



Deconvolution estimation of onset of pregnancy with replicate observations

Fabienne Comte, Adeline Samson, Julien J. Stirnemann

► To cite this version:

Fabienne Comte, Adeline Samson, Julien J. Stirnemann. Deconvolution estimation of onset of pregnancy with replicate observations. *Scandinavian Journal of Statistics*, 2014, 41 (2), pp.325-345. 10.1111/sjos.12029 . hal-00588235v3

HAL Id: hal-00588235

<https://hal.science/hal-00588235v3>

Submitted on 22 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DECONVOLUTION ESTIMATION OF ONSET OF PREGNANCY WITH REPLICATE OBSERVATIONS

FABIENNE COMTE⁽¹⁾, ADELIN 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. 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.

En général, la date précise de début de grossesse est inconnue et est seulement estimée à partir de mesures de l'embryon par ultrasons. Nous voulons estimer la densité de la variable aléatoire correspondant à l'intervalle entre la date des dernières règles et le début réel de la grossesse. Les observations dont nous disposons correspondent aux variables d'intérêt mesurées avec un bruit additif. Nous proposons une procédure d'estimation basée sur des techniques de déconvolution. Cela requiert la connaissance de la loi du bruit mais elle n'est pas disponible. En revanche, nous disposons d'un échantillon d'observations répétées obtenues à partir de grossesses gémellaires. Cela nous permet à la fois d'estimer la loi du bruit et d'améliorer la procédure de déconvolution. La vitesse de convergence de l'estimateur résultant est étudiée et comparée à celles obtenues dans d'autres contextes. Notre estimateur contient un paramètre de troncature pour lequel nous proposons une sélection par validation croisée. Pour finir, nous étudions notre estimateur en simulation puis nous l'appliquons pour estimer la densité visée sur notre échantillon de grossesses spontanées avec estimation du bruit via des données répétées obtenues sur un échantillon de grossesses gémellaires.

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

$$(1) \quad Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n$$

are available. Here, the X_j and the ε_j for $j = 1, \dots, n$ are independent identically distributed and the sequences $(X_j)_{1 \leq j \leq n}$ and $(\varepsilon_j)_{1 \leq j \leq n}$ 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 :

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

with X_{-j} , ε_{-j1} and ε_{-j2} , for $j = 1, \dots, M$, independent and identically distributed. The sequences $(X_{-j})_{1 \leq j \leq M}$, $(\varepsilon_{-j1})_{1 \leq j \leq M}$ and $(\varepsilon_{-j2})_{1 \leq j \leq M}$ 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 a similar 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, 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 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.

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 independent 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.

Assumption(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:

$$\mathbb{E}(e^{it(\varepsilon_{-j1}-\varepsilon_{-j2})}) = |\mathbb{E}(e^{it\varepsilon_{-j1}})|^2 \stackrel{(A1)}{=} \left(\mathbb{E}(e^{it\varepsilon_{-j1}})\right)^2 = (f_\varepsilon^*(t))^2.$$

Therefore, given that $\mathbb{E}(e^{it(\varepsilon_{-j1}-\varepsilon_{-j2})}) = \mathbb{E}(e^{it(Y_{-j1}-Y_{-j2})})$ and under the hypothesis **(A1)**, Delaigle et al. [2008] propose the following estimator of $(f_\varepsilon^*)^2$

$$\left| \frac{1}{M} \sum_{j=1}^M \cos(t(Y_{-j1} - Y_{-j2})) \right|^{1/2}$$

As we will see shortly, our estimator of f_ε^* is used over a domain for which f_ε^* is strictly positive. Therefore we have the following estimation of $(f_\varepsilon^*)^2$ without the absolute value:

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

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

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

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

$$(5) \quad \hat{f}_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixt} \frac{\hat{f}_Y^*(t)}{\tilde{f}_\varepsilon^*(t)} dt, \text{ where } \hat{f}_Y^*(t) = \frac{1}{n+M} \left(\sum_{j=1}^n e^{itY_j} + \sum_{j=1}^M e^{itY_{-j1}} \right).$$

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_{-j2} are not used in $\hat{f}_Y^*(t)$. Indeed, using the $n+2M$ variables increases the variance of the estimator of $f_Y^*(t)$ due to the dependency between Y_{-j1} and Y_{-j2} (see Appendix).

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:

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

where \hat{f}_Y^* is defined as previously.

(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

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

The estimate of f in this case is

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

where \hat{f}_Y^* is defined as previously with $M = 0$. The additional sample with size M can not be used here to improve the estimation of f_Y^* since it concerns the noise only. 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.

The estimator $\hat{f}_m(x)$ also differs from Delaigle et al. [2008] in several ways:

(a) 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.

(b) As already mentioned earlier, we only use the estimation of f_ε^* where f_ε^* 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 $f_m^* = f^* \mathbf{1}_{[-\pi m, \pi m]}(\cdot)$. 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 . Delaigle et al. [2008] study the pointwise risk of their estimator. In this paper, we propose to complete their theoretical study by considering

the integrated risk of our estimator. 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*

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

Proof is given in 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*

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

Proof is given in appendix.

The precise decomposition is

$$(7) \quad \begin{aligned} \mathbb{E}(\|f - \hat{f}_m\|^2) \leq & \|f - f_m\|^2 + C \left(\frac{\Delta_2(m)}{n+M} + \frac{\Delta_4(m)}{(n+M)\sqrt{M}} \wedge \frac{\Delta_6(m)}{(n+M)M} \wedge \frac{\Delta_2(m)}{n+M} \right. \\ & \left. + \frac{\Delta_2^{(f)}(m)}{\sqrt{M}} \wedge \frac{\Delta_4^{(f)}(m)}{M} \right). \end{aligned}$$

This decomposition underlines the different terms involved in the bound of the integrated risk. We recognize in this last inequality (7) the bias $\|f - f_m\|^2$ and variance denoted $\text{Var}(m) := Q_1(m) + Q_1(m) \wedge Q_2(m) + Q_3(m)$ with

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

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]) and for sample size $n + M$. 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

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

Moreover, for $M \geq n$, Comte and Lacour [2011] easily obtain a variance term bounded by $Q_1(m)$. Therefore, compared to the case with observed noise or known density, the term $Q_3(m)$ resulting from the estimation of $(f_\varepsilon^*)^2$ with replicate observations is a substantial step in complexity. Moreover, an interesting question would be to see if $Q_2(m)$ would not be smaller than $Q_1(m)$, at least in an empirical way. 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^* :

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

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

The noise distribution is called ordinary smooth if $\gamma = 0$ and supersmooth otherwise. A Gaussian noise is supersmooth 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 [see Comte et al., 2006, Comte and Lacour, 2011]:

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

Such orders are non standard for variance terms in nonparametric estimation and in particular larger than orders m/n 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) $f \in \mathcal{A}_{a,b,c}(l) = \{f \in \mathbb{L}^1 \cap \mathbb{L}^2, \int |f^*(x)|^2 (x^2 + 1)^a \exp(2b|x|^c) dx \leq l\}$ 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:

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

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

$$Q_3(m) \leq C \left(\frac{m^{2(\alpha-a)_+} e^{(2\beta(\pi m)^\gamma - 2b(\pi m)^c)_+}}{\sqrt{M}} \right) \wedge \left(\frac{m^{2(2\alpha-a)_+} e^{(4\beta(\pi m)^\gamma - 2b(\pi m)^c)_+}}{M} \right).$$

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 (6) 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 $m_{opt} = M^{1/(2a+2\alpha+1)}$,

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

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

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

then

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

The first inequality is a consequence of (6), (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*

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

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

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

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 supersmooth case.

3.3. Discussion about the dominating variance term. We have shown that the optimal rate could be reached both for $M \leq n$ or $M \geq n$, depending on the configuration of the parameters. Now, for cross validation choice of the cutoff parameter, it is useful to determine which term in the variance decomposition is dominating. In particular, we hypothesize that the main variance term defined in (6) may be simply $\text{Var}(m) = CQ_1(m) = C\Delta_2(m)/(n+M)$ where C is a numerical constant.

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

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

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, $\hat{Q}_1(m)$ is larger than 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 $\widehat{\text{Var}}(m) \simeq C\hat{\Delta}_2(m)/(n+M)$ an appropriate choice regardless of the respective values of n and M . The previous considerations will become prominent in the choice of a penalty in Section 4.

4. CROSS-VALIDATION CUT-OFF SELECTOR

The parameter m belongs to the set of admissible values

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

for fixed $\kappa > 0$ and C_0 , and $\lfloor z \rfloor$ denoting the integer part of z . The general outline of the method used to select 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, we have

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

The theoretical optimal choice of m is defined as:

$$m^{th} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \left(\|f - f_m\|^2 + \text{Var}(m) \right) = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \left(-\|f_m\|^2 + \text{Var}(m) \right)$$

The previous value of m^{th} may only be estimated since $\|f - f_m\|$ and $\text{Var}(m)$ are both unknown. Using the estimator \hat{f}_m defined by (5), we will consider the following preliminary

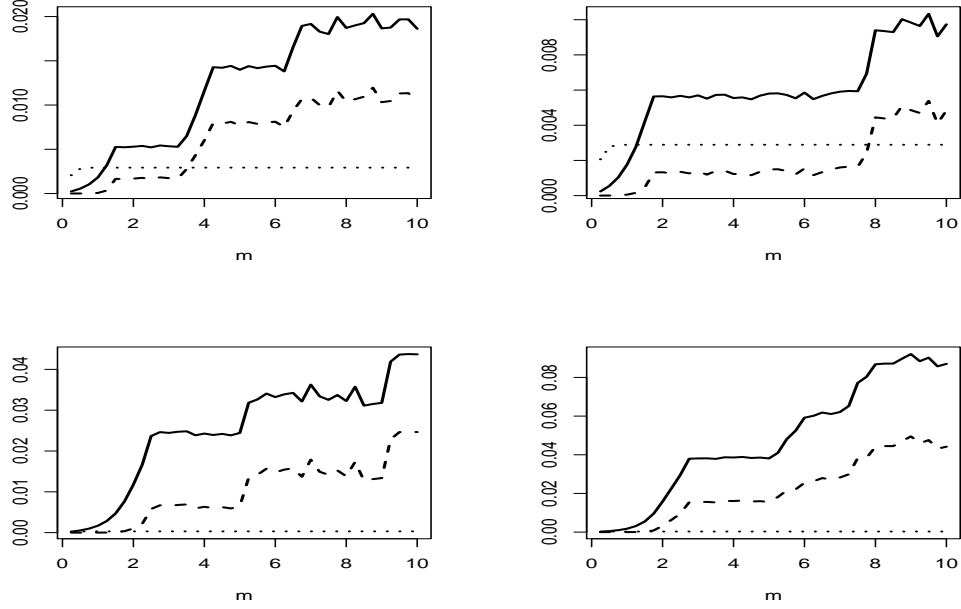


FIGURE 1. Empirical behavior of $\hat{Q}_1(m)$, $\hat{Q}_2(m)$ and $\hat{Q}_3(m)$ as a function of m when $M \ll n$ and $M \gg n$ for f ordinary smooth and supersmooth, chosen as Laplace (left figures) and Gaussian (right figures) respectively. The influence of M and n is illustrated by $M = 200$, $n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

estimate of m :

$$(13) \quad \hat{m} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \left(-\|\hat{f}_m\|^2 + \operatorname{pen}(m) \right)$$

where $\operatorname{pen}(m)$ 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 $\operatorname{Var}(m)$ as defined by (6). When $M \geq n$ and f is sufficiently more regular than f_ε , the optimal rates computations show that $\operatorname{Var}(m) \simeq C\Delta_2(m)/(n+M)$. The empirical behavior of $\hat{Q}_1(m)$, $\hat{Q}_2(m)$ and $\hat{Q}_3(m)$ seems to suggest that $\widehat{\operatorname{Var}}(m) \simeq C\hat{\Delta}_2(m)/(n+M)$ even when $M \ll n$ (see Figure 1). Therefore, based upon Theorem 1 in Comte and Lacour [2011], we define $\operatorname{pen}(m)$ as

$$(14) \quad \operatorname{pen}(m) = K_0 \left(\frac{\log(\Delta_2(m))}{\log(m+1)} \right)^2 \frac{\Delta_2(m)}{n+M}$$

where K_0 is a numerical constant. Note that the term $\log(\Delta_2(m))/\log(m+1)$ 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.

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

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

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 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 $\varphi_{m,j}^*(x) = e^{-ixj/m} \mathbf{1}_{[-\pi m, \pi m]}(x)/\sqrt{m}$. We also recall from Comte and Lacour [2011] the estimated projection coefficients

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

The estimator (5) may then be expressed:

$$(15) \quad \hat{f}_m = \sum_{\ell \in \mathbb{Z}} \hat{a}_{m,\ell} \varphi_{m,\ell}$$

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 is unknown we consider an estimation of this domain,

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

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

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

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

$$(16) \quad \hat{m} = \underset{m \in \{1, \dots, \hat{m}_n\}}{\text{argmin}} \left(-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m) \right)$$

Finally, by plugging (16) in (15) we obtain $\hat{f}_{\hat{m}}$ 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. 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.

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

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).

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. 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].

Table 2 presents the comparison of the penalized estimator and the estimator given by Delaigle et al. [2008] for the Gaussian mixture density (v). The integrated squared

TABLE 1. Results of simulations presented as $\text{MISE} \times 100$. In each case the MISE was averaged over 100 estimations. The case “ f_ε known” corresponds to the estimator denoted \hat{f}_m^{KN} with 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
	$M = \sqrt{n}$	0.401	0.118	0.562	0.236
	$M = n$	0.241	0.102	0.411	0.158
f Cauchy	f_ε known	0.371	0.044	0.728	0.624
	$M = \sqrt{n}$	0.422	0.161	0.753	0.234
	$M = n$	0.287	0.152	0.461	0.177
f Laplace	f_ε known	2.066	0.588	3.506	1.815
	$M = \sqrt{n}$	2.348	2.211	3.814	2.574
	$M = n$	2.288	2.195	2.636	2.223
f Gaussian	f_ε known	0.191	0.041	0.355	0.846
	$M = \sqrt{n}$	0.196	0.050	0.686	0.087
	$M = n$	0.133	0.037	0.212	0.051
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
	$M = \sqrt{n}$	0.441	0.133	0.598	0.186
	$M = n$	0.250	0.107	0.357	0.117
f Cauchy	f_ε known	0.339	0.167	0.420	0.149
	$M = \sqrt{n}$	0.526	0.171	0.652	0.200
	$M = n$	0.310	0.157	0.378	0.168
f Laplace	f_ε known	1.110	0.367	1.967	0.664
	$M = \sqrt{n}$	2.384	2.264	3.348	2.267
	$M = n$	2.312	2.200	2.330	2.208
f Gaussian	f_ε known	0.511	0.219	0.720	0.386
	$M = \sqrt{n}$	0.330	0.045	0.550	0.102
	$M = n$	0.126	0.037	0.226	0.036

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 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

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 $\times 100$ (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)	0.33 (0.13)	0.25 (0.06)
ε Gaussian	2.09 (1.33)	1.42 (0.92)	0.44 (0.87)	0.27 (0.06)

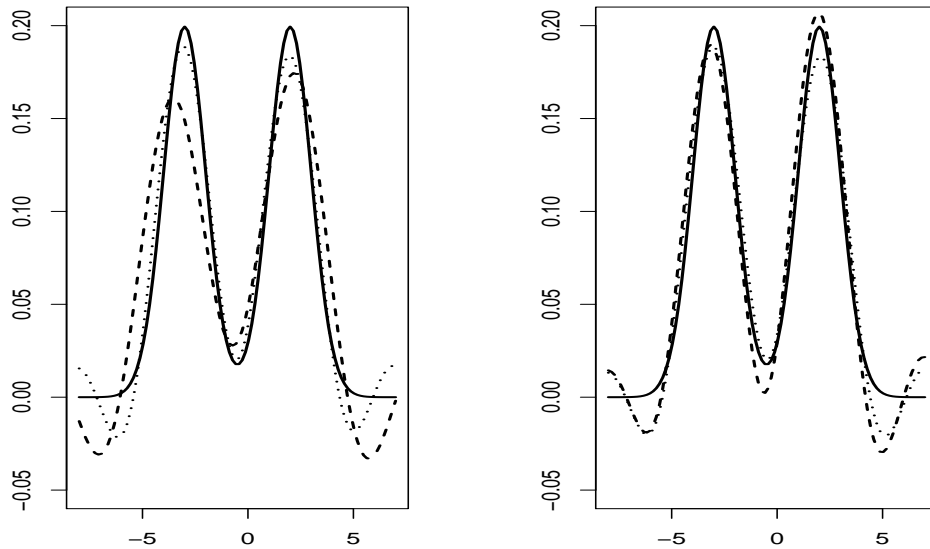


FIGURE 2. Estimations for $n = M = 200$ (dashed line) and $n = M = 500$ (dotted line) for the Gaussian mixture density (plain line) (v). (left): Laplace noise; (right): Gaussian noise. Two independent samples were used, of size n and M respectively.

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

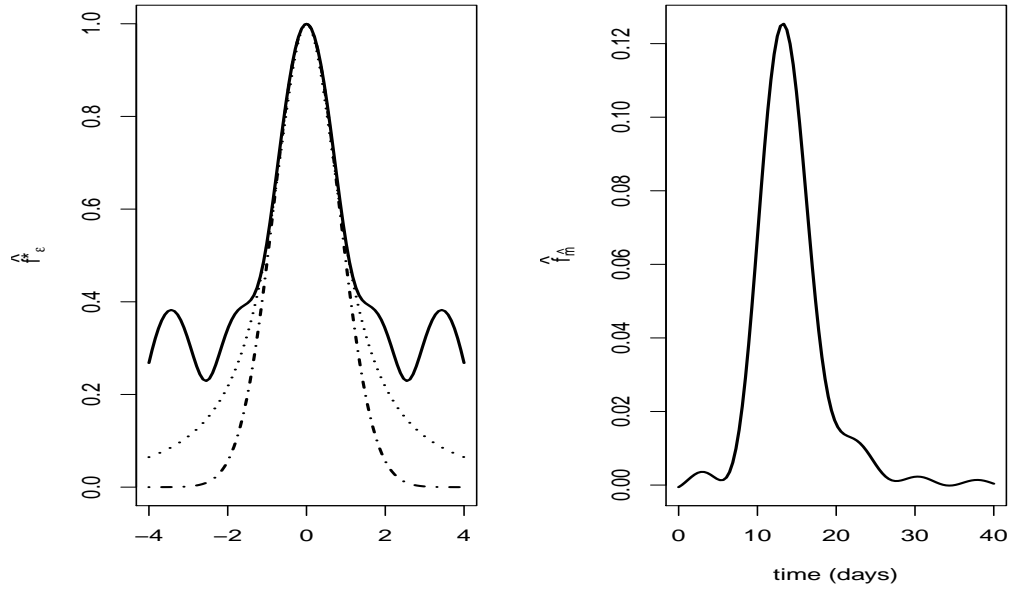


FIGURE 3. (left): estimation of the characteristic function of the noise in twin pregnancies (plain line). Laplace (dashed line) and Gaussian (dotted line) characteristic functions are plotted for comparison. (right): estimation of the density of onset of pregnancy.

Figure 3 (b) presents the penalized estimator $\hat{f}_{\hat{m}}(\cdot)$. 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 Laplace and Gaussian characteristic functions). Under this simulation model, the risk $\text{MISE} \times 100$ 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 module. 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. 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.

APPENDIX: PROOFS

.1. Proof of Remark 1. Let us denote the estimator of f_Y^* based on the samples (Y_j) , (Y_{-j1}) and (Y_{-j2}) :

$$\tilde{f}_Y^*(t) = \frac{1}{n + 2M} \left(\sum_{j=1}^n e^{itY_j} + \sum_{j=1}^M e^{itY_{-j1}} + \sum_{j=1}^M e^{itY_{-j2}} \right)$$

Due to the dependency between (Y_{-j1}) and (Y_{-j2}) , we have

$$\begin{aligned} \text{Var } \tilde{f}_Y^*(t) &= \frac{1}{(n+2M)^2} \left(\text{Var} \left(\sum_{j=1}^n e^{itY_j} \right) + \text{Var} \left(\sum_{j=1}^M e^{itY_{-j1}} \right) + \text{Var} \left(\sum_{j=1}^M e^{itY_{-j2}} \right) \right. \\ &\quad \left. + 2\text{cov} \left(\sum_{j=1}^M e^{itY_{-j1}}, \sum_{j=1}^M e^{itY_{-j2}} \right) \right) \\ &= \frac{n+4M}{(n+2M)^2} |f_\varepsilon^*(t)|^2 (1 - |f_X^*(t)|^2) \end{aligned}$$

The variance of $\hat{f}_Y^*(t)$ is equal to

$$\frac{1}{n+M} |f_\varepsilon^*(t)|^2 (1 - |f_X^*(t)|^2).$$

Easy calculations provide $(n+4M)/(n+2M)^2 \geq 1/(n+M)$, which end the proof of Remark 1.

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

Firt, write the decomposition

$$\begin{aligned} \mathbb{E}(|R(t)|^2) &= \mathbb{E} \left(\frac{\mathbf{1}_{(\widehat{f_\varepsilon^*}^2(t) < M^{-1/2})}}{(f_\varepsilon^*)^2(t)} \right) \\ &\quad + \frac{1}{(f_\varepsilon^*)^2(t)} \mathbb{E} \left(\mathbf{1}_{(\widehat{f_\varepsilon^*}^2(t) \geq M^{-1/2})} \frac{((\widehat{f_\varepsilon^*}^2(t) - (f_\varepsilon^*)^2(t))^2}{(\widehat{f_\varepsilon^*}^2(t)(\sqrt{(\widehat{f_\varepsilon^*}^2(t) + f_\varepsilon^*(t))^2})} \right) \end{aligned}$$

On the one hand, we get

$$\begin{aligned} \mathbb{E}(|R(t)|^2) &\leq \frac{1}{(f_\varepsilon^*)^2(t)} + \frac{M}{(f_\varepsilon^*)^2(t)} \mathbb{E}[(\widehat{f_\varepsilon^*}^2(t) - (f_\varepsilon^*)^2(t))^2] \\ &\leq \frac{2}{(f_\varepsilon^*)^2(t)} \end{aligned}$$

On the other hand, let us distinguish two cases.

(i) If $(f_\varepsilon^*)^2(t) \leq 2M^{-1/2}$, we have both $1/(f_\varepsilon^*)^2(t) \leq 2M^{-1/2}/(f_\varepsilon^*)^4(t)$ and $M^{-1/2}/(f_\varepsilon^*)^4(t) \leq 2M^{-1}/(f_\varepsilon^*)^6(t)$ and

$$\begin{aligned} \mathbb{E}(|R(t)|^2) &\leq \frac{1}{(f_\varepsilon^*)^2(t)} + \frac{M^{1/2}}{(f_\varepsilon^*)^4(t)} \mathbb{E}[(\widehat{f_\varepsilon^*}^2(t) - (f_\varepsilon^*)^2(t))^2] \\ &\leq \frac{1}{(f_\varepsilon^*)^2(t)} + \frac{M^{-1/2}}{(f_\varepsilon^*)^4(t)} \leq 3 \frac{M^{-1/2}}{(f_\varepsilon^*)^4(t)} = 3 \left(\frac{M^{-1/2}}{(f_\varepsilon^*)^4(t)} \right) \wedge \left(\frac{M^{-1}}{(f_\varepsilon^*)^6(t)} \right) \end{aligned}$$

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

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

and with the same decomposition as above, this yields

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

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. \square

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

$$(17) \quad \|f - \hat{f}_m\|^2 = \|f - f_m\|^2 + \|f_m - \hat{f}_m\|^2.$$

Moreover, by applying the Parseval formula, we obtain

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

It follows that

$$(18) \quad \|f_m - \hat{f}_m\|^2 \leq \frac{1}{\pi} \int_{-\pi m}^{\pi m} |\hat{f}_Y^*(u)|^2 |R(u)|^2 du + \frac{1}{\pi} \int_{-\pi m}^{\pi m} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|f_\varepsilon^*(u)|^2} du.$$

The last term of the right-hand-side of (18) 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 (18):

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

Applying Lemma 1 yields:

$$\begin{aligned} \mathbb{E}(\|f_m - \hat{f}_m\|^2) &\leq \frac{2}{\pi} \int_{-\pi m}^{\pi m} (\mathbb{E}(|\hat{f}_Y^*(u) - f_Y^*(u)|^4) \mathbb{E}(|R(u)|^4))^{1/2} du \\ &\quad + \frac{2}{\pi} \int_{-\pi m}^{\pi m} |f_Y^*(u)|^2 |f_\varepsilon^*(u)|^2 \mathbb{E}(|R(u)|^2) du + 2 \frac{\Delta_2(m)}{n+M} \\ &\leq \frac{2}{\pi} \int_{-\pi m}^{\pi m} (n+M)^{-1} \frac{1}{|f_\varepsilon^*(u)|^2} du \\ (19) \quad &\quad + \frac{2C_1}{\pi} \int_{-\pi m}^{\pi m} |f_Y^*(u)|^2 |f_\varepsilon^*(u)|^2 \left(\frac{M^{-1/2}}{|f_\varepsilon^*(u)|^4} \wedge \frac{M^{-1}}{|f_\varepsilon^*(u)|^6} \right) du + 2 \frac{\Delta_2(m)}{n+M} \end{aligned}$$

By gathering (17) and (19), we obtain the result. \square

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, September 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, January 1993.
- J Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272, September 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, October 2009.
- T Li and Q Vuong. Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65:139165, 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, 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, October 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, November 2000.

MAP5, UMR CNRS 8145, UNIVERSITÉ PARIS DESCARTES, 45 RUE DES SAINTS PÈRES, 75006 PARIS, FRANCE