in \mathbb{N};\ \mathrm{Var}(m) \leq C_0\}$$

for fixed C_0. The general outline of the method used to select among all considered indexes $\mathcal{M}$ is borrowed from Comte and Lacour [2011]. Our approach aims to select $m \in \mathcal{M}$ based upon an adequate bias-variance compromise. First, notice that in our case $\|f - f_m\|^2 = \|f\|^2 - \|f_m\|^2$. Indeed, denoting $\langle \cdot, \cdot \rangle$ the scalar product, we have

$$\begin{aligned} \|f - f_m\|^2 &= \|f\|^2 - 2\langle f^*, f_m^* \rangle + \|f_m\|^2 \\ &= \|f\|^2 - 2 \int f^*(u)(\bar{f}^* \mathbf{1}_{[-\pi m, \pi m]})(u) du + \|f_m\|^2 = \|f\|^2 - \|f_m\|^2. \end{aligned}$$

The theoretical optimal choice of m is defined as:

$$m^{th} = \underset{m\in\mathcal{M}}{\operatorname{argmin}}\Big(\|f - f_m\|^2 + \mathrm{Var}(m)\Big) = \underset{m\in\mathcal{M}}{\operatorname{argmin}}\Big(-\|f_m\|^2 + \mathrm{Var}(m)\Big).$$

The previous value of m^{th} may only be estimated since $\|f - f_m\|$ and $\mathrm{Var}(m)$ are both unknown. Using the estimator $\hat{f}_m$ defined by (6), we would like to consider the following preliminary estimate of m:

$$\underset{m\in\mathcal{M}}{\operatorname{argmin}}\Big(-\|\hat{f}_m\|^2 + \widehat{\mathrm{Var}}(m)\Big), \tag{13}$$

where $\widehat{\mathrm{Var}}(m)$ would be an adequate estimate of the variance. These are the ideas leading to the following proposals.

4.2. **Theoretical result.** Let us now give our theoretical result. Let

$$\mathcal{M}_{n,M} = \{k \in \mathbb{N},\ k = 1, \dots, m(n, M)\}$$

be a discrete set of cutoff with $m(n, M)$ such that

$$\frac{\Delta_2(m)}{n+M} \vee \frac{\Delta_2^{(f)}(m)}{\sqrt{M}} \vee \frac{\Delta_4^{(f)}(m)}{M} \leq C,\ \forall m \in \mathcal{M}_{n,M}.$$

It is easy to see that $\hat{f}_m$ can also be defined as the minimizer of

$$\gamma_{n,M}(t) = \|t\|^2 - \frac{1}{\pi}\int \frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} t^*(-u)du$$

over the functions t belonging to S_m, where

$$S_m = \{t \in \mathbb{L}^2(\mathbb{R}), \mathrm{supp}(t^*) = [-\pi m, \pi m]\}.$$

Here, $\mathrm{supp}(t)$ denotes the support of the function t i.e. the domain where it is nonzero. We recall that $(\varphi_{m,j})_{j\in\mathbb{Z}}$ and $\varphi_{m,j} = \sqrt{m}\varphi(mx - j)$, $\varphi(x) = \sin(\pi x)/(\pi x)$, is an orthonormal basis of S_m. In other words, we have

$$\hat{f}_m = \arg\min_{t\in S_m} \gamma_{n,M}(t) \text{ and } \gamma_{n,M}(\hat{f}_m) = -\|\hat{f}_m\|^2.$$

Next, let us define

$$\mathrm{pen}(m) = K_0\left\{\log(M)\log(n+M)\frac{\Delta_2(m)}{n+M} + \left(\log(M)\frac{\Delta_2^{(f)}(m)}{\sqrt{M}} \wedge \log^{3/2}(M)\frac{\Delta_4^{(f)}(m)}{M}\right)\right\}.$$

Then we take

$$\hat{m} = \arg\min_{m\in\mathcal{M}_{n,M}}\Big(\gamma_{n,M}(\hat{f}_m) + \mathrm{pen}(m)\Big). \tag{14}$$

We can prove the following result for the theoretical estimator $\hat{f}_{\hat{m}}$ (the proof is given in supplementary material).

Theorem 1. *Assume that **(A1)-(A3)** hold and $\hat{f}_{\hat{m}}$ is defined by (6) with $\hat{m}$ as in (14). Then there exists a constant C such that for any $m \in \mathcal{M}_{n,M}$,*

$$\mathbb{E}(\|f - \hat{f}_{\hat{m}}\|^2) \leq C\left(\|f - f_m\|^2 + \text{pen}(m)\right) + \frac{C'}{n} + \frac{C"}{M}. \tag{15}$$

where C is a numerical constant and $C', C"$ are constants which do not depend on n, M.

We can see that logarithmic losses occur in all terms of the penalty $\text{pen}(m)$ compared to the terms appearing in the variance. The resulting rate may be deteriorated accordingly, but this deterioration will remain negligible.

To complete the procedure, we suggest then a plug-in method to replace the terms in the penalty and in the admissible set $\mathcal{M}_{n,M}$ by estimators

$$\hat{\Delta}_2(m) = \int_{-\pi m}^{\pi m} \frac{du}{(\tilde{f}_\varepsilon^*)^2(u)} \text{ and } \hat{\Delta}_k^{(f)}(m) = \int_{-\pi m}^{\pi m} \frac{|\hat{f}_Y^*(u)|^2}{(\tilde{f}_\varepsilon^*)^{k+2}(u)} du \text{ since } f^* = \frac{f_Y^*}{f_\varepsilon^*},$$

see Section 5.2.

We do not prove any theoretical result concerning this complete procedure. Nevertheless, we can mention that the result proved in Comte and Lacour [2011] can be applied to a complete procedure based on a slightly different estimator of f. Let us consider an estimator $\check{f}_m$ of f where only the n observations of the non-paired sample are used to estimate f_Y^*. Let us also consider the case where the number M of paired observations is such that $M \geq n^{2+\epsilon}$ for some $\epsilon > 0$. Then the two terms $\check{Q}_1(m)$ and $\check{Q}_2(m)$ involved in the bound of the variance of $\check{f}_m$ are such that $\check{Q}_1(m) = \Delta_2(m)/n$ and $\check{Q}_2(m) \leq \Delta_2^{(f)}(m)/n \leq \Delta_2(m)/n$. We can thus construct a complete procedure for $\check{f}_m$, as follows. We first define an estimate penalty $\check{\text{pen}}(m)$ by:

$$\check{\text{pen}}(m) = K_1 \left(\frac{\log\left(\hat{\Delta}_2(m)\right)}{\log(m+1)}\right)^2 \frac{\hat{\Delta}_2(m)}{n}$$

and $\check{\mathcal{M}}_n = \{1, \dots, \check{m}_n\}$ with $\check{m}_n = \arg\max\{m \in \{1, \dots, n\},\ 1/4 \leq \hat{\Delta}_2(m)/n \leq 1/2\}$. Then the model is selected by

$$\check{m} = \arg \min_{m \in \{1, \dots, \check{m}_n\}} \{-\|\check{f}_m\|^2 + \check{\text{pen}}(m)\}$$

and we get a final estimator $\check{f}_{\check{m}}$. The result proved in Comte and Lacour [2011] can be extended to this complete procedure and we can prove that

$$\mathbb{E}(\|\check{f}_{\check{m}} - f\|^2) \leq C_1 \inf_{1 \leq m \leq m(n,0)} \left\{ \|f_m - f\|^2 + \left(\frac{\log\left(\Delta_2(m)\right)}{\log(m+1)}\right)^2 \frac{\Delta_2(m)}{n} \right\} + \frac{C_2}{n}$$

where C_1 is a numerical constant, C_2 a constant depending on f and f_ε.

This result allows to exhibit a convergent complete procedure. However it hides a substantial loss in the rate of the estimator $\check{f}_m$. Indeed, the rate is related to a sample size of order n, while the number of observations is in fact of order $n+M$ with $M \geq n^{2+\epsilon}$. This is the reason why we consider the estimator $\hat{f}_m$ with an estimator of f_Y^* based on the $n+M$ observations (6), which may avoid the previous loss. Nevertheless, as the complete convergence result is proved for a complete procedure with the penalty $\check{\text{pen}}(m)$, we decide to use the same penalty where n is replaced by $n+M$, as explained below.

5. Simulation study

5.1. **Discussion about the dominating variance term.** We have shown in Section 3 that the optimal rate could be reached both for $M \leq n$ or $M \geq n$, depending on the configuration of the parameters. The dominating term in the bound of $\text{Var}(m)$ is studied numerically for these two configurations.

An estimation of $\text{Var}(m)$ is $\widehat{\text{Var}}(m) = \hat{Q}_1(m) + \hat{Q}_2(m)$ with

$$\hat{Q}_1(m) = \frac{\hat{\Delta}_2(m)}{n+M}, \quad \hat{Q}_2(m) = \frac{\hat{\Delta}_2^{(f)}(m)}{\sqrt{M}} \wedge \frac{\hat{\Delta}_4^{(f)}(m)}{M}.$$

An example of empirical behavior of $\hat{Q}_1(m)$ and $\hat{Q}_2(m)$ is depicted in Figure 1 for $M \ll n$ and $M \gg n$ and in the setting of ordinary smooth or super smooth functions for f chosen as Laplace and Gaussian respectively, together with a super smooth Gaussian noise. The Figure with ordinary smooth Gaussian noise is not reported since it is exactly similar to Figure 1 with only a slight difference in the vertical scales. Expectedly, $\hat{Q}_1(m)$ is larger than the other term when $M \gg n$. Interestingly, this seems to be also true when $M \ll n$, at least empirically: the only case where it is not true is for m less than one, and we can check that in practice, in all the examples considered in Section 5.4, selected values of m are larger than 2. This finding was invariable throughout simulations, thus making $\widehat{\text{Var}}(m) \simeq C\hat{\Delta}_2(m)/(n+M)$ an appropriate choice regardless of the respective values of n and M. The previous considerations become prominent in the choice of a penalty and reinforce the justification of a penalty similar to $\check{\text{pen}}(m)$.

5.2. **Estimation procedure.** In practice we define the estimated penalty $\widehat{\text{pen}}(m)$ as:

$$\widehat{\text{pen}}(m) = K_1 \left(\frac{\log\left(\hat{\Delta}_2(m)\right)}{\log(m+1)} \right)^2 \frac{\hat{\Delta}_2(m)}{n+M}. \tag{16}$$

Throughout numerical estimations we will consider $K_1 = 1$, after a set of simulation experiments to calibrate it. The computation of $\|\hat{f}_m\|$ is performed by using the following expression of the estimator (6) as an orthogonal projection

$$\hat{f}_m = \sum_{\ell \in \mathbb{Z}} \hat{a}_{m,\ell} \varphi_{m,\ell} \tag{17}$$

where $\{\varphi_{m,\ell}\}_{\ell \in \mathbb{Z}}$ is the orthonormal sinus cardinale basis defined in Section 4.2 and $\varphi_{m,\ell}^*(u) = e^{-iu\ell/m} \mathbf{1}_{[-\pi m, \pi m]}(u)/\sqrt{m}$. We also recall that the estimated projection coefficients can be computed by the following formula

$$\hat{a}_{m,\ell} = \frac{\sqrt{m}}{2}(-1)^\ell \int_0^2 e^{i\ell\pi u} \frac{\hat{f}_Y^*}{\tilde{f}_\varepsilon^*}(\pi m(u-1))du.$$

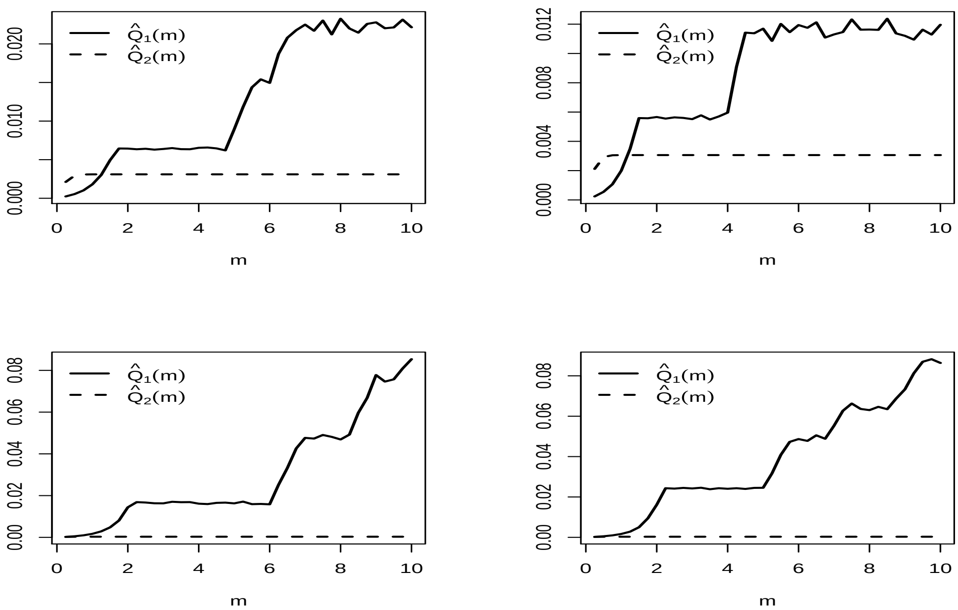

Figure 1. Empirical behavior of $\hat{Q}_1(m)$ and $\hat{Q}_2(m)$ as a function of m when $M \ll n$ and $M \gg n$ for f ordinary smooth and super smooth, chosen as Laplace (left figures) and Gaussian (right figures) respectively. The influence of M and n is illustrated by $M = 200$, $n = 2000$ (top figures) and $M = 2000$, $n = 200$ (bottom figures). The noise is Gaussian.

This expression of the estimator allows us to use Inverse Fast Fourier Transform (IFFT) Algorithms in the estimation process. Therefore, for numerical tractability we use only a finite sample of projection coefficients with $\hat{f}_m = \sum_{|\ell| \leq K_n} \hat{a}_{m,\ell}\varphi_{m,\ell}$. Theoretical results assert that $K_n = n$ always suits (see Comte et al. [2006]) but we make the constant choice $K_n = 255$.

Since $\mathcal{M}_{n,M}$ is unknown we consider an estimation of this domain,

$$\widehat{\mathcal{M}}_{n,M} = \{k/\kappa,\ k = 1, \ldots, \kappa\, \hat{m}_{n,M}\}$$

where κ is a fixed positive integer and $\hat{m}_{n,M}$ is such that

$$\hat{m}_{n,M} = \operatorname{argmax}\Big(m \in \mathbb{N}, \frac{\hat{\Delta}_2(m)}{n+M} \leq 2\Big)$$

Following, we have the final estimation of m^{th} defined by:

$$\widehat{\hat{m}} = \underset{m\in\{1,\dots,\hat{m}_{n,M}\}}{\text{argmin}} \Big(- \|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)\Big) \tag{18}$$

with $\widehat{\text{pen}}(m)$ given by (16). Finally, by considering (17) and (18), we obtain $\hat{f}_{\widehat{\hat{m}}}$ which is our final estimator. The choice of the constant κ will influence the quality of the final estimation since it governs the number of models that are proposed before selection. Choosing κ small will offer only a restricted number of models for the algorithm to choose from, whereas choosing κ large will allow a more refined estimation of $\hat{m}_n$. For the simulations in Section 5, we choose $\kappa = 4$ to keep the computing time reasonable. Conversely, for the real data application in Section 6, we choose $\kappa = 30$.

5.3. **Design of simulation.** Noise was given a Laplace or a Gaussian density with variance σ^2 as follows:

- Laplace noise.

$$f_\varepsilon(x) = \frac{\sigma}{2}e^{-\sigma|x|} \text{ and } f_\varepsilon^*(u) = \frac{\sigma^2}{\sigma^2 + u^2}$$

- Gaussian noise.

$$f_\varepsilon(x) = \frac{1}{\sqrt{2\pi}}e^{-0.5x^2/\sigma^2} \text{ and } f_\varepsilon^*(u) = e^{-0.5\sigma^2u^2}$$

We compared our results to estimations under the assumption of a known noise density for the description of the estimation procedure and penalties for Gaussian and Laplace noises). We considered the following four different densities of the X_j's:

(i) Mixed Gamma distribution: $X = 1/\sqrt{5.48}W$ with $W \sim 0.4\Gamma(5, 1) + 0.6\Gamma(13, 1)$
(ii) Cauchy distribution: $f(x) = (1/\pi)/(1 + x^2)$
(iii) Laplace distribution: $f(x) = e^{-\sqrt{2}|x|}/\sqrt{2}$
(iv) Gaussian distribution: $X \sim \mathcal{N}(0, 1)$
(v) Beta distribution: $X \sim 2\sqrt{7}B(3, 3)$

Except the case of the Cauchy density, these densities are normalized with unit variance, thus allowing the ratio $1/\sigma^2$ to represent the signal-to-noise ratio, denoted $s2n$. We considered signal to noise ratios of $s2n = 5$ and $s2n = 10$ in our simulations. To study the influence of the relationship between n and M, we considered several values of n and values of $M = n$ and $M = \sqrt{n}$. Additionally, we considered the density (vi) defined by $X \sim 0.5\mathcal{N}(-3, 1)+0.5\mathcal{N}(2, 1)$ with a signal-to-noise ratio of 4 for comparison with Delaigle et al. [2008].

5.4. **Results.** The values of the MISE risk multiplied by 100 for each density and simulation scenario and computed from 100 simulated data sets, are given in Table 1. As expected, the risk decreases as n or M increases. Similarly, when increasing the level of contamination of the data by reducing the signal-to-noise ratio, the risk increases. Compared to Gaussian noise, Laplace noise demonstrates overall lower risks whatever the other simulation parameters. Indeed, Gaussian noise is super smooth whereas Laplace noise is ordinary smooth thus explaining the improvement in risk. Strikingly, in most cases the estimation of the square of the characteristic function of noise density f_ε^* has reduced the risk compared to the known density case. This phenomenon is counter-intuitive and we do not have a clear explanation. However, this has been noticed in Comte and Lacour [2011]. Following Remark 1, we have also implemented the estimator of f_Y^* is based on the samples (Y_j), $(Y_{n+j,1})$ and $(Y_{n+j,2})$. The MISE are of the same order than those presented in Table 1.

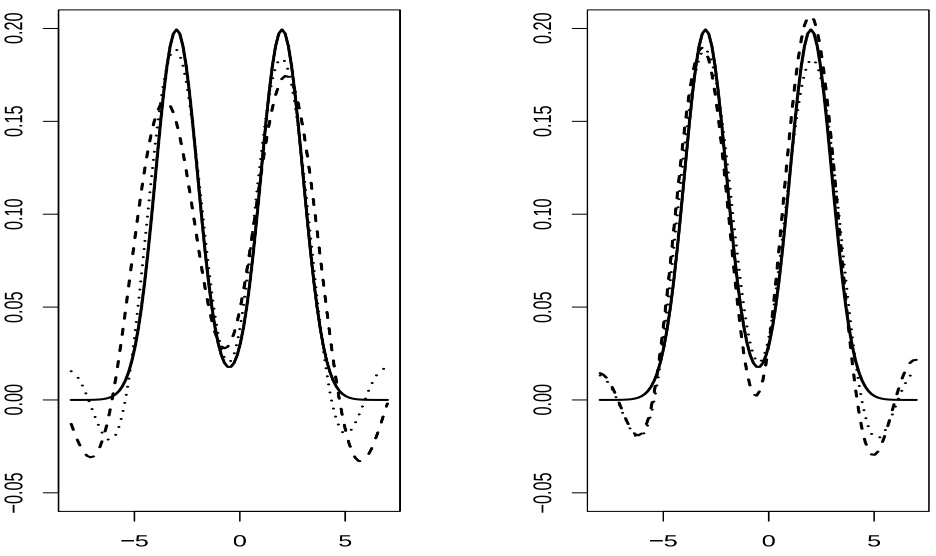

Figure 2. Estimations for $n = M = 200$ (dashed line) and $n = M = 500$ (dotted line) for the Gaussian mixture density (plain line) (vi). (left): Laplace noise; (right): Gaussian noise. Two independent samples were used, of size n and M respectively.

Table 2 presents the comparison of the penalized estimator and the estimator given by Delaigle et al. [2008] for the Gaussian mixture density (vi). The integrated squared error (ISE) is computed over 100 estimations and we present the results using the median and inter-quartile range (IQR). In all cases, the penalized estimator enjoys lower risks compared to those given by Delaigle et al. [2008].

In Figure 2, we present an estimation of f using the penalized estimator. We considered the Gaussian mixture distribution (vi) contaminated by Gaussian and Laplace noise with a signal-to-noise ratio of 4 with $n = M = 200$ and $n = M = 500$. The bimodal character of distribution (v) is well described by the estimation in both cases whereas the increase in precision for $n = M = 500$ is mostly visible in the Laplace noise case which closely matches the theoretical density in that case.

TABLE 1. Results of simulations presented as MISE×100. In each case the MISE was averaged over 100 estimations. The case "f_ε known" corresponds to the estimator given in (3) with Y-sample of size $2n$ (to be compared to the case $M = n$).

ε **Gaussian**		$s2n = 10$		$s2n = 5$	
		$n = 200$	$n = 2000$	$n = 200$	$n = 2000$
f Mixed Gamma	f_ε known	0.447	0.104	0.700	0.668
	$M = \sqrt{n}$	0.631	0.092	0.788	0.277
	$M = n$	0.374	0.059	0.508	0.149
f Cauchy	f_ε known	0.371	0.044	0.728	0.624
	$M = \sqrt{n}$	0.513	0.107	0.808	0.274
	$M = n$	0.511	0.075	0.445	0.123
f Laplace	f_ε known	2.066	0.588	3.506	1.815
	$M = \sqrt{n}$	1.469	0.592	3.908	2.504
	$M = n$	1.088	0.405	2.391	1.209
f Gaussian	f_ε known	0.191	0.041	0.355	0.846
	$M = \sqrt{n}$	0.848	0.120	0.847	0.175
	$M = n$	0.681	0.093	0.476	0.133
f Beta	$M = \sqrt{n}$	0.236	0.049	0.517	0.097
	$M = n$	0.129	0.037	0.242	0.043
ε **Laplace**		$s2n = 10$		$s2n = 5$	
		$n = 200$	$n = 2000$	$n = 200$	$n = 2000$
f Mixed Gamma	f_ε known	0.349	0.062	0.588	0.107
	$M = \sqrt{n}$	0.570	0.095	0.771	0.198
	$M = n$	0.395	0.062	0.535	0.108
f Cauchy	f_ε known	0.339	0.167	0.420	0.149
	$M = \sqrt{n}$	0.612	0.107	0.685	0.195
	$M = n$	0.550	0.053	0.498	0.174
f Laplace	f_ε known	1.110	0.367	1.967	0.664
	$M = \sqrt{n}$	1.376	0.479	3.359	1.595
	$M = n$	0.936	0.385	1.726	0.578
f Gaussian	f_ε known	0.511	0.219	0.720	0.386
	$M = \sqrt{n}$	0.839	0.122	0.859	0.182
	$M = n$	0.594	0.066	0.720	0.222
f Beta	$M = \sqrt{n}$	0.256	0.048	0.466	0.065
	$M = n$	0.126	0.036	0.224	0.039

6. Density estimation of onset of pregnancy

As defined previously, X denotes the interval between last menstrual period (LMP) and the true onset of pregnancy. We denote Y the interval between LMP and the onset of pregnancy estimated by the sonographic measurement of the crown-rump length (CRL)

Table 2. Comparison of the ISE between the estimators of Delaigle et al. [2008] and the penalized estimator for the Gaussian mixture density (vi). For the sake of comparison, the results are presented by the median $\times$ 100 (inter-quartile range $\times$ 100) of 100 estimations with $M = n$.

	Delaigle et al. [2008]		Penalized estimator	
	$n = 200$	$n = 500$	$n = 200$	$n = 500$
ε Laplace	1.41 (0.94)	0.89 (0.51)	0.35 (0.17)	0.24 (0.07)
ε Gaussian	2.09 (1.33)	1.42 (0.92)	0.47 (0.90)	0.27 (0.08)

with $Y = X + \varepsilon$. Two separate independent samples are available: the first is an M-sample of spontaneous twin pregnancies, $M = 86$, each embryo with its own CRL measurement; the second is an n-sample of spontaneous singleton pregnancies, $n = 1378$, with $Y_j = X_j + \varepsilon_j$. Each of these samples is a sample of the general unselected population and was obtained from the screening unit of the department of obstetrics and maternal-fetal medicine of the children's hospital Necker - Enfants Malades in Paris, France. Since the onset of pregnancy is identical for both twins, we thus have replicate noisy observations $Y_{n+j,1} = X_j + \varepsilon_{n+j,1}$ and $Y_{n+j,2} = X_{-j} + \varepsilon_{n+j,2}$, $-j = 1, \ldots, M$. We wish to estimate f which represents the distribution of probability of onset of pregnancy within a female cycle.

Figure 3 (b) presents the penalized estimator $\hat{f}_{\widehat{m}}(.)$. As expected, the mode of the distribution is at around 13 days, meaning that the likelihood of onset of pregnancy is greatest at 13 days following the last menstrual period. However this distribution looks positively skewed with a significant remaining probability of onset after 20 days. The risk was assessed by simulation in the setting of our data by considering $X \sim \Gamma(16, 1.2)$ and a Laplace $\varepsilon \sim \text{Lap}(0, 0.95)$ or Gaussian $\varepsilon \sim \mathcal{N}(0, 1.2)$ noise. These densities were chosen empirically because they fitted our estimate (see Figure 3 (a) for the comparison of the empirical characteristic function of ε with Laplace and Gaussian characteristic functions). Under this simulation model, the risk MISE$\times$100 was 0.038 and 0.034 for Laplace and Gaussian noise respectively over 100 estimations. We emphasize that the strong side-effects which are observed on the estimated characteristic function in Figure 3 (a) can also be seen on simulated data (for a size sample 86) and mainly appear when going from the direct noise observation to the replicate case (where only differences of noise are observed).

7. Concluding remarks

We have presented an adaptive deconvolution estimator of a density when the noise density is unknown. Instead, a sample of noisy replicate observations is available. Although this estimator seems to perform nicely in simulation, it can exhibit poorer theoretical rates than in other settings. This expected loss is directly related to the use of replicate observations for the estimation of the characteristic function of noise density or more precisely the square of its modulus. Simulations show that the influence of the relative values of M and n is likely to be small. We also find that the gain in precision for increasing values

(a) (b)

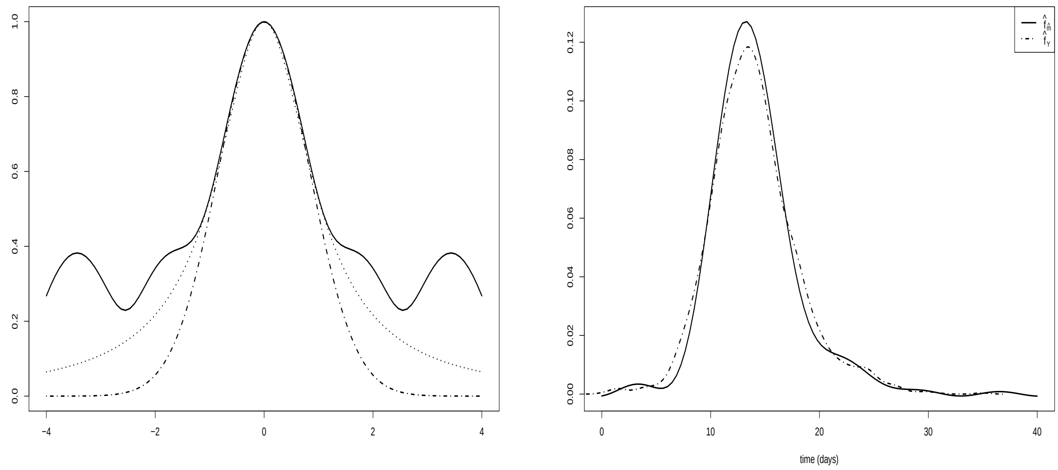

Figure 3. (a): estimation of the characteristic function of the noise in twin pregnancies (plain line). Laplace (dashed line) and Gaussian (dotted line) characteristic functions are plotted for comparison. (b): estimation of the density of onset of pregnancy (plain line) and estimated density of the observations (dashed line).

of M may be small. Whereas this may be of little value in the field of engineering, it is of importance in biomedical applications or clinical research. Indeed, obtaining a sample of ε is often difficult or impossible in these applications, as well as a strong prior assumption regarding its density. However, replicate data may be found in clinical or biomedical applications. Nevertheless they are likely to be scarce since they involve multiple measurements/observations in one patient. In the case of dating pregnancy this is dealt with by using twin pregnancies instead. The estimation of a density of onset of pregnancy may find multiple clinical applications. The knowledge of the underlying variability of onset of pregnancy may help clinicians in the follow-up of pregnancies and mostly regarding growth monitoring by ultrasound and delivery since both these aspects rely upon an accurate estimation of onset of pregnancy. Furthermore, this density is of interest for the physiology of the female cycle, confirming with simple clinical data the variation in onset of pregnancy that could be expected from biological experiments (Wilcox et al. [2000]).

Acknowledgement

The authors thank Pr Jean-Christophe Thalabard for his advice, suggestions and support.

8. Appendix: Proofs

8.1. **Proof of Remark 1.** Let us denote the estimator of f_Y^* based on the samples (Y_j), $(Y_{n+j,1})$ and $(Y_{n+j,2})$:

$$\tilde{f}_Y^*(u) = \frac{1}{n+2M}\left(\sum_{j=1}^n e^{iuY_j} + \sum_{j=1}^M e^{iuY_{n+j,1}} + \sum_{j=1}^M e^{iuY_{n+j,2}}\right).$$

Due to the dependency between $(Y_{n+j,1})$ and $(Y_{n+j,2})$, we have

$$\begin{aligned} Var\, \tilde{f}_Y^*(u) &= \frac{1}{(n+2M)^2}\left(Var\left(\sum_{j=1}^n e^{iuY_j}\right) + Var\left(\sum_{j=1}^M e^{iuY_{n+j,1}}\right) + Var\left(\sum_{j=1}^M e^{iuY_{n+j,2}}\right)\right. \\ &\qquad \left. + 2\mathrm{cov}\left(\sum_{j=1}^M e^{iuY_{n+j,1}}, \sum_{j=1}^M e^{iuY_{n+j,2}}\right)\right) \\ &= \frac{1}{n+2M}(1-(f_\varepsilon^*)^2(u)|f^*(u)|^2) + \frac{2M}{(n+2M)^2}(f_\varepsilon^*)^2(u)(1-|f^*(u)|^2) \end{aligned}$$

The variance of $\hat{f}_Y^*(u)$ is equal to $\frac{1}{n+M}(1-(f_\varepsilon^*)^2(u)|f^*(u)|^2)$. The main term of the variance of $\tilde{f}_Y^*$ is $1/(n+2M)$ while it is $1/(n+M)$ for $\hat{f}_Y^*(u)$, which explains Remark 1.

8.2. **Proof of Lemma 1.** Let us define

$$R(u) = \frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} \text{ and } S(u) = R(u) + \frac{1}{f_\varepsilon^*(u)}\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)<M^{-1/2}}. \tag{19}$$

We first prove the following result, which is useful for the proof of Theorem 1.

Lemma 2. *Consider $R(u)$ and $S(u)$ as defined by (19). Then we have:*

$$(a)\ \ |S(u)|^2 \le \frac{M}{(f_\varepsilon^*)^2(u)}((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u))^2 \quad \text{and} \quad \mathbb{E}[|S(u)|^2] \le \frac{1}{(f_\varepsilon^*)^2(u)}.$$

$$(b)\ \ |S(u)|^2 \le \frac{M^{1/2}}{(f_\varepsilon^*)^4(u)}\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^2 \quad \text{and} \quad \mathbb{E}\left[|S(u)|^2\right] \le \frac{M^{-1/2}}{(f_\varepsilon^*)^4(u)}.$$

$$\begin{aligned}(c)\ \ |S(u)|^2 &\le \frac{\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^2}{(f_\varepsilon^*)^6(u)} + \frac{M^{1/2}}{(f_\varepsilon^*)^6(u)}\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^3 \\ &\quad \text{and } \mathbb{E}[|S(u)|^2] \le 2\frac{M^{-1}}{(f_\varepsilon^*)^6(u)}.\end{aligned}$$

Proof of Lemma 2.

First remark that $S(u) = \mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)\ge M^{-1/2}}\left(\frac{1}{\sqrt{\widehat{(f_\varepsilon^*)^2}(u)}} - \frac{1}{f_\varepsilon^*(u)}\right)$.

Proof of (a). The following equality holds

$$S(u) = \mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) \geq M^{-1/2}} \frac{(f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)}{(f_\varepsilon^*(u) + \sqrt{\widehat{(f_\varepsilon^*)^2}(u)})f_\varepsilon^*(u)\sqrt{\widehat{(f_\varepsilon^*)^2}(u)}}.$$

Thus using $f_\varepsilon^*(u) + \sqrt{\widehat{(f_\varepsilon^*)^2}(u)} \geq \sqrt{\widehat{(f_\varepsilon^*)^2}(u)}$ and the bound given by the indicator, we get $|S(u)|^2 \leq [M/(f_\varepsilon^*)^2(u)]((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u))^2$, and the expectation follows.
Proof of (b). We can also write

$$\begin{aligned}
|S(u)|^2 &= \frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) \geq M^{-1/2}}}{(f_\varepsilon^*)^2(u)} \frac{\left(f_\varepsilon^*(u) - \sqrt{\widehat{(f_\varepsilon^*)^2}(u)}\right)^2}{\widehat{f_\varepsilon^{*2}}(u)} \\
&= \frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) \geq M^{-1/2}}}{(f_\varepsilon^*)^2(u)} \frac{\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^2}{\widehat{f_\varepsilon^{*2}}(u)\left(f_\varepsilon^*(u) + \sqrt{\widehat{(f_\varepsilon^*)^2}(u)}\right)^2} \leq \frac{M^{1/2}}{(f_\varepsilon^*)^4(u)}\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^2.
\end{aligned}$$

Therefore $\mathbb{E}\left[|S(u)|^2\right] \leq M^{-1/2}/(f_\varepsilon^*)^4(u)$, which completes the proof of (b).

Proof of (c). We can also write

$$\begin{aligned}
|S(u)|^2 &= \frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) \geq M^{-1/2}}}{(f_\varepsilon^*)^2(u)} \frac{\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^2}{\widehat{(f_\varepsilon^*)^2}(u)\left(f_\varepsilon^*(u) + \sqrt{\widehat{(f_\varepsilon^*)^2}(u)}\right)^2} \\
&= \frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) \geq M^{-1/2}}}{(f_\varepsilon^*)^2(u)} \frac{\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^2}{\left(f_\varepsilon^*(u) + \sqrt{\widehat{(f_\varepsilon^*)^2}(u)}\right)^2}\left[\frac{1}{\widehat{(f_\varepsilon^*)^2}(u)} - \frac{1}{(f_\varepsilon^*)^2(u)} + \frac{1}{(f_\varepsilon^*)^2(u)}\right].
\end{aligned}$$

Thus

$$\begin{aligned}
|S(u)|^2 &= \frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) \geq M^{-1/2}}}{(f_\varepsilon^*)^4(u)} \frac{\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^2}{\left(f_\varepsilon^*(u) + \sqrt{\widehat{(f_\varepsilon^*)^2}(u)}\right)^2} + \frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) \geq M^{-1/2}}\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^3}{(f_\varepsilon^*)^4(u)\widehat{(f_\varepsilon^*)^2}(u)\left(f_\varepsilon^*(u) + \sqrt{\widehat{(f_\varepsilon^*)^2}(u)}\right)^2} \\
&\leq \frac{\left((f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right)^2}{(f_\varepsilon^*)^6(u)} + \frac{M^{1/2}\left|(f_\varepsilon^*)^2(u) - \widehat{(f_\varepsilon^*)^2}(u)\right|^3}{(f_\varepsilon^*)^6(u)}.
\end{aligned}$$

This implies $\mathbb{E}\left[|S(u)|^2\right] \leq 2M^{-1}/|f_\varepsilon^*(u)|^6$, which completes the proof of (c). □.

Proof of Lemma 1. We give the proof for $p=2$ but the extension to any p is straightforward. First, write the decomposition

$$\mathbb{E}(|R(u)|^2) \;=\; \mathbb{E}\left(\frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)<M^{-1/2}}}{(f_\varepsilon^*)^2(u)}\right)+\mathbb{E}(|S(u)|^2).$$

Clearly, the bound for $\mathbb{E}[(S(u))^2]$ obtained in Lemma 2 is the bound announced for $\mathbb{E}(R^2(u))$. Therefore, we only have to prove that the first term of the decomposition has the same bound. Clearly,

$$\mathbb{E}\left(\frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)<M^{-1/2}}}{(f_\varepsilon^*)^2(u)}\right)\le\frac{1}{(f_\varepsilon^*)^2(u)}.$$

For the other terms to obtain as bounds, let us distinguish two cases. (i) If $(f_\varepsilon^*)^2(u)\le 2M^{-1/2}$, we have both $1/(f_\varepsilon^*)^2(u)\le 2M^{-1/2}/(f_\varepsilon^*)^4(u)$ and $M^{-1/2}/(f_\varepsilon^*)^4(u)\le 2M^{-1}/(f_\varepsilon^*)^6(u)$. Therefore, in this case

$$\frac{1}{(f_\varepsilon^*)^2(u)}=\frac{1}{(f_\varepsilon^*)^2(u)}\wedge\frac{2M^{-1/2}}{(f_\varepsilon^*)^4(u)}\wedge\frac{2M^{-1}}{(f_\varepsilon^*)^6(u)}.$$

(ii) If $(f_\varepsilon^*)^2(u)>2M^{-1/2}$, using the Bernstein Inequality as in Neumann yields:

$$\begin{aligned}\mathbb{P}(|(\hat{f_\varepsilon^*})^2(u)|<M^{-1/2}) &\le \mathbb{P}(|\widehat{(f_\varepsilon^{*2}}(u)-(f_\varepsilon^*)^2(u)|>(f_\varepsilon^*)^2(u)-M^{-1/2})\\ &\le \mathbb{P}(|\widehat{(f_\varepsilon^*)^2}(u)-(f_\varepsilon^*)^2(u)|>(f_\varepsilon^*)^2(u)/2)\\ &\le 2\exp(-M(f_\varepsilon^*)^4(u)/16)\le O\left((M^{-1}(f_\varepsilon^*(u))^{-4})^p\right).\end{aligned}$$

Consequently,

$$\mathbb{E}\left(\frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)<M^{-1/2}}}{(f_\varepsilon^*)^2(u)}\right)\le\frac{1}{(f_\varepsilon^*)^2(u)}\mathbb{P}(\widehat{(f_\varepsilon^*)^2}(u)<M^{-1/2})\le c\frac{M^{-1}}{(f_\varepsilon^*)^6(u)}$$

Thus, in that case where $1/(f_\varepsilon^*)^2(u)\ge 2M^{-1/2}/(f_\varepsilon^*)^4(u)\ge 4M^{-1}/(f_\varepsilon^*)^6(u)$, we get

$$\mathbb{E}\left(\frac{\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)<M^{-1/2}}}{(f_\varepsilon^*)^2(u)}\right)\le C\frac{M^{-1}}{(f_\varepsilon^*)^6(u)}=\frac{1}{(f_\varepsilon^*)^2(u)}\wedge\frac{2M^{-1/2}}{(f_\varepsilon^*)^4(u)}\wedge\frac{4M^{-1}}{(f_\varepsilon^*)^6(u)}.$$

This ends the proof of the lemma. □

8.3. **Proof of Proposition 1.** Let us study the integrated mean square risk. By writing in the Fourier domain that

$$f^*-\hat f_m^*=(f^*-f_m^*)+(f_m^*-\hat f_m^*)=f^*\mathbf{1}_{[-\pi m,\pi m]^c}+(f_m^*-\hat f_m^*)\mathbf{1}_{[-\pi m,\pi m]},$$

we get, as $\|f-\hat f_m\|^2=(2\pi)^{-1}\|f^*-\hat f_m^*\|^2=(2\pi)^{-1}(\|f^*\mathbf{1}_{[-\pi m,\pi m]^c}\|^2+\|(f_m^*-\hat f_m^*)\mathbf{1}_{[-\pi m,\pi m]}\|^2)$, that

$$\|f-\hat f_m\|^2=\|f-f_m\|^2+\|f_m-\hat f_m\|^2. \tag{20}$$

Moreover, by applying the Parseval formula, we obtain

$$\|f_m - \hat{f}_m\|^2 = \frac{1}{2\pi}\int_{-\pi m}^{\pi m}\left|\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)}\right|^2 du.$$

It follows that

$$\|f_m - \hat{f}_m\|^2 \leq \frac{1}{\pi}\int_{-\pi m}^{\pi m}|\hat{f}_Y^*(u)|^2|R(u)|^2 du + \frac{1}{\pi}\int_{-\pi m}^{\pi m}\frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{(f_\varepsilon^*)^2(u)}du. \tag{21}$$

The last term of the right-hand-side of (21) is the usual term that is found when f_ε^* is known, and the first one is specific to the framework with estimated f_ε^*.

We take the expectation of (21):

$$\begin{aligned}\mathbb{E}(\|f_m - \hat{f}_m\|^2) &\leq \frac{2}{\pi}\int_{-\pi m}^{\pi m}\mathbb{E}(|\hat{f}_Y^*(u) - f_Y^*(u)|^2|R(u)|^2)du \\ &\quad + \frac{2}{\pi}\int_{-\pi m}^{\pi m}|f_Y^*(u)|^2\mathbb{E}(|R(u)|^2)du + \frac{1}{\pi}\int_{-\pi m}^{\pi m}\frac{(n+M)^{-1}}{(f_\varepsilon^*)^2(u)}du.\end{aligned}$$

Applying Lemma 1 yields:

$$\begin{aligned}\mathbb{E}(\|f_m - \hat{f}_m\|^2) &\leq \frac{2}{\pi}\int_{-\pi m}^{\pi m}(\mathbb{E}(|\hat{f}_Y^*(u) - f_Y^*(u)|^4)\mathbb{E}(|R(u)|^4))^{1/2}du \\ &\quad + \frac{2}{\pi}\int_{-\pi m}^{\pi m}|f^*(u)|^2(f_\varepsilon^*)^2(u)\mathbb{E}(|R(u)|^2)du + 2\frac{\Delta_2(m)}{n+M} \\ &\leq \frac{2}{\pi}\int_{-\pi m}^{\pi m}(n+M)^{-1}\frac{1}{(f_\varepsilon^*)^2(u)}du \\ &\quad + \frac{2C_1}{\pi}\int_{-\pi m}^{\pi m}|f^*(u)|^2(f_\varepsilon^*)^2(u)\left(\frac{M^{-1/2}}{(f_\varepsilon^*)^4(u)} \wedge \frac{M^{-1}}{(f_\varepsilon^*)^6(u)}\right)du + 2\frac{\Delta_2(m)}{n+M} \qquad (22)\end{aligned}$$

By gathering (20) and (22), we obtain the result. □

References

C Butucea. Deconvolution of supersmooth densities with smooth noise. *Canadian Journal of Statistics*, 32(2):181–192, 2004.

C Butucea and A B Tsybakov. Sharp optimality in density deconvolution with dominating bias. i. *Theory of Probability and its Applications*, 52(1):24, 2008a.

C Butucea and A B Tsybakov. Sharp optimality in density deconvolution with dominating bias. II. *Theory of Probability and its Applications*, 52(2):237, 2008b.

F. Comte and C. Lacour. Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):601–627, 2011.

F Comte, Y Rozenholc, and M L Taupin. Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, 34(3):431–452, 2006.

A Delaigle and I Gijbels. Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, 56(1):19–47, 2004.

A Delaigle, P Hall, and A Meister. On deconvolution with repeated measurements. *The Annals of Statistics*, 36(2):665–685, 2008.

P J Diggle and P Hall. A fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(2): 523–531, 1993.

J Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272, 1991.

J Fan and J Y Koo. Wavelet deconvolution. *Information Theory, IEEE Transactions on*, 48(3):734–747, 2002.

C H Hesse. Data-driven deconvolution. *Journal of Nonparametric Statistics*, 10(4):343–373, 1999.

J Johannes. Deconvolution with unknown error distribution. *The Annals of Statistics*, 37 (5):2301–2323, 2009.

T Li and Q Vuong. Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65:139–165, 1998.

M C Liu and R L Taylor. A consistent nonparametric density estimator for the deconvolution problem. *Canadian Journal of Statistics*, 17(4):427–438, 1989.

A. Meister and M. Neumann. Deconvolution from non-standard error densities under replicated measurements. *Statistica Sinica*, 20 (4):1609–1936, 2009.

M H Neumann. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330, 1997.

M Pensky and B Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, pages 2033–2053, 1999.

H P Robinson. Sonar measurement of fetal crown-rump length as means of assessing maturity in first trimester of pregnancy. *British Medical Journal*, 4(5883):28–31, 1973.

P Sladkevicius, S Saltvedt, H Almström, M Kublickas, C Grunewald, and L Valentin. Ultrasound dating at 12-14 weeks of gestation. a prospective cross-validation of established dating formulae in in-vitro fertilized pregnancies. *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 26(5):504–511, 2005.

L A Stefanski and R J Carroll. Deconvoluting kernel density estimators. *Statistics*, pages 169—184, 1990.

A J Wilcox, D Dunson, and D D Baird. The timing of the "fertile window" in the menstrual cycle: day specific estimates from a prospective study. *BMJ (Clinical Research Ed.)*, 321(7271):1259–1262, 2000.

Supplementary Materials: Proof of Theorem 1

Let us begin the proof as in Comte and Lacour (2011). We observe that for all t, t' in spaces S_m's

$$\gamma_{n,M}(t) - \gamma_{n,M}(t') = \|t - f\|^2 - \|t' - f\|^2 - 2\nu_{n,M}(t - t')$$

where

$$\nu_{n,M}(t) = (n+M)^{-1} \sum_j \left\{ \tilde{v}_t(Y_j) - \int t(x)f(x)dx \right\}, \quad \tilde{v}_t(x) = \frac{1}{2\pi} \int e^{iux} \frac{t^*(-u)}{\tilde{f}_\varepsilon^*(u)} du,$$

with convention $Y_{n+k} = Y_{-k,1}$ for $k = 1, \dots, M$.

Let us fix $m \in \mathcal{M}_{n,M}$ and recall that f_m is the orthogonal projection of f on S_m. Since $\gamma_{n,M}(\hat{f}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma_{n,M}(f_m) + \text{pen}(m)$, we have

$$\begin{aligned} \|\hat{f}_{\hat{m}} - f\|^2 &\leq \|f_m - f\|^2 + 2\nu_{n,M}(\hat{f}_{\hat{m}} - f_m) + \text{pen}(m) - \text{pen}(\hat{m}) \\ &\leq \|f_m - f\|^2 + 2\|\hat{f}_{\hat{m}} - f_m\| \sup_{t \in B(m,\hat{m})} \nu_{n,M}(t) + \text{pen}(m) - \text{pen}(\hat{m}) \end{aligned}$$

where, for all m, m', $B(m, m') = \{t \in S_m + S_{m'}, \quad \|t\| = 1\}$. Then, using inequality $2xy \leq x^2/4 + 4y^2$,

$$\|\hat{f}_{\hat{m}} - f\|^2 \leq \|f_m - f\|^2 + \frac{1}{4}\|\hat{f}_{\hat{m}} - f_m\|^2 + 4 \sup_{t \in B(m,\hat{m})} \nu_{n,M}^2(t) + \text{pen}(m) - \text{pen}(\hat{m}). \tag{23}$$

But $\|\hat{f}_{\hat{m}} - f_m\|^2 \leq 2\|\hat{f}_{\hat{m}} - f\|^2 + 2\|f - f_m\|^2$ so that, introducing a function $p(.,.)$

$$\|\hat{f}_{\hat{m}} - f\|^2 \leq 3\|f_m - f\|^2 + 8[\sup_{t \in B(m,\hat{m})} \nu_{n,M}^2(t) - p(m,\hat{m})] + 8p(m,\hat{m}) + 2\text{pen}(m) - 2\text{pen}(\hat{m}).$$

If p is such that for all m, m',

$$4p(m, m') \leq \text{pen}(m) + \text{pen}(m') \tag{24}$$

then

$$\mathbb{E}\|\hat{f}_{\hat{m}} - f\|^2 \leq 3\|f_m - f\|^2 + 8\mathbb{E}[\sup_{t \in B(m,\hat{m})} \nu_{n,M}^2(t) - p(m,\hat{m})] + 4\text{pen}(m). \tag{25}$$

Our study is now dedicated to find $p(m, m')$ such that $\mathbb{E}[\sup_{t \in B(m,\hat{m})} \nu_{n,M}^2(t) - p(m,\hat{m})]$ is small.

Recall that $R(u)$ and $S(u)$ are defined in (19). We write the following decomposition

$$\begin{aligned}
2\pi\nu_{n,M}(t) &= \int t^*(-u)\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)}du - \langle t^*, f^*\rangle \\
&= \int t^*(-u)\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)}du - \langle t^*, f^*\rangle + \int t^*(-u)\hat{f}_Y^*(u)R(u)du \\
&= \int t^*(-u)\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)}du - \langle t^*, f^*\rangle \\
&\quad + \int t^*(-u)(\hat{f}_Y^*(u) - f_Y^*(u))R(u)du + \int t^*(-u)f^*(u)f_\varepsilon^*(u)R(u)du \\
&= \int t^*(-u)\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)}du - \langle t^*, f^*\rangle + \int t^*(-u)(\hat{f}_Y^*(u) - f_Y^*(u))S(u)du \\
&\quad + \int t^*(-u)f^*(u)f_\varepsilon^*(u)S(u)du - \int t^*(-u)(\hat{f}_Y^*(u) - f_Y^*(u))\frac{1}{f_\varepsilon^*(u)}\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)<M^{-1/2}}du \\
&\quad - \int t^*(-u)f^*(u)\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)<M^{-1/2}}du \\
&:= \nu_{n,M,1}(t) + \nu_{n,M,2}(t) + \nu_{n,M,3}(t) + \nu_{n,M,4}(t) + \nu_{n,M,5}(t)
\end{aligned}$$

The term $\nu_{n,M,1}(t)$ is the empirical process corresponding to the case of known f_ε^* and has already been studied in several works (see Comte *et al.* (2006)). It satisfies

$$\mathbb{E}\left[\sup_{t\in B(m,\hat{m})} \nu_{n,M,1}^2(t) - p_1(m,\hat{m})\right] \le \frac{c}{n+M}$$

with $p_1(m,m') = p_1(m) + p_1(m')$ and

$$p_1(m) = \kappa_1\left(\frac{\log(\Delta_2(m))}{\log(m+1)}\right)^2\frac{\Delta_2(m)}{n+M}.$$

This gives a first contribution to the penalty $\text{pen}(m)$.

We can prove the following Lemma:

Lemma 3. *Under the assumptions of Theorem 1, we have:*

$$(i)\ \mathbb{E}[\sup_{t\in B(m,\hat{m})} \nu_{n,M,2}^2(t) - p_2(m,\hat{m})]_+ \le C\left(\frac{1}{n+M} + \frac{1}{M}\right) \tag{26}$$

with $p_2(m,m') = p_2(m) + p_2(m')$ *and*

$$p_2(m) = \kappa_2\log(M)\log(n+M)\frac{\Delta_2(m)}{n+M}.$$

$$(ii)\ \mathbb{E}[\sup_{t\in B(m,\hat{m})} \nu_{n,M,3}^2(t) - p_3(m,\hat{m})]_+ \le \frac{c}{M}.$$

with $p_3(m,m') = p_3(m) + p_3(m')$ *and*

$$p_3(m) = \kappa_3 \left(\log(M)\frac{\Delta_2^{(f)}(m)}{\sqrt{M}} \wedge \log^{3/2}(M)\frac{\Delta_4^{(f)}(m)}{M} \right).$$

$$(iii)\ \mathbb{E}[\sup_{t\in B(m,\hat m)} \nu_{n,M,4}^2(t) - p_4(m,\hat m)]_+ \le \frac{c}{n+M}$$

with $p_4(m,m') = p_4(m) + p_4(m')$ *and*

$$p_4(m) = \kappa_4 \left(\frac{\log(\Delta_2(m))}{\log(m+1)} \right)^2 \frac{\Delta_2(m)}{n+M}.$$

$$(iv)\ \mathbb{E}[\sup_{t\in B(m,\hat m)} \nu_{n,M,5}^2(t) - p_3(m,\hat m)]_+ \le \frac{c}{M}$$

The bounds partly rely on Lemma 2 proved in Section 8.2.

8.4. **Proof of Lemma 3.** Proof of (i). Let $m^* = m\wedge \hat m$. We have, by using (a) of Lemma 2, for $t\in B(m,\hat m)$,

$$\begin{aligned} |\nu_{n,M,2}(t)|^2 &= \left| \int t^*(-u)(\hat f_Y^*(u) - f_Y^*(u))S(u)du \right|^2 \le \int_{-\pi m^*}^{\pi m^*} |\hat f_Y^*(u) - f_Y^*(u)|^2 |S(u)|^2 du \\ &\le \int_{-\pi m^*}^{\pi m^*} |\hat f_Y^*(u) - f_Y^*(u)|^2 \frac{M}{(f_\varepsilon^*)^2(u)} (\widehat{(f_\varepsilon^*)^2}(u) - (f_\varepsilon^*)^2(u))^2 du. \end{aligned}$$

Let us define

$$\Omega_1(u) = \{u,\ \ |\hat f_Y^*(u) - f_Y^*(u)| \le 4\sqrt{\log(n+M)}/\sqrt{n+M}\}$$

and

$$\Omega_2(u) = \{u,\ \ |\widehat{(f_\varepsilon^*)^2}(u) - (f_\varepsilon^*)^2(u)| \le 4\sqrt{\log(M)}/\sqrt{M}\}.$$

We have

$$\begin{aligned} \sup_{t\in B(m,\hat m)} |\nu_{n,M,2}(t)|^2 &\le \frac{\log(M)\log(n+M)}{n+M} \int_{-\pi m^*}^{\pi m^*} \frac{1}{(f_\varepsilon^*)^2(u)} du \\ &\quad + \int_{-\pi m^*}^{\pi m^*} |\hat f_Y^*(u) - f_Y^*(u)|^2 \frac{M}{(f_\varepsilon^*)^2(u)} (\widehat{(f_\varepsilon^*)^2}(u) - (f_\varepsilon^*)^2(u))^2 \mathbf{1}_{(\Omega_1(u)\cap\Omega_2(u))^c} du. \end{aligned}$$

Now, by using Bernstein Inequality, there exists a constant C such that

$$\mathbb{P}[(\Omega_1(u))^c] \le \frac{C}{(n+M)^2}, \quad \text{and} \quad \mathbb{P}[(\Omega_2(u))^c] \le \frac{C}{M^2}.$$

Moreover, Rosenthal Inequality yields

$$\mathbb{E}^{1/4}[|\hat f_Y^*(u) - f_Y^*(u)|^8] \le C/(n+M), \text{ and } \mathbb{E}^{1/4}[(\widehat{(f_\varepsilon^*)^2}(u) - (f_\varepsilon^*)^2(u))^8] \le C/M.$$

Therefore

$$
\begin{aligned}
&\mathbb{E}\left[\left(\sup_{t\in B(m,\hat{m})}|\nu_{n,M,2}(t)|^2-\frac{\log(M)\log(n+M)}{n+M}\Delta_2(m^*)\right)_+\right]\\
\leq\;& \int_{-\pi m(n,M)}^{\pi m(n,M)}\mathbb{E}^{1/4}[|\hat{f}_Y^*(u)-f_Y^*(u)|^8]\frac{M}{f_\varepsilon^{*2}(u)}\mathbb{E}^{1/4}[(\widehat{(f_\varepsilon^*)^2}(u)-(f_\varepsilon^*)^2(u))^8]\\
&\qquad\qquad\left\{\mathbb{P}^{1/2}[(\Omega_1(u))^c]+\mathbb{P}^{1/2}[(\Omega_2(u))^c]\right\}du\\
\leq\;& C\frac{\Delta_2(m(n,M))}{n+M}(\frac{1}{n+M}+\frac{1}{M})\leq C(\frac{1}{n+M}+\frac{1}{M})
\end{aligned}
$$

since $m(n,M)$ is chosen to ensure that $\Delta_2(m(n,M))/(n+M)\leq C$. This ends the proof of (i) of Lemma 3. □

Proof of (ii).
We have $\nu_{n,M,3}(t)=\int t^*(-u)f^*(u)f_\varepsilon^*(u)S(u)du$ and therefore, for $t\in B(m,\hat{m})$,

$$|\nu_{n,M,3}(t)|^2\leq\int|f^*(u)f_\varepsilon^*(u)S(u)|^2du.$$

With the same method as in (i) but using only $\Omega_2(u)$, we can obtain, by using (b) of Lemma 2, the part $\log(M)\Delta_2^{(f)}(m)/\sqrt{M}$, and by using (c), the part $(\log(M))^{3/2}\Delta_4^{(f)}(m)/M$. □

Proof of (iii). We have

$$\nu_{n,M,4}(t)=\int t^*(-u)(\hat{f}_Y^*(u)-f_Y^*(u))\frac{1}{f_\varepsilon^*(u)}\mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u)<M^{-1/2}}du$$

and thus

$$
\begin{aligned}
|\nu_{n,M,4}(t)|^2\leq\int_{-\pi m^*}^{\pi m^*}\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^2}{|f_\varepsilon^*(u)|^2}du &= \sup_{t\in B(m,\hat{m})}\left|\int t^*(-u)\frac{\hat{f}_Y^*(u)-f_Y^*(u)}{f_\varepsilon^*(u)}du\right|^2\\
&= \sup_{t\in B(m,\hat{m})}|\nu_{n,M,1}(t)|^2
\end{aligned}
$$

which implies the result. □

Proof of (iv). We have, for $t \in B(m, \hat{m})$,

$$
\begin{aligned}
|\nu_{n,M,5}(t)|^2 &= \left| \int t^*(-u) f^*(u) \mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) < M^{-1/2}} du \right|^2 \leq \int_{\pi m^*}^{\pi m^*} |f^*(u)|^2 \mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) < M^{-1/2}} du \\
&\leq \int_{\pi m^*}^{\pi m^*} |f^*(u)|^2 \mathbf{1}_{|(f_\varepsilon^*)^2(u)| \leq 2/\sqrt{M}} \mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) < M^{-1/2}} du \\
&\quad + \int_{\pi m^*}^{\pi m^*} |f^*(u)|^2 \mathbf{1}_{|(f_\varepsilon^*)^2(u)| > 2/\sqrt{M}} \mathbf{1}_{\widehat{(f_\varepsilon^*)^2}(u) < M^{-1/2}} du \\
&\leq \left(\frac{2}{\sqrt{M}} \int_{\pi m^*}^{\pi m^*} \frac{|f^*(u)|^2}{(f_\varepsilon^*)^2(u)} du \right) \wedge \left(\frac{4}{M} \int_{\pi m^*}^{\pi m^*} \frac{|f^*(u)|^2}{f_\varepsilon^{*4}(u)} du \right) \\
&\quad + \int_{\pi m^*}^{\pi m^*} |f^*(u)|^2 \mathbf{1}_{|\widehat{(f_\varepsilon^*)^2}(u) - (f_\varepsilon^*)^2(u)| > (f_\varepsilon^*)^2(u)/2} du \\
&\leq 4 \frac{\Delta_2^{(f)}(m^*)}{\sqrt{M}} \wedge \frac{\Delta_4^{(f)}(m^*)}{M} \\
&\quad + 4 \int_{\pi m^*}^{\pi m^*} |f^*(u)|^2 \left(\frac{|\widehat{(f_\varepsilon^*)^2}(u) - (f_\varepsilon^*)^2(u)|}{(f_\varepsilon^*)^2(u)} \wedge \frac{|\widehat{(f_\varepsilon^*)^2}(u) - (f_\varepsilon^*)^2(u)|^2}{f_\varepsilon^{*4}(u)} \right) du
\end{aligned}
$$

and using the set $\Omega_2(u)$ again yields the result.□

MAP5, UMR CNRS 8145, Université Paris Descartes, 45 rue des Saints Pères, 75006 Paris, France