Deconvolution estimation of onset of pregnancy with replicate observations

Fabienne Comte, Adeline Samson, Julien J. Stirnemann

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 Paris Descartes.*

Abstract. Except in the specific case of in vitro fertilization, the precise date of onset of pregnancy is unknown. In clinical practice, the date of pregnancy may only be estimated, and most commonly from ultrasound biometric measurements of the embryo. Denoting X the interval between last menstrual period and true onset of pregnancy and Y the interval between last menstrual period and the date estimated by ultrasound, we wish to estimate the density f of X. Only noisy observations $Y_j = X_j + \varepsilon_j, j = 1, \dots, n$ are observed and the density f_ε of ε is unknown. Because the noise itself cannot be sampled for the estimation of its density, we consider the specific setting of replicate noisy observations $Y_{-j1} = X_{-j} + \varepsilon_{-j1}$ and $Y_{-j2} = X_{-j} + \varepsilon_{-j2}$, $j = 1, \dots, M$. We suggest an adaptive non-parametric estimator of f built following a deconvolution device. Convergence rates are studied and compared to other settings that do not involve replicates. Lastly, we estimate the density f 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 without assisted reproductive techniques, 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 (US) 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 analyses 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 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 desnities 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:

$$Y_{-j1} = X_{-j} + \varepsilon_{-j1}, \quad Y_{-j2} = X_{-j} + \varepsilon_{-j2}, \quad j = 1, \dots, M \tag{2}$$

with X_{-j}, ε_{-j1} and ε_{-j2}, for $j = 1, \dots, M$, independent and identically distributed. Moreover the sequences $(X_{-j})_{1 \le j \le M}$, $(\varepsilon_{-j1})_{1 \le j \le M}$ and $(\varepsilon_{-j2})_{1 \le j \le M}$ are independent. These noisy observations could be replicate measurements of CRL of the same embryo or measurements of CRL in twin pregnancies for example. 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 [2010]. Our approach suggests a different estimator that is related to the truncated estimator of Neumann [1997]. We also use an adaptive cut-off selection with a penalization device, following the procedure described in Comte and Lacour [2011].

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 when M is large compared to n. 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 respective 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.

The convolution problem may be written as $f_Y(x) = f \star f_\varepsilon(x) = \int f(x-t)f_\varepsilon(t)dt$ where $\star$ denotes the convolution operator. Using the characteristic functions, we have $f_Y^*(u) = f^*(u)f_\varepsilon^*(u)$. 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.

In the following, we consider the model described by the two samples (1) and (2). A noise sample is not available and we suggest that the estimation of f_ε^* relies upon replicate observations given in (2) instead.

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

A1. *We assume ε is symmetric and that its characteristic function never vanishes.*

Since assuming ε symmetric is equivalent to assuming f_ε^* real-valued, we can see that **A1** is equivalent to the assumption 2.2 in Delaigle et al. [2008]. Therefore, assumption **A1** implies that

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

Under this reasonable assumption, we have:

$$\mathbb{E}\big(e^{it(\varepsilon_{-j1}-\varepsilon_{-j2})}\big) = |\mathbb{E}\big(e^{it\varepsilon_{-j1}}\big)|^2 \overset{\text{A1}}{=} \Big(\mathbb{E}\big(e^{it\varepsilon_{-j1}}\big)\Big)^2 = \big(f_\varepsilon^*(t)\big)^2.$$

Therefore, given that $\mathbb{E}\big(e^{it(\varepsilon_{-j1}-\varepsilon_{-j2})}\big) = \mathbb{E}\big(e^{it(Y_{-j1}-Y_{-j2})}\big)$ and under the hypothesis **A1**, we have the following estimation of $(f_\varepsilon^*)^2$:

$$\widehat{(f_\varepsilon^*)^2}(t) = \frac{1}{M}\sum_{j=1}^{M} \cos\big(t(Y_{-j1} - Y_{-j2})\big) \tag{3}$$

We also define a truncated estimate of $1/f_\varepsilon^*$:

$$\frac{1}{\tilde{f}_\varepsilon^*(t)} = \frac{\mathbb{1}_{\widehat{(f_\varepsilon^*)^2}(t)\geq M^{-1/2}}}{\sqrt{\widehat{(f_\varepsilon^*)^2}(t)}} \tag{4}$$

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

$$\hat{f}_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-ixt}\frac{\hat{f}_Y^*(t)}{\tilde{f}_\varepsilon^*(t)}dt, \text{ where } \hat{f}_Y^*(t) = \frac{1}{n}\sum_{j=1}^n e^{itY_j}. \tag{5}$$

For further discussion, we recall the definitions of the estimators given in (3), (4) and (5) for the cases (i) with known density of noise, denoted with superscript KN and (ii) with an independent sample (ε_{-j}) of noise observations, denoted with superscript ON (see Comte and Lacour [2011]).
(i) In the case of a known density f_ε of the noise, the estimate of f is:

$$\hat{f}_m^{KN}(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-ixt}\frac{\hat{f}_Y^*(t)}{f_\varepsilon^*(t)}dt$$

(ii) In the case of an unknown density but with a sample of (ε_{-j}), the estimate of the characteristic function f_ε^* is

$$\hat{f}_\varepsilon^{ON*}(t) = \frac{1}{M}\sum_{j=1}^M e^{it\varepsilon_{-j}}$$

and its truncated estimate

$$\frac{1}{\tilde{f}_\varepsilon^{ON*}(t)} = \frac{\mathbb{1}_{|\hat{f}_\varepsilon^{ON*}(t)|\geq M^{-1/2}}}{\hat{f}_\varepsilon^{ON*}(t)}.$$

The estimate of f in this case is

$$\hat{f}_m^{ON}(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-ixt}\frac{\hat{f}_Y^*(t)}{\tilde{f}_\varepsilon^{ON*}(t)}dt$$

These estimators differ from (5) in several ways:

(a) First, because of replications, we can only estimate $(f_\varepsilon^*)^2$ and not directly f_ε^* as in case (ii).

(b) Subsequently, the levels of truncations required will also differ in (4) and (ii). This directly arises from the previous comment since the truncation will be in $M^{-1/4}$ in (4), whereas it is in $M^{-1/2}$ in the ON case (ii).

(c) Finally, compared to our setting and (ii) which share the fact that f_ε is unknown, the case (i) does not require a truncation and the estimator is the classical expression of density estimation by deconvolution with a Fourier transform of compact support. The risk bound obtained in this case is considered as a benchmark of the best reachable bound.

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

Let us define f_m such that $f_m^* = f^* \mathbb{1}_{[-\pi m, \pi m]}(.)$. 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*

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

Proof. See appendix

Let us define, for $k \in \mathbb{N}^*$,

$$\Delta_k(m) = \int_{-\pi m}^{\pi m} \frac{dt}{|f_\varepsilon^*(t)|^k} \text{ and } \Delta_k^{(f)}(m) = \int_{-\pi m}^{\pi m} \frac{|f^*(t)|^2}{|f_\varepsilon^*(t)|^k} dt$$

as well as their estimates

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

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 (5). Then there exists a constant C such that*

$$\mathbb{E}(\|f - \hat{f}_m\|^2) \leq \|f - f_m\|^2 + C\left(\frac{\Delta_2(m)}{n} + \frac{\Delta_4(m)}{n\sqrt{M}} \wedge \frac{\Delta_6(m)}{nM} + \frac{\Delta_2^{(f)}(m)}{\sqrt{M}} \wedge \frac{\Delta_4^{(f)}(m)}{M}\right). \quad (6)$$

Proof. See appendix

We recognize in this last inequality (6) the bias $\|f - f_m\|^2$ and variance denoted $\mathrm{Var}(m) := Q_1(m) + Q_2(m) + Q_3(m)$ with

$$\begin{aligned} Q_1(m) &:= \Delta_2(m)/n \\ Q_2(m) &:= \Delta_4(m)/(n\sqrt{M}) \wedge \Delta_6(m)/(nM) \\ Q_3(m) &:= \Delta_2^{(f)}(m)/\sqrt{M} \wedge \Delta_4^{(f)}(m)/M. \end{aligned}$$

We can also recognize $Q_1(m)$ as the variance term that arises alone when f_ε^* is assumed as known (see Comte et al. [2006]). The following terms $Q_2(m)$ and $Q_3(m)$ are specific

to our setting involving replicate observations and show the loss in the resulting rates. In the case of observed noise (see Comte and Lacour [2011]), the upper bound is

$$\|f - f_m\|^2 + CQ_1(m) + (C+2)\Delta_2^{(f)}(m)/M.$$

Therefore, compared to the case with observed noise or known density, the terms $Q_2(m)$ and $Q_3(m)$ resulting from the estimation of $(f_\varepsilon^*)^2$ with replicate observations are a substantial step in complexity. A theoretical and an empirical study of the behavior of each of these terms is presented later.

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

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*

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

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

Let us also consider the set of admissible values of m, defined as follows. Let $\kappa > 0$ and C_0 be fixed, and $\lfloor z \rfloor$ denoting the integer part of z,

$$m \in \mathcal{M}_n = \{m = k/\kappa, k = 1, \ldots, \lfloor n\kappa \rfloor, \ \ \mathrm{Var}(m) \leq C_0\}.$$

Then, under **A2**, Lemma 3 in Comte and Lacour [2011] states that: $\forall p \geq 1/2$, there exists a constant c_p such that

$$\Delta_{4p}(m) \leq c_p \left(\Delta_2(m)\right)^{2p} \left(\log(n)\right)^{(2p-1)\mathbb{1}_{\gamma>1}}. \tag{7}$$

This enables us to prove the following result

Proposition 2. *Assume that assumptions* **A1** *and* **A2** *hold and assume that* $M \geq n^2(\log(n))^{2\mathbb{1}_{\gamma>1}}$*. Then, there exists a constant* C *such that,* $\forall m \in \mathcal{M}_n$*,*

$$\mathbb{E}(\|f - \hat{f}_m\|^2) \leq \|f - f_m\|^2 + C\frac{\Delta_2(m)}{n} \tag{8}$$

Proof. First, if $m \in \mathcal{M}_n$, then $\Delta_2(m)/n \leq C_0$. Next, by (7) recalled above, we have

$$Q_2(m) \leq \frac{\Delta_4(m)}{n\sqrt{M}} \leq c_1 \frac{\Delta_2^2(m)(\log(n))^{\mathbb{1}_{\gamma>1}}}{n\sqrt{M}}.$$

Then as $M \geq n^2(\log(n))^{2\mathbb{1}_{\gamma>1}}$ and with the initial remark, we get

$$Q_2(m) \leq c_1 \left(\frac{\Delta_2(m)}{n}\right)^2 \leq c_1 C_0 \frac{\Delta_2(m)}{n}.$$

Lastly, as $|f^*(t)| \leq 1$, $\forall t \in \mathbb{R}$, we have $\Delta_2^{(f)}(m) \leq \Delta_2(m)$ and with the condition on M:

$$Q_3(m) \leq \frac{\Delta_2^{(f)}(m)}{\sqrt{M}} \leq \frac{\Delta_2(m)}{n}.$$

It follows that $\mathrm{Var}(m) \leq c_1(C_0+2)\Delta_2(m)/n$ which implies the result of Proposition 2. □

Note that in addition, we know that, under **A2**, the dominating variance term has the following order (see Comte et al. [2006], Comte and Lacour [2011]):

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

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

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 regularity of f^*. Classically, we consider the following smoothness spaces for density f on $\mathbb{R}$:

A3. $f \in \mathcal{A}_{a,b,c}(l) = \{f \in \mathbb{L}^1 \cap \mathbb{L}^2,\ \int |f^*(x)|^2(x^2+1)^a \exp(2b|x|^c)dx \leq l\}$ *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 supersmooth, and as ordinary smooth otherwise. The spaces of ordinary smooth functions correspond to classical Sobolev classes, while supersmooth 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:

$$\|f - f_m\|^2 \leq Cm^{-2a}\exp(-2b(\pi m)^c). \tag{9}$$

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

Proposition 3. *We consider assumptions* **A1**, **A2**, **A3** *and the ordinary smooth case for both* f *and* f_ε *with* $c = \gamma = 0$. *The bound (6) then becomes*

$$\mathbb{E}(\|f - \hat{f}_m\|^2) \leq C\left(m^{-2a} + \frac{m^{2\alpha+1}}{n}\left(1 + \frac{m^{2\alpha}}{\sqrt{M}} \wedge \frac{m^{4\alpha}}{M}\right) + \frac{m^{2(\alpha-a)_+}}{\sqrt{M}} \wedge \frac{m^{2(2\alpha-a)_+}}{M}\right).$$

If moreover $M \geq n^2$, $\forall m \in \mathcal{M}_n$, *we have*

$$\mathbb{E}(\|f - \hat{f}_m\|^2) \leq C\left(m^{-2a} + \frac{m^{2\alpha+1}}{n}\right). \tag{10}$$

The first inequality is a consequence of (6), (9) and elementary bounds for the variance. Despite assumptions, the resulting rate in the ordinary smooth case presented in Proposition 3 is still practically intractable. Inequality (10) is a consequence of Proposition 2 and allows us to recover deconvolution rates as if the noise had known density.

Indeed, we can see on (10) that the estimator can reach the rate $n^{-2a/(2a+2\alpha+1)}$ for $m \propto n^{1/(2\alpha+2a+1)}$. 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 result (10) of Proposition 3 can be extended to the general case of $f \in \mathcal{A}_{a,b,c}(l)$ with more stringent conditions on M with regard to n.

Proposition 4. *Assume that assumptions* **A1**, **A2**, **A3** *are fulfilled. If* $M \geq (n\log(n))^2$, $\forall m \in \mathcal{M}_n$, *then there exists a constant* C *such that*

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

3.3. **Discussion about the relative orders of M and n.** In the case $M \geq (n\log(n)^{\mathbb{1}_{\gamma>1}})^2$, Proposition 2 clearly implies that the variance term defined in (6) may be simply $\mathrm{Var}(m) = CQ_1(m) = C\Delta_2(m)/n$ where C is a numerical constant.

In the general case, the relationship between n and M may influence the resulting rates especially when $M \ll n$.

An estimation of $\mathrm{Var}(m)$ is $\widehat{\mathrm{Var}}(m) = \hat{Q}_1(m) + \hat{Q}_2(m) + \hat{Q}_3(m)$ with

$$\begin{aligned}
\hat{Q}_1(m) &= \frac{\hat{\Delta}_2(m)}{n} \\
\hat{Q}_2(m) &= \frac{\hat{\Delta}_4(m)}{n\sqrt{M}} \wedge \frac{\hat{\Delta}_6(m)}{nM} \\
\hat{Q}_3(m) &= \frac{\hat{\Delta}_2^{(f)}(m)}{\sqrt{M}} \wedge \frac{\hat{\Delta}_4^{(f)}(m)}{M}
\end{aligned}$$

An example of empirical behavior of $\hat{Q}_1(m)$, $\hat{Q}_2(m)$ and $\hat{Q}_3(m)$ is depicted in Figure 1 for $M \ll n$ and $M \gg n$ and in the setting of ordinary smooth or supersmooth functions for f chosen as Laplace and Gaussian respectively, together with an ordinary smooth Laplace noise. Expectedly, following Proposition 2, $\hat{Q}_1(m)$ majorates the other terms when $M \gg n$. Interestingly, this seems to be also true when $M \ll n$, at least empirically. This finding was invariable throughout simulations, thus making $\widehat{\mathrm{Var}}(m) \simeq C\hat{\Delta}_2(m)/n$ an appropriate choice regardless of the respective values of n and M. The previous considerations will become prominent in the choice of a penalty in Section 4.

4. Model selection

The general outline of the method used to select the parameter m among all considered indexes $\mathcal{M}_n$ is borrowed from Comte and Lacour [2011]. Our approach aims to select $m \in \mathcal{M}_n$ 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

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

The theoretical optimal choice of m is defined as:

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

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 (5), we will consider the following preliminary

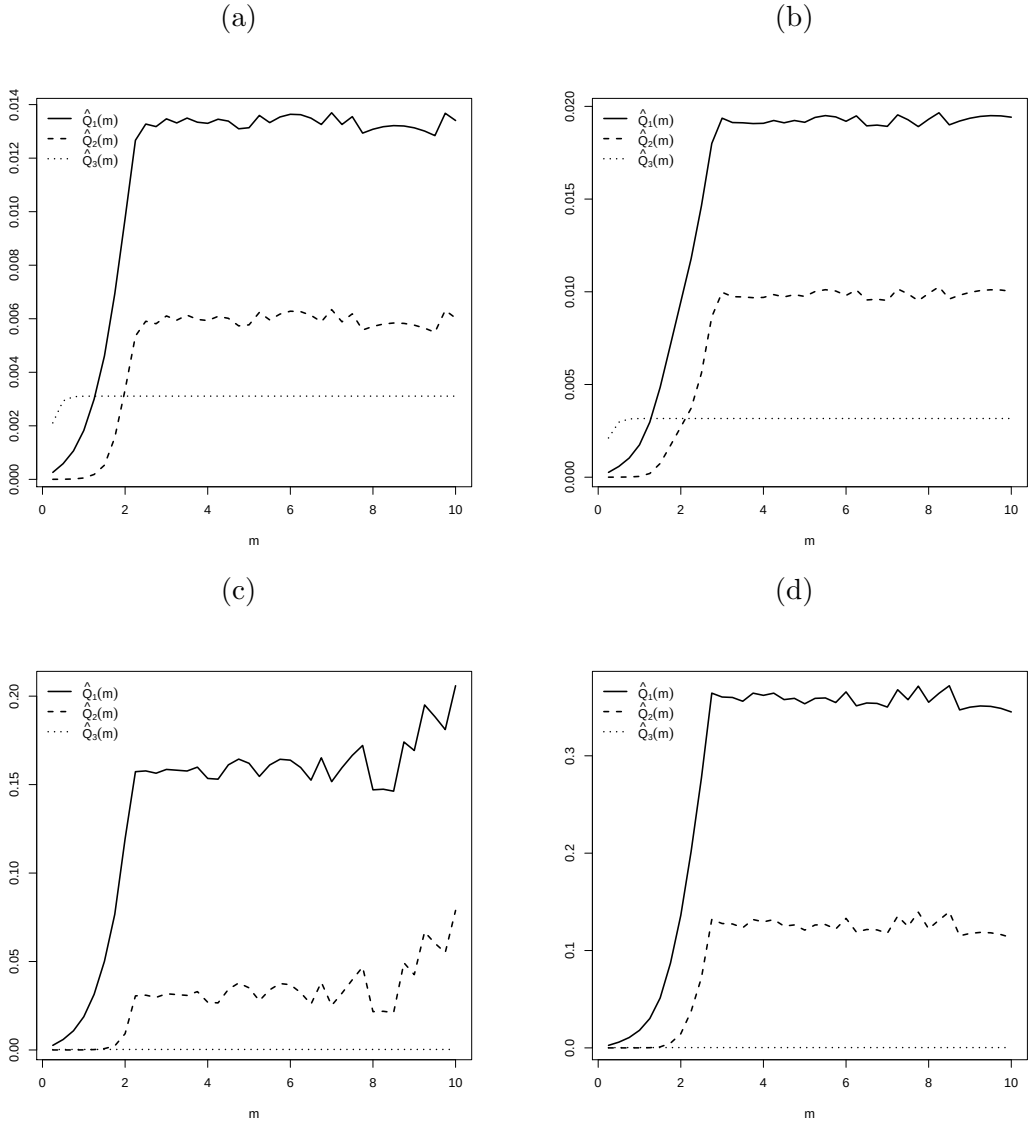

FIGURE 1. Empirical behavior of $\hat{Q}_1(m)$, $\hat{Q}_2(m)$ and $\hat{Q}_3(m)$ as a function of m when $M \ll n$ and $M \gg n$ for f ordinary smooth and supersmooth, chosen as Laplace (panels (a) and (c)) and Gaussian (panels (b) and (d)) respectively. The influence of M and n is illustrated by $M = 200$, $n = 2000$ (figures (a) and (b)) and $M = 2000$, $n = 200$ (figures (c) and (d)).

estimate of m:

$$\hat{m} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \Big(- \|\hat{f}_m\|^2 + \operatorname{pen}(m) \Big) \tag{11}$$

where $\text{pen}(m)$ is a penalty term related to the variance that remains to be defined.

4.1. **Choice of an appropriate penalty.** The penalty term is related to the variance $\text{Var}(m)$ as defined by (6). Therefore, as discussed in the previous section, the choice of an appropriate penalty will depend on the relative values of n and M:

(i) When $M \geq (n\log(n)^{\mathbb{1}_{\gamma>1}})^2$, following Proposition 2, we have shown that $\text{Var}(m) \simeq C\Delta_2(m)/n$. Therefore, based upon Theorem 1 in Comte and Lacour [2011], we define $\text{pen}(m)$ as

$$\text{pen}(m) = K_0 \left(\frac{\log\left(\Delta_2(m)\right)}{\log(m+1)} \right)^2 \frac{\Delta_2(m)}{n} \tag{12}$$

where K_0 is a numerical constant. Note that the term $\log\left(\Delta_2(m)\right)/\log(m+1)$ has the order of a constant when the noise is ordinary smooth and yields only a slight correction when the noise is supersmooth. This quantity has been introduced by Comte and Lacour [2011] to take into account negligible loss in the rates that sometimes (not systematically) occur in this last case.

(ii) When $M \ll n$, the choice of an appropriate penalty is difficult since we did not obtain a majoration under reasonable assumptions in this case. However, the empirical behavior of $\hat{Q}_1(m)$, $\hat{Q}_2(m)$ and $\hat{Q}_3(m)$ seems to suggest that $\widehat{\text{Var}}(m) \simeq C\hat{\Delta}_2(m)/n$ (see Figure 1). Therefore, the penalty defined in (12) may be a safe choice to be plugged into (11) in this case as well.

In the following we will consider that the penalty defined by (12) is an appropriate choice regardless of the respective values of M and n.

4.2. **Estimation procedure.** We define $\widehat{\text{pen}}(m)$ the estimator of $\text{pen}(m)$ as:

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

Throughout numerical estimations we will consider $K_1 = 2$, after a set of simulation experiments to calibrate it. The computation of $\|\hat{f}_m\|$ is performed following Comte and Lacour [2011] and Comte et al. [2006] by considering the following expression of the estimator (5) as an orthogonal projection. We denote $\varphi(x) = \sin(\pi x)/\pi x$ the sinus cardinal function. The orthonormal basis $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ is thus defined by $\varphi_{m,j}(x) = \sqrt{m}\varphi(mx - j)$, $m \in \mathbb{N}^*, j \in \mathbb{Z}$ and $\varphi^*_{m,j}(x) = e^{-ixj/m}\mathbb{1}_{[-\pi m,\pi m]}(x)/\sqrt{m}$. We also recall from Comte and Lacour [2011] the estimated projection coefficients

$$\hat{a}_{m,j} = \frac{\sqrt{m}}{2}(-1)^l \int_0^2 e^{il\pi x} \frac{\hat{f}^*_Y}{\tilde{f}^*_\varepsilon}(\pi m(x-1))dx.$$

The estimator (5) may then be expressed:

$$\hat{f}_m = \sum_{\ell\in\mathbb{Z}} \hat{a}_{m,\ell}\varphi_{m,\ell} \tag{13}$$

This expression of the estimator allows us to use Inverse Fast Fourrier Transform (IFFT) Algorithms in the estimation process. Therefore, for numerical tractability we use only

a finite sample of projection coefficients with $\hat{f}_m = \sum_{|\ell| \leq K_n} \hat{a}_{m,\ell} \varphi_{m,\ell}$. Theroetical results assert that $K_n = n$ always suits (see Comte et al. [2006]) but we make the constant choice $K_n = 255$.

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

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

by defining an integer $\hat{m}_n$ such that

$$\hat{m}_n = \text{argmax}\Big(m \in \mathbb{N}, \frac{\hat{\Delta}_2(m)}{n} \leq 1\Big)$$

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

$$\widehat{\hat{m}} = \underset{m \in \{1,\ldots,\hat{m}_n\}}{\text{argmin}} \Big(- \|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)\Big) \tag{14}$$

Finally, by plugging (14) in (13) 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. Simulation study

5.1. **Design of simulation.**

5.1.1. *Density of noise.* Noise was given a Laplace or a Gaussian density with variance σ^2 as follows:

- Laplace noise.

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

- Gaussian noise.

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

We compared our results to estimations under the assumption of a known noise density (see Comte et al. [2006], Comte and Lacour [2011] for the description of the estimation procedure and penalties for Gaussian and Laplace noises).

5.1.2. *Densities for f.* Following Comte et al. [2006], we considered the following four different densities:

(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: $X \sim \mathcal{N}(0,1)$

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 (v) 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].

5.2. **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. The results for the known noise case are similar with those reported in Comte et al. [2006], Comte and Lacour [2011]. 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 supersmooth 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].

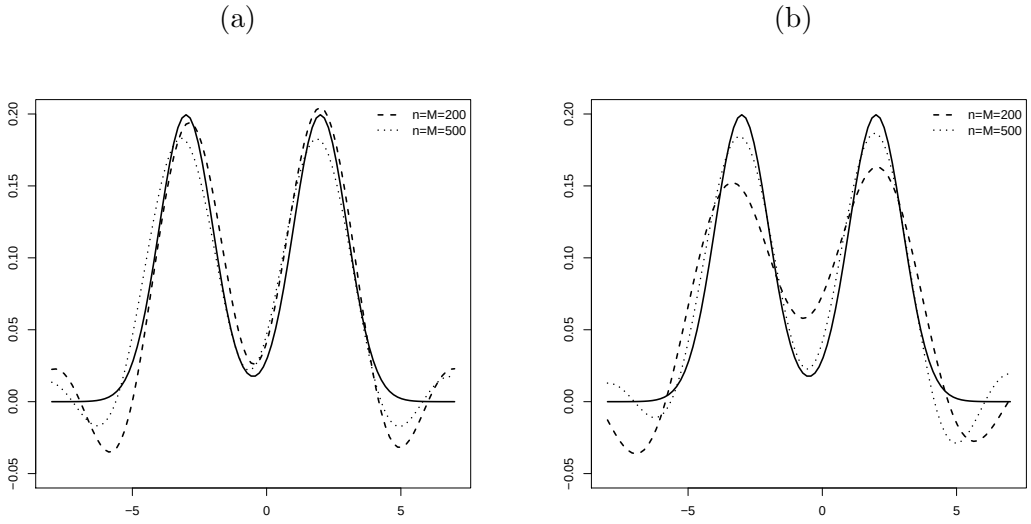

FIGURE 2. Estimations for $n = M = 200$ (dotted line) and $n = M = 500$ (dashed line) for the Gaussian mixture density (v). (a): Laplace noise; (b): Gaussian noise. Two independent samples were used, of size n and M respectively.

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 denoted $\hat{f}_m^{KN}$.

ε **Gaussian**		$s2n = 10$		$s2n = 5$	
		$n = 200$	$n = 2000$	$n = 200$	$n = 2000$
f Mixed Gamma	f_ε known	0.746	0.171	0.901	0.740
	$M = \sqrt{n}$	0.535	0.126	0.837	0.261
	$M = n$	0.540	0.126	0.797	0.264
f Cauchy	f_ε known	0.642	0.081	1.052	0.561
	$M = \sqrt{n}$	0.395	0.175	0.861	0.213
	$M = n$	0.394	0.175	0.834	0.204
f Laplace	f_ε known	3.045	0.841	4.440	3.768
	$M = \sqrt{n}$	2.408	2.203	3.862	2.690
	$M = n$	2.403	2.201	3.726	2.232
f Gaussian	f_ε known	0.269	0.061	0.826	0.543
	$M = \sqrt{n}$	0.301	0.052	0.752	0.092
	$M = n$	0.291	0.052	0.558	0.085
ε **Laplace**		$s2n = 10$		$s2n = 5$	
		$n = 200$	$n = 2000$	$n = 200$	$n = 2000$
f Mixed Gamma	f_ε known	0.740	0.099	0.910	0.173
	$M = \sqrt{n}$	0.526	0.121	0.758	0.203
	$M = n$	0.535	0.121	0.743	0.206
f Cauchy	f_ε known	0.555	0.179	0.750	0.238
	$M = \sqrt{n}$	0.417	0.175	0.697	0.198
	$M = n$	0.404	0.175	0.658	0.194
f Laplace	f_ε known	1.581	0.580	2.808	1.222
	$M = \sqrt{n}$	2.408	2.199	3.603	2.362
	$M = n$	2.377	2.196	2.936	2.217
f Gaussian	f_ε known	0.741	0.330	0.661	0.427
	$M = \sqrt{n}$	0.313	0.054	0.632	0.079
	$M = n$	0.302	0.053	0.555	0.076

Our estimator differs from the one provided by Delaigle et al. [2008] in several ways. Whereas Delaigle et al. [2008] uses a single sample containing replicate noisy observations for the estimation of f, we use two independent samples. The first sample, of size M, comprises replicate noisy observations, and is used for the estimation of f_ε^* and the second, of size n, containing single observations is used for the estimation of f. This may be useful when only a small amount of replicate data is available. Also, our method is based on a data driven selection of the cutoff parameter through a penalization device whereas Delaigle et al. [2008] uses kernel methods with a plug-in bootstrap estimation of

TABLE 2. Comparison of the ISE between the estimators of Delaigle et al. [2008] and the penalized estimator for the Gaussian mixture density (v). For the sake of comparison, the results are presented by the median × 100 (inter-quartile range × 100) of 100 estimations.

	Delaigle et al. [2008]		Penalized estimator with two samples		Penalized estimator with one sample	
	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
ε Laplace	1.41 (0.94)	0.89 (0.51)	1.28 (0.24)	1.16 (0.85)	1.22 (0.55)	0.29 (0.13)
ε Gaussian	2.09 (1.33)	1.42 (0.92)	1.28 (0.31)	1.19 (0.11)	1.27 (0.19)	1.13 (0.93)

the bandwidth parameter (see Delaigle and Gijbels [2004]). Table 2 presents the comparison of the penalized estimator and the estimator given by Delaigle et al. [2008] for the Gaussian mixture density (v). We studied two cases: (i) two separate samples are used for the estimation of f_ε^* and f respectively; (ii) the same sample, containing two replicate observations is used for the estimation of f_ε^* and f as in the estimator of Delaigle et al. [2008]. The integrated squared error (ISE) is computed over 100 estimations and we present the results using the median and inter-quartile range (IQR). In most cases, the penalized estimator enjoys lower risks compared to those given by Delaigle et al. [2008]. Furthermore, these risks seem to be improved when using only one sample compared to the two samples setting.

In Figure 2, we present an estimation of f using the penalized estimator in the two samples setting as in the simulated case (i) aforementioned. We considered the Gaussian mixture distribution (v) 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.

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) 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 $Y_{-j1} = X_j + \varepsilon_{-j1}$ and $Y_{-j2} = X_{-j} + \varepsilon_{j2}$, $-j = 1, \ldots, M$. We wish to estimate f which represents the distribution of probability of onset of pregnancy within a female cycle.

(a) (b)

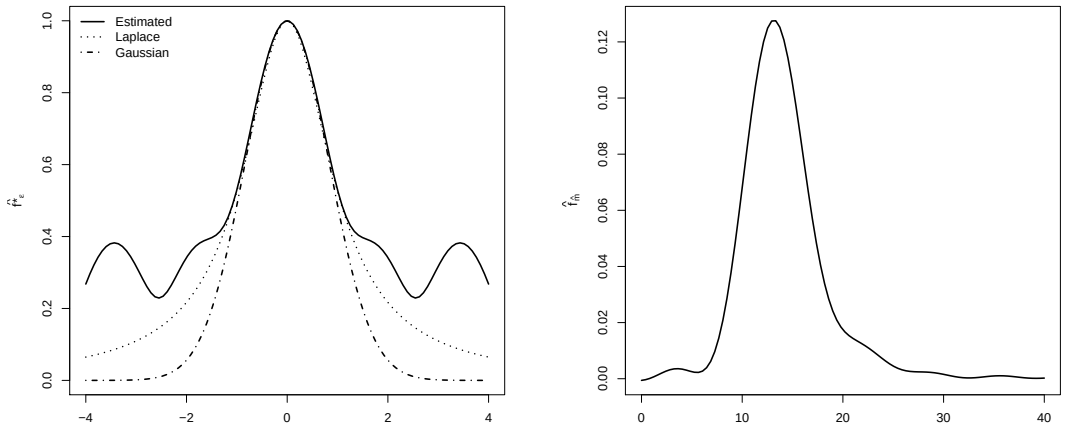

FIGURE 3. (a): estimation of the characteristic function of the noise in twin pregnancies. Gaussian and Laplace characteristic functions are plotted for comparison. (b): estimation of the density of onset of pregnancy.

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 is 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 Gaussian and Laplace characteristic functions). Under this simulation model, the risk MISE$\times$100 was 0.05 and 0.04 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 exhibits 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 module. However, when the sample size M of replicate observations is

large compared to the sample size n of single observations, the resulting rates are satisfying compared to usual deconvolution settings. Regardless of theoretical convergence rates, 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 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, but 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]).

Appendix A. Proofs

A.1. **Proof of Lemma 1.** The proof is given for $p = 1$. Let us denote $R(t) = \dfrac{1}{\tilde{f}^*_\varepsilon(t)} - \dfrac{1}{f^*_\varepsilon(t)}$.

$$\begin{aligned}
\mathbb{E}(|R(t)|^2) &= \mathbb{E}\left(\frac{\mathbb{1}_{\widehat{(f^*_\varepsilon)^2}(t)<M^{-1/2}}}{(f^*_\varepsilon)^2(t)}\right) \\
&\quad + \frac{1}{(f^*_\varepsilon)^2(t)}\mathbb{E}\left(\mathbb{1}_{(\hat{f^*_\varepsilon})^2(t)\geq M^{-1/2}}\frac{(\widehat{(f^*_\varepsilon)^2}(t)-(f^*_\varepsilon)^2(t))^2}{\widehat{(f^*_\varepsilon)^2}(t)(\sqrt{\widehat{(f^*_\varepsilon)^2}(t)}+f^*_\varepsilon(t))^2}\right) \\
&\leq \frac{1}{(f^*_\varepsilon)^2(t)} + \frac{M^{1/2}}{(f^*_\varepsilon)^4(t)}\mathbb{E}[(\widehat{(f^*_\varepsilon)^2}(t)-(f^*_\varepsilon)^2(t))^2] \\
&\leq \frac{1}{(f^*_\varepsilon)^2(t)} + \frac{M^{-1/2}}{(f^*_\varepsilon)^4(t)} \leq \frac{3}{2}\frac{M^{-1/2}}{(f^*_\varepsilon)^4(t)}
\end{aligned}$$

(i) if $(f^*_\varepsilon)^2(t) \leq 2M^{-1/2}$, we have $M^{-1/2}/(f^*_\varepsilon)^4(t) \leq 2M^{-1}/(f^*_\varepsilon)^6(t)$.
(ii) If $(f^*_\varepsilon)^2 > 2M^{-1/2}$, using the Bernstein Inequality as in Neumann yields:

$$\begin{aligned}
\mathbb{P}(|(\hat{f^*_\varepsilon})^2(t)| < M^{-1/2}) &\leq \mathbb{P}(|(\hat{f^*_\varepsilon})^2(t)-(f^*_\varepsilon)^2(t)| > (f^*_\varepsilon)^2(t) - M^{-1/2}) \\
&\leq \mathbb{P}(|(\hat{f^*_\varepsilon})^2(t)-(f^*_\varepsilon)^2(t)| > (f^*_\varepsilon)^2(t)/2) \\
&\leq 2\exp(-M(f^*_\varepsilon)^4(t)/16) \\
&\leq O\left(\left(M^{-1}(f^*_\varepsilon(t))^{-4}\right)^p\right)
\end{aligned}$$

and with the same decomposition as above, this yields

$$\begin{aligned}
\mathbb{E}(|R(t)|^2) &\leq \mathbb{E}\left(\frac{\mathbb{1}_{\widehat{(f_\varepsilon^*)^2}(t)<M^{-1/2}}}{(f_\varepsilon^*)^2(t)}\right) \\
&\quad + \frac{1}{(f_\varepsilon^*)^4(t)}\mathbb{E}\left(\mathbb{1}_{(\hat{f_\varepsilon^*})^2(t)\geq M^{-1/2}}\frac{(\widehat{(f_\varepsilon^*)^2}(t)-(f_\varepsilon^*)^2(t))^2}{(\sqrt{\widehat{(f_\varepsilon^*)^2}(t)}+f_\varepsilon^*(t))^2}\right) \\
&\quad + \frac{1}{(f_\varepsilon^*)^2(t)}\mathbb{E}\left(\mathbb{1}_{(\hat{f_\varepsilon^*})^2(t)\geq M^{-1/2}}\frac{(\widehat{(f_\varepsilon^*)^2}(t)-(f_\varepsilon^*)^2(t))^2}{(\sqrt{\widehat{(f_\varepsilon^*)^2}(t)}+f_\varepsilon^*(t))^2}\left(\frac{1}{\widehat{f_\varepsilon^*)^2}(t)}-\frac{1}{(f_\varepsilon^*)^2(t)}\right)\right) \\
&\leq \frac{1}{(f_\varepsilon^*)^2(t)}\mathbb{P}(\widehat{(f_\varepsilon^*)^2}(t)<M^{-1/2})+\frac{M^{-1}}{(f_\varepsilon^*)^6(t)} \\
&\quad + \frac{M^{-1/2}}{(f_\varepsilon^*)^6(t)}\mathbb{E}[(|\widehat{f_\varepsilon^*)^2}(t)-(f_\varepsilon^*)^2(t)|^3] \\
&\leq \frac{M^{-1}}{(f_\varepsilon^*)^6(t)}+\frac{M^{-1}}{(f_\varepsilon^*)^6(t)}+\frac{M^{1/2}}{(f_\varepsilon^*)^6(t)}M^{-3/2} \\
&\leq c\frac{M^{-1}}{(f_\varepsilon^*)^6(t)}
\end{aligned}$$

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

$$\mathbb{E}(|R(t)|^2) \leq \frac{M^{-1}}{|f_\varepsilon^*(t)|^6}.$$

This ends the proof of the lemma. □

A.2. **Proof of Proposition 1.** Let us study the integrated mean square risk. By writing in the Fourier domain that

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

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

$$\|f - \hat{f}_m\|^2 = \|f - f_m\|^2 + \|f_m - \hat{f}_m\|^2. \tag{15}$$

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

$$\|f_m - \hat{f}_m\|^2 \leq \frac{1}{\pi}\int_{-\pi m}^{\pi m}|\hat{f_Y^*}(u)|^2|R(u)|^2du + \frac{1}{\pi}\int_{-\pi m}^{\pi m}\frac{|\hat{f_Y^*}(u)-f_Y^*(u)|^2}{|f_\varepsilon^*(u)|^2}du. \tag{16}$$

The last term of the right-hand-side of (16) 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 (16):

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

Applying Lemma 1 yields:

$$\begin{aligned}\mathbb{E}(\|f_m - \hat{f}_m\|^2) &\leq \frac{2}{\pi}\int_{-\pi m}^{\pi m} \mathbb{E}(|\hat{f_Y^*}(u) - f_Y^*(u)|^2)\mathbb{E}(|R(u)|^2)du \\ &\quad + \frac{2}{\pi}\int_{-\pi m}^{\pi m} |f^*(u)|^2|f_\varepsilon^*(u)|^2\mathbb{E}(|R(u)|^2)du + 2\frac{\Delta_2(m)}{n} \\ &\leq \frac{2C_1}{\pi}\int_{-\pi m}^{\pi m} n^{-1}\left(\frac{M^{-1/2}}{|f_\varepsilon^*(u)|^4} \wedge \frac{M^{-1}}{|f_\varepsilon^*(u)|^6}\right)du \\ &\quad + \frac{2C_1}{\pi}\int_{-\pi m}^{\pi m} |f^*(u)|^2|f_\varepsilon^*(u)|^2\left(\frac{M^{-1/2}}{|f_\varepsilon^*(u)|^4} \wedge \frac{M^{-1}}{|f_\varepsilon^*(u)|^6}\right)du + 2\frac{\Delta_2(m)}{n} \qquad (17)\end{aligned}$$

By gathering (15) and (17), we obtain the result. □

Acknowledgements

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

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):24–39, 2008a.

C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. II. *Theory of Probability and its Applications*, 52(2):336–349, 2008b.

F. Comte and C. Lacour. Data driven density estimation in presence of unknown convolution operator. *Journal of the Royal Statistical Society B, to appear*, 2011.

F. Comte, Y. Rozenholc, and M. L. Taupin. Penalized contrast estimator for adaptive density deconvolution. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 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(2):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. H. Neumann. Deconvolution from non-standard error densities under replicated measurements. *Statist. Sinica*, 20(4):1609–1636, 2010.

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. *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.

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