HAL
archives-ouvertes.fr

Hazard estimation for censored data contaminated with additive measurement error: application to length of pregnancy

Fabienne Comte, Adeline Samson, Julien Stirnemann

▶ **To cite this version:**

Fabienne Comte, Adeline Samson, Julien Stirnemann. Hazard estimation for censored data contaminated with additive measurement error: application to length of pregnancy. MAP5 2015-18. 2015. <hal-01150296v2>

HAL Id: hal-01150296
https://hal.archives-ouvertes.fr/hal-01150296v2

Submitted on 4 Dec 2015

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.

HAZARD ESTIMATION FOR CENSORED DATA CONTAMINATED WITH ADDITIVE MEASUREMENT ERROR: APPLICATION TO LENGTH OF PREGNANCY

FABIENNE COMTE[(1)], ADELINE SAMSON[(1,2)], AND JULIEN J STIRNEMANN[(1,3)]

[(1)] *MAP5, UMR CNRS 8145, Université Paris Descartes*
[(2)] *Laboratoire Jean Kuntzmann, UMR CNRS 5224, Univ Grenoble-Alpes, France*
[(3)] *Obstetrics and Maternal - Fetal Medicine, GHU Necker-Enfants Malades, Université Paris Descartes.*

ABSTRACT. We consider random variables which can be subject to both censoring and measurement errors. We focus on the case where the measurement errors affect both the variable of interest and the censoring variable, which is the case of the timing of spontaneous delivery among pregnant women. We propose an estimation strategy to estimate the hazard rate of the underlying variable of interest. We explain the model and this strategy and provide L^2-risk bound for the data driven resulting estimator. Simulations illustrate the performances of the estimator. Lastly, the method is applied to a real data set of length of pregnancy.

KEYWORDS. Censored data; Measurement error; Survival function estimation; Hazard rate function estimation; Nonparametric methods; Deconvolution;

1. INTRODUCTION

The length of pregnancy is defined as the time between the spontaneous fertilization of the oocyte and spontaneous delivery following labor. However, despite the simplicity of the definition, the estimation of the physiological length of pregnancy is a challenging problem since both the time origin and the time of onset of labor are only partly observed. Several events may occur during pregnancy and interfere with its physiological course prior to spontaneous labor, such as medical induction of labor or cesarean section, intrauterine fetal death or any fetal or maternal medical condition that leads to delivery before its "physiological" time.

Using time-to-event terminology, the time to spontaneous labor is therefore randomly right-censored. Right-censored data, in its standard presentation, involves independent observations $((X_j \wedge C_j), \mathbf{1}_{X_j \leq C_j})$ for $j = 1, \ldots, n$ where the variable X denotes the true time between the origin and the occurrence of the event of interest and the variable C denotes the true time between the origin and the occurrence of censoring. We classically assume that X and C are independent.

As stated above, the time of fertilization (the time origin) is ill-observed and can never be directly measured, except in the very special case of in vitro fertilization, which we do not consider here, as this case cannot be considered as physiological. In practice, in spontaneous pregnancies, the time of onset of pregnancy is predicted either from the last menstrual period, by simply adding 14 days or, better, by the sonographic measurement of the length of the embryo in the first trimester.

Because it is not directly observed and estimated from biological data, the time of fertilization is known up to an error, the magnitude of which is of several days, regardless of the method.

In this paper, we consider the problem of censored data prone to measurement error. As the time origin of pregnancy is known up to an additive (random) error, both variables X (time between the true onset and the natural childbirth) and C (time between the true onset and any censoring event) are observed up to this additive error.

Let us define the model more precisely. Let ε denote the random error variable assumed to be independent of X and C. We assume that the observations are properly classified as censored or uncensored and that both the censored and uncensored observations are measured with error:

$$\begin{aligned} Y_j &= (X_j \wedge C_j) + \varepsilon_j = (X_j + \varepsilon_j) \wedge (C_j + \varepsilon_j), \quad j = 1, \dots, n \\ \delta_j &= \mathbf{1}_{X_j \leq C_j}, \end{aligned} \tag{1}$$

Note that the censoring indicator δ_j is unchanged by the measurement error: $\mathbf{1}_{X_j \leq C_j} = \mathbf{1}_{X_j + \varepsilon_j \leq C_j + \varepsilon_j}$.

The purpose of this work is to provide a nonparametric estimator of the hazard function h_X of X, based on the observations (Y_j, δ_j).

Nonparametric methods have been proposed in the frameworks of both censored data and deconvolution. Regarding censoring, Antoniadis et al. [1999] consider a wavelet hazard estimator which is not adaptive, whereas Li [2007, 2008] suggests estimators based on wavelet with hard or block thresholding. Estimators based on model selection via penalization have also been proposed: Dohler and Ruschendorf [2002] estimate the log-hazard function using a penalized likelihood-based criterion, Brunel and Comte [2005, 2008] consider penalized contrast estimators for both the density and the hazard rate using either the Nelson-Aalen estimator of the cumulative hazard function or the Kaplan-Meier cumulative hazard estimator, Reynaud-Bouret [2006] proposes a penalized projection estimator of the Aalen multiplicative intensity process with adaptive results and minimax rates and Akakpo and Durot [2010] consider a histogram selection for both density and hazard rate estimation.

We can also consider our estimation problem in the setting of deconvolution. Deconvolution has been widely studied in various contexts. We hereby restrain to references with a known density of the noise. Kernel estimators have been proposed by Stefanski and Carroll [1990], Fan [1991], with bandwidth selection strategies [Delaigle and Gijbels, 2004]. Wavelet estimators [Pensky and Vidakovic, 1999, Fan and Koo, 2002], and projection methods with model selection [Comte et al., 2006] have also been developed. A pointwise estimation method for S_X has been proposed by Dattner et al. [2011] when the data are noisy but not censored. In this work, we estimate the hazard rate h_X based on the ratio of deconvolution estimators, using the developments of Dattner et al. [2011] in the setting of distribution functions (with no censoring) for the denominator.

The paper is organized as follows: Section 2 studies a quotient estimator of the hazard rate h_X. The estimator is illustrated with a simulation study in Section 3 and is compared to results obtained when neither measurement error nor censoring are considered. The motivating application of estimation of length of pregnancy is illustrated by an analysis of real data in Section 4. In Section 5, we discuss the alternative problem where the noise affects only the variable X, a problem that remains open. Proofs are gathered in Appendix.

Notations. We denote f_U the density of a variable U. We denote $S_U(t) = \mathbb{P}(U \geq t)$ the survival function at point t of a random variable U, $h_U(t) = f_U(t)/S_U(t)$ the hazard ratio at point t

and f_U^* the characteristic function. We denote $g^*(t) = \int e^{itx} g(x)dx$ the Fourier transform of any integrable function g. For a function $g : \mathbb{R} \mapsto \mathbb{R}$, we denote $\|g\|^2 = \int_{\mathbb{R}} g^2(x)dx$ the $\mathbb{L}^2$ norm. For two integrable and square-integrable functions g and h, we denote $g \star h$ the convolution product $g \star h(x) = \int g(x-u)h(u)du$. For two real numbers a and b, we denote $a \wedge b = \min(a, b)$.

2. Censored data and measurement noise

2.1. **Setting.** We observe for $j = 1, \dots, n$

$$Y_j = (X_j \wedge C_j) + \varepsilon_j,\ \delta_j = \mathbf{1}_{X_j \leq C_j}.$$

We assume that the law of the noise is known and that its characteristic function is such that

$$\forall u \in \mathbb{R},\ \ f_\varepsilon^*(u) \neq 0.$$

The following assumption, which is verified by exponential or Gamma distributions for example, will be considered fulfilled throughout this section:

Assumption (A1) We assume both X and C to be nonnegative random variables. We also assume $\mathbb{E}(X) < +\infty$ and $\mathbb{E}(C) < +\infty$.

In this section, we want to estimate the hazard rate h_X of X. This hazard rate may be expressed as the following nonstandard quotient, where S_X is the survival function:

$$h_X = \frac{f_X}{S_X} = \frac{f_X S_C}{S_X S_C} = \frac{f_X S_C}{S_{X \wedge C}}.$$

The idea is to estimate separately the numerator $f_X S_C$ and the denominator $S_{X \wedge C}$.

2.2. **Construction of the estimator for the numerator.** It is rather easy to get an estimator of the numerator $f_X S_C$, and more precisely of its projection on the space

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

For a square-integrable function g, let us denote g_m its orthogonal projection on S_m, such that $g_m^*(x) = g^*(x)\mathbf{1}_{|x| \leq \pi m}$. Then $(f_X S_C)_m$ is estimated by the following deconvolution estimator:

$$\widehat{(f_X S_C)}_m(x) = \frac{1}{2\pi n} \sum_{j=1}^{n} \int_{-\pi m}^{\pi m} \frac{e^{-iux}\delta_j e^{iuY_j}}{f_\varepsilon^*(u)} du. \tag{3}$$

Indeed

$$\begin{aligned} \mathbb{E}(\delta_1 e^{iuY_1}) &= \mathbb{E}(\mathbf{1}_{X_1 \leq C_1} e^{iu(X_1 \wedge C_1)} e^{iu\varepsilon_1}) = \mathbb{E}(\mathbf{1}_{X_1 \leq C_1} e^{iuX_1}) f_\varepsilon^*(u) \\ &= \mathbb{E}(S_C(X_1) e^{iuX_1}) f_\varepsilon^*(u) = (f_X S_C)^*(u) f_\varepsilon^*(u). \end{aligned}$$

Therefore

$$\mathbb{E}(\widehat{(f_X S_C)}_m(x)) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} (f_X S_C)^*(u) du := (f_X S_C)_m(x).$$

Under integrability conditions, $(f_X S_C)_m(x)$ tends to $f_X S_C(x)$ when m tends to infinity by the Fourier inverse formula. Then the risk bound of $\widehat{(f_X S_C)}_m$ can easily be deduced from Comte *et al.* (2006):

$$\mathbb{E}(\|\widehat{(f_X S_C)}_m - (f_X S_C)\|^2) \leq \|f_X S_C - (f_X S_C)_m\|^2 + \frac{\mathbb{E}(\delta_1)}{2\pi} \int_{-\pi m}^{\pi m} \frac{du}{|f_\varepsilon^*(u)|^2}$$

where the bias term

$$\|f_X S_C - (f_X S_C)_m\|^2 = \frac{1}{2\pi} \int_{|u| \geq \pi m} |(f_X S_C)^*(u)|^2 du$$

is decreasing with m while the variance term, $(2\pi)^{-1}\mathbb{E}(\delta_1) \int_{-\pi m}^{\pi m} du/|f_\varepsilon^*(u)|^2$, obviously increases. As a consequence, we wish to choose m so that a good compromise is reached. This is done by estimating each term. The variance is estimated by a quantity $\mathrm{pen}_1(m)$ proportional to the bound on the variance. Since $\|f_X S_C - (f_X S_C)_m\|^2 = \|f_X S_C\|^2 - \|(f_X S_C)_m\|^2$, the bias term can be replaced by $-\|(f_X S_C)_m\|^2$ in the search for an optimal m, as it amounts to omit a term which does not depend on m. Therefore, this term is estimated by $-\|\widehat{(f_X S_C)}_m\|^2$. This explains why we select $\hat{m}_1$ as follows.

$$\hat{m}_1 = \arg \min_{m \in \{1,\dots,m_{n,1}\}} (-\|\widehat{(f_X S_C)}_m\|^2 + \mathrm{pen}_1(m)), \tag{4}$$

where $m_{n,1}$ is such that $m_{n,1} \leq n$ and

$$\mathrm{pen}_1(m) = \frac{\kappa_1}{n} \left(\frac{1}{n} \sum_{k=1}^{n} \delta_k \right) \log(J_1(m)) J_1(m), \text{ with } J_1(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{du}{|f_\varepsilon^*(u)|^2}. \tag{5}$$

In $\mathrm{pen}_1(m)$, the constant κ_1 is calibrated from preliminary simulations.

Following Comte et al. [2006], applying Talagrand's Inequality,

$$\mathbb{E}(\|\widehat{(f_X S_C)}_{\hat{m}_1} - f_X S_C\|^2) \leq C \inf_{m \in \{1,\dots,m_{n,1}\}} \left(\|(f_X S_C)_m - f_X S_C\|^2 + \mathrm{pen}_1(m) \right) + \frac{C'}{n}$$

for C and C' two constants which do not depend on n. In other words, the estimator $\widehat{(f_X S_C)}_{\hat{m}_1}$ realizes the adequate squared-bias/variance trade off, up to the multiplicative factor C and the negligible residual term C'/n.

2.3. **Construction of the estimator for the denominator.** We now wish to estimate the denominator $S_{X \wedge C} = S_X S_C$. Note that under assumption **(A1)**, the survival functions are square-integrable over $\mathbb{R}^+$ (thus over $\mathbb{R}$ if they are extended by 0), as opposed to cumulative distribution functions. This is true, for example, for exponential distributions, classically used in survival analysis: the associated survival functions are clearly square integrable.

We define for $x \geq 0$, the following estimator of $S_{X \wedge C}$, as proposed by Dattner et al. [2011]:

$$\widehat{(S_{X \wedge C})}_m(x) = \frac{1}{2} + \frac{1}{\pi n} \sum_{j=1}^{n} \mathrm{Re} \int_0^{\pi m} \frac{1}{iu} \left(\frac{e^{iu(Y_j - x)}}{f_\varepsilon^*(u)} \right) du. \tag{6}$$

While only the pointwise risk of this estimator is studied in Dattner et al. [2011], we want hereby to compute the integrated $\mathbb{L}^2$-risk of $\widehat{(S_{X \wedge C})}_m$. It is not trivial from (6) why this integrated risk is

properly defined. Therefore, before proceeding to the study of the integrated risk we consider an alternate expression of $(\widehat{S_{X\wedge C}})_m$ using $(1/\pi)\int_0^{+\infty}\sin(v)/v dv = 1/2$, as follows:

$$(\widehat{S_{X\wedge C}})_m(x) = \mathrm{Re}\left(\frac{1}{2\pi n}\sum_{j=1}^{n}\int_{-\pi m}^{\pi m}\frac{e^{-iux}}{iu}\left(\frac{e^{iuY_j}}{f_\varepsilon^*(u)}-1\right)du\right)+\psi_m(x) \tag{7}$$

with

$$\psi_m(x) = -\frac{1}{2i\pi}\int_{|u|\geq\pi m}\frac{e^{-iux}}{u}du = \frac{1}{\pi}\int_{\pi m}^{+\infty}\frac{\sin(ux)}{u}du.$$

Note that $\psi_m^*(u) = -1/(iu)\mathbf{1}_{|u|\geq\pi m}$. This implies by Parseval formula that $\int_0^{+\infty}|\psi_m(x)|^2dx = (1/2\pi^2)m^{-1}$. Moreover, the real part $\mathrm{Re}(\cdot)$ in formula (7) is not mandatory because the integral is real, given the symmetry of the domain and the properties of the function under integration.

Then, in order to compute the integrated $\mathbb{L}^2$-risk of $(\widehat{S_{X\wedge C}})_m$, we can see (7) as a deconvolution estimator of $S^*_{X\wedge C}$. First, notice that $S^*_{X\wedge C}(u) = \int_0^{+\infty}e^{iuv}S_{X\wedge C}(v)dv$ is well defined under assumption **(A1)** because $S_{X\wedge C}$ is integrable and square integrable on $\mathbb{R}^+$, its support. Then, let us introduce the following estimate of $S^*_{X\wedge C}(u)$: for all u,

$$\hat{S}^*_{X\wedge C}(u) = \frac{1}{n}\frac{1}{iu}\sum_{j=1}^{n}\left(\frac{e^{iuY_j}}{f_\varepsilon^*(u)}-1\right). \tag{8}$$

Lemma 1. *The estimator $\hat{S}^*_{X\wedge C}$ given by (8) is well defined on $\mathbb{R}$ and is an unbiased estimate of $S^*_{X\wedge C}(u)$.*

The estimator $(\widehat{S_{X\wedge C}})_m$ written as (7) can be seen as the Fourier inversion of (8). Here, however, the Fourier inversion is done with a cutoff πm on the first part of the estimator, which is not integrable, and on the whole real line for the non random part which has a known value. This allows us to write

$$\begin{aligned}(\widehat{S_{X\wedge C}})_m(x) &= \mathrm{Re}\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m}e^{-iux}\hat{S}^*_{X\wedge C}(u)du\right)+\psi_m(x)\\ &= \frac{1}{2\pi}\int_{-\pi m}^{\pi m}e^{-iux}\hat{S}^*_{X\wedge C}(u)du+\psi_m(x).\end{aligned}$$

We emphasize that, in practice, the $\psi_m(x)$ term is a very useful correction of the estimator for $x\in[0,1]$. We are now able to study the integrated $\mathbb{L}^2$-risk and prove the following result.

Proposition 1. *Let $(\widehat{S_{X\wedge C}})_m$ be defined by (6). Under assumption* **(A1)**, *we have*

$$\mathbb{E}(\|(\widehat{S_{X\wedge C}})_m - S_{X\wedge C}\|^2) \leq \frac{1}{\pi}\int_{|u|\geq\pi m}|S^*_{X\wedge C}(u)|^2du + \frac{2}{\pi^2 m}+\frac{4}{\pi n}\int_1^{\pi m}\frac{du}{u^2|f_\varepsilon^*(u)|^2}+\frac{c}{n}$$

where c is a positive constant, $c = (\mathbb{E}(Y_1^2)/\pi)\int_0^1 du/|f_\varepsilon^*(u)|^2$.

The first two terms in the upper bound are squared bias terms decreasing when m increases, the third is a variance term which increases with m; the last term is a negligible residual. Contrary to

the numerator estimator, the bias decreases at a slower rate. This is due to the term $1/(\pi^2 m)$ and to

$$S^*_{X\wedge C}(u) = \frac{f^*_{X\wedge C}(u) - 1}{iu} = \frac{f^*_{X\wedge C}(u)}{iu} - \frac{1}{iu}$$

which implies

$$\frac{1}{2\pi}\int_{|u|\geq \pi m} |S^*_{X\wedge C}(u)|^2 du = O\left(\frac{1}{m}\right).$$

This slow bias order is due to the discontinuity in 0 of survival functions for positive random variables, while computing a global risk over $\mathbb{R}^+$. Even with a slow bias decrease rate, we could still obtain a satisfactory convergence rate for the estimator. Indeed, the noises we have in mind must also have lower bounded supports. For example, an exponential distribution for ε yields a variance term of order m/n. The optimal value m_{opt} for the cutoff given by the bias-variance compromise is such that $m_{opt} = O(\sqrt{n})$ and the resulting rate is $O(n^{-1/2})$, which is good for a nonparametric deconvolution problem.

All these considerations being asymptotic, we propose a finite sample model selection strategy for choosing m. Let us denote by

$$J_2(m) := \frac{1}{\pi}\int_1^{\pi m} \frac{du}{u^2|f^*_\varepsilon(u)|^2}, \text{ and } \quad \mathrm{pen}_2(m) = \kappa_2 \log(n)\frac{J_2(m)}{n}, \tag{9}$$

where κ_2 is a constant to be calibrated by simulations. Note that the lower bound of the integral is 1 so that the integral is properly defined. Then, setting

$$\hat{m}_2 = \arg\min_{m\in\{1,\dots,m_{n,2}\}} (-\|(\widehat{S_{X\wedge C}})_m\|^2 + \frac{3}{2\pi^2 m} + \mathrm{pen}_2(m)), \tag{10}$$

for $m_{n,2}$ such that $m_{n,2} \leq n$ and $J_2(m_{n,2}) \leq n$, we obtain an adaptive estimator of $S_{X\wedge C}$, which is rather simple to implement, compared to the pointwise procedure of [Dattner et al., 2011].

We can prove

Theorem 1. *Let $(\widehat{S_{X\wedge C}})_m$ be defined by (6) and $\hat{m}_2$ by (10). Then there exists a numerical constant κ_0, such that for $\kappa_2 \geq \kappa_0$, we have*

$$\mathbb{E}(\|(\widehat{S_{X\wedge C}})_{\hat{m}_2} - S_{X\wedge C}\|^2) \quad \leq \quad \inf_{m\in\{1,\dots,m_{n,2}\}} \left(\frac{3}{\pi}\int_{|u|\geq\pi m} |S^*_{X\wedge C}(u)|^2 du + \frac{2}{\pi^2 m} + 4\mathrm{pen}_2(m)\right) + \frac{C}{n}$$

where C is a constant depending on f^*_ε.

From the proof we find that $\kappa_0 = 48$ suits, but this theoretical value is too large in practice (see Section 3).

Our adaptive procedure $(\widehat{S_{X\wedge C}})_{\hat{m}_2}$ has the simplicity of choosing a unique global cutoff $\hat{m}_2$ for m. This is an advantage compared to the pointwise selection procedure described in [Dattner et al., 2011], where the selection is repeated for each point. As a counterpart, we obtain a theoretical global rate which is not as good as the pointwise one. This is due to the fact the pointwise strategy avoids the point $x = 0$ where a discontinuity occurs.

2.4. **Construction of the estimator of h_X.** The two estimators $(\widehat{f_X S_C})_{\hat{m}_1}$ and $(\widehat{S_{X\wedge C}})_{\hat{m}_2}$ allow us to build the final estimator of the hazard rate h_X as a quotient estimator. To prevent the denominator to get small, a truncation is required when computing the quotient. The estimator of $h_X(x)$ is finally

$$\hat{h}_{\hat{m}_1,\hat{m}_2}(x) = \frac{(\widehat{f_X S_C})_{\hat{m}_1}(x)}{(\widehat{S_{X\wedge C}})_{\hat{m}_2}(x)} \mathbf{1}_{(\widehat{S_{X\wedge C}})_{\hat{m}_2}(x) \geq \lambda/\sqrt{n}} \tag{11}$$

where λ is a constant to be calibrated. Note that, heuristically, the resulting risk of $\hat{h}_{\hat{m}_1,\hat{m}_2}$ is the addition of the risks of the numerator and the denominator, up to a multiplicative constant.

3. Simulation

3.1. **Design of simulation.** Simulations are used to evaluate the performances of the estimator. For each design of simulations, 500 datasets are simulated. We consider samples of size $n = 400, 1000$. Data are simulated with a Laplace noise with variance $\sigma^2 = 2b^2$ as follows:

$$f_\varepsilon(x) = \frac{1}{2b} e^{-|x|/b} \text{ and } f^*_\varepsilon(x) = \frac{1}{1 + b^2 x^2}$$

with $b = 1/(2\sqrt{5})$ or $b = 1/(\sqrt{5})$. We consider four densities for X.

(1) Mixed Gamma distribution: $X = 1/\sqrt{5.48}W$ with $W \sim 0.4\Gamma(5,1) + 0.6\Gamma(13,1)$
(2) Beta distribution: $X \sim \mathcal{B}(2,5)/\sqrt{0.025}$
(3) Gaussian distribution: $X \sim |\mathcal{N}(5,1)|$
(4) Gamma distribution: $X \sim \Gamma(5,1)/\sqrt{5}$

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 $s2n = 2.5$ and $s2n = 10$ in our simulations.

The censoring variable C is simulated with an exponential distribution, with parameter chosen to ensure 20% or 40% of censored observations.

3.2. **Estimator implementation.** We first describe the implementation of the numerator $(\widehat{f_X S_C})_{\hat{m}_1}$. The penalty depends on $J_1(m)$, which is computed by discretization of the integral. Then we compute $\text{pen}_1(m)$ defined by (5) with the choice $\kappa_1 = 2$, obtained after a set of simulation experiments to calibrate it. We consider $m_{n,1} = \text{argmax}(m \in \mathbb{N}, J_1(m)/n \leq 1)$. Following, we have the final estimation of $\hat{m}_1$ defined by (4). By plugging (4) into (3) we obtain $(\widehat{f_X S_C})_{\hat{m}_1}$ which is our estimator for the numerator.

For the implementation of the denominator $(\widehat{S_{X\wedge C}})_{\hat{m}_2}$, the penalty depends on $J_2(m)$, which is computed by discretization of the integral. We take $\text{pen}_2(m)$ as defined by (9) with $\kappa_2 = 5$, after a set of simulation experiments to calibrate it. We define $m_{n,2} = \text{argmax}(m \in \mathbb{N}, J_2(m)/n \leq 1)$. Following, we have the final estimation of $\hat{m}_2$ defined by (10). By plugging this in (6) we obtain our estimator for the denominator $(\widehat{S_{X\wedge C}})_{\hat{m}_2}$.

Finally, we estimate h_X as a quotient:

$$\hat{h}_{\hat{m}_1,\hat{m}_2}(x) = \frac{(\widehat{f_X S_C})_{\hat{m}_1}(x)}{(\widehat{S_{X\wedge C}})_{\hat{m}_2}(x)} \mathbf{1}_{(\widehat{S_{X\wedge C}})_{\hat{m}_2}(x)\geq \lambda/\sqrt{n}},$$

with the numerical constant $\lambda = 0.1$.

TABLE 1. MISE×100 of the estimation of h_X, compared with the MISE obtained when data are not censored, or not noisy, or neither censored nor noisy. MISE was averaged over 500 samples. Data are simulated with a Laplace noise, and an exponential censoring variable.

$s2n = 10$		0% censoring		20% censoring		40% censoring	
		$n = 400$	$n = 1000$	$n = 400$	$n = 1000$	$n = 400$	$n = 1000$
f_X Mixed Gamma	with noise	0.723	0.285	0.857	0.423	1.426	0.726
	without noise	0.793	0.287	1.120	0.381	2.071	0.690
f_X Beta	with noise	1.405	0.813	1.780	1.019	2.328	1.354
	without noise	1.288	0.641	1.767	0.793	2.454	1.050
f_X Gaussian	with noise	0.598	0.250	1.238	0.703	6.580	6.077
	without noise	0.656	0.201	2.009	0.628	8.934	5.808
f_X Gamma	with noise	0.805	0.361	0.843	0.364	0.988	0.408
	without noise	0.684	0.275	0.865	0.268	1.058	0.327
$s2n = 2.5$		0% censoring		20% censoring		40% censoring	
		$n = 400$	$n = 1000$	$n = 400$	$n = 1000$	$n = 400$	$n = 1000$
f_X Mixed Gamma	with noise	1.129	0.497	1.235	0.645	1.727	0.915
	without noise	0.793	0.287	1.120	0.381	2.071	0.690
f_X Beta	with noise	2.075	1.154	2.878	1.455	3.437	1.847
	without noise	1.288	0.641	1.767	0.793	2.454	1.050
f_X Gaussian	with noise	0.950	0.437	1.978	1.031	5.526	4.940
	without noise	0.656	0.201	2.009	0.628	8.934	5.808
f_X Gamma	with noise	1.173	0.613	1.379	0.660	1.538	0.821
	without noise	0.684	0.275	0.865	0.268	1.058	0.327

3.3. **Results.** The values of the MISE are computed from 500 simulated data sets, for each density and simulation scenario and are given (multiplied by 100) in Table 1. Note that even based on 500 simulated datasets, a variability remains in the MISE results, that might be due to the quotient estimator. Therefore, only the trends of the MISE should be interpreted.

Results are compared to estimators obtained in the three following cases: 1/ data with no noise and no censoring, 2/ data with no noise but censoring, 3/ data with noise but no censoring. These three cases can be considered as benchmarks for our situation including both noise and censoring. For case 1, h_X is estimated as a quotient of a projection estimator of f_X with a trigonometric basis and a Kaplan-Meier estimator of S_X. For case 2, h_X is estimated as a quotient with numerator

and denominator adapted from $(\widehat{f_X S_C})_{\hat{m}_1}$ and $(\widehat{S_{X\wedge C}})_m$ (removing the noise $1/|f_\varepsilon^*|$). For case 3, h_X is estimated as the quotient of the projection estimator of f_X with the trigonometric basis and an estimator of S_X directly deduced from $(\widehat{S_{X\wedge C}})_m$. Note that trigonometric polynomials are easy to implement but are sometimes subject to bad side-effects.

Table 1 shows that the MISE obtained with the new estimator are close to the MISE obtained with the more standard estimators without noise or without censoring. The results are satisfactory for the four distributions of X. The MISE are reduced when n increases, whatever the censoring level and the signal to noise ratio. Similarly, the MISE decreases when the censoring level decreases, whatever the value of n and the signal to noise ratio.

We also compare the MISE obtained for $\hat{h}_{\hat{m}_1,\hat{m}_2}$ with the MISE obtained on the same noisy and censored data but modeling either only the noise, or only the censoring, or neither the noise nor the censoring. Results are presented in Table 2 for data with 20% of censoring, small noise ($s2n = 10$) and $n = 400, 1000$. We see that when the model is misspecified, the MISE increases. This is especially true when censoring is neglected (two last columns). Neglecting the noise increases the MISE in the Gaussian and the Gamma case. For the Mixed Gamma and the Beta distributions, the MISE are of the same order in the first two columns, when censoring is appropriately modeled.

TABLE 2. MISE×100 of the estimation of h_X, compared with the MISE on the same noisy and censored data but assuming in the modeling either only the noise, or only the censoring, or neither the noise nor the censoring. MISE was averaged over 500 samples. Data are simulated with a Laplace noise, and an exponential censoring variable with 20% of censoring, small noise ($s2n = 10$) and $n = 400$ or $n = 1000$.

estimation assuming		noise censor	no noise censor	noise no censor	no noise no censor
f_X Mixed Gamma	$n = 400$	0.857	1.196	1.833	2.049
	$n = 1000$	0.423	0.397	1.378	1.307
f_X Beta	$n = 400$	1.780	1.877	2.682	1.760
	$n = 1000$	1.019	1.017	2.002	1.134
f_X Gaussian	$n = 400$	1.238	2.552	5.751	5.774
	$n = 1000$	0.703	1.178	5.017	4.982
f_X Gamma	$n = 400$	0.843	1.116	1.499	1.110
	$n = 1000$	0.364	0.506	0.992	0.675

4. Application to length of pregnancy

This work was motivated by the problem of estimating the physiological length of pregnancy, i.e. the time between spontaneous oocyte fertilization and spontaneous delivery.

Although many estimates have been reported, usually of around 40 weeks following last menstrual period (i.e. around 38 weeks after conception), they all rely on imperfect dating of the time origin since the precise time of conception remains unknown in spontaneously conceived pregnancies. In practice, the onset of pregnancy may be estimated by adding two weeks to the last menstrual period,

TABLE 3. Outcomes and main reasons for censoring in a sample of 8960 singleton pregnancies followed in Necker Enfants Malade teaching hospital

Termination of pregnancy or intrauterine fetal demise	110
Induction at term ($\geq$ 41 weeks)	903
Induction for maternal or fetal reasons before term	1490
Planned cesarean section	907
Spontaneous labor	5550

by biochemical tests and also by fetal ultrasound, which is in many cases the preferred method Stirnemann et al. [2013]. Prediction of the time origin using ultrasonographic measurement of fetal crown-rump is not exact and one should take into account this prediction error.

In such data, censoring may occur because of medically planned deliveries whenever maternal or fetal conditions require delivery prior to spontaneous labor. For example, induction (or cesarean section) may be offered whenever women reach 41 weeks and is medically indicated at 41 weeks and 3 days in the Obstetrics unit of Paris Necker Hospital, to avoid fetal complications of prolonged pregnancy. Other medical indications for timely delivery include maternal conditions such as preeclampsia or fetal conditions such as growth restriction or fetal malformations. For obvious reasons, termination of pregnancy or intrauterine fetal demise also require preterm induction of labor.

As already explained, the prediction error using ultrasonographic measurement of fetal crown-rump affects the time origin. Therefore it impacts both censoring times and the variable of interest which is the occurrence of a spontaneous onset of labor. This situation refers to the model studied in Section 2. In the following, we consider the measurement error as a Gaussian distribution with mean=0 and standard deviation of 0.35 weeks, as estimated by [Stirnemann et al., 2013].

The data we consider here is a sample of 8960 consecutive singleton pregnancies followed in the department of obstetrics, Necker teaching hospital in Paris, from the routine first trimester ultrasound around 12 weeks to delivery between 2011 and 2014. Dating of conception was performed by ultrasonographic measurement of crown-rump length in all cases. In this dataset, censoring occurred in 3410/8960 (38%) cases. The outcomes in the sample are presented in Table 3.

We estimated the hazard rate of the length of pregnancy for spontaneous delivery using estimator (11). The resulting hazard rate for spontaneous delivery is presented in Figure 1. This function increases rapidly from 37 weeks onwards reaching its maximum at 40 weeks and 6.5 days followed by a rapid decrease. In this population this result is markedly different from the usual estimate of 40 weeks that is considered in clinical practice. Therefore, our results would suggest that the true underlying length of pregnancy is longer than observed using noisy data. For comparison, in Figure 1, we also added the hazard rate function for misspecified models that do not model censoring and measurement error. In agreement with the simulation study (Table 2), the misspecified models show significant deviations from our estimator. Interestingly, the model that neglects dating error yields a maximum hazard at 40 weeks and 5 days, 1.5 days earlier than what we found using our estimator. Such a difference may impact on several aspects of clinical care and pregnancy management, when

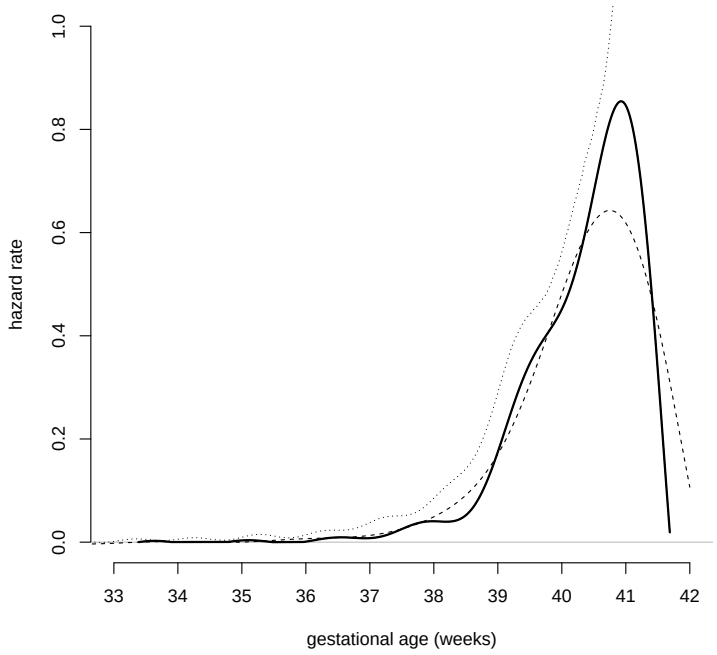

FIGURE 1. Hazard rate for spontaneous delivery estimated from the noisy and censored dataset of 8960 pregnancies. Solid line: our estimator (11) modeling both measurement error and censoring; broken line: censoring is modeled but not measurement error on time origin; dotted line: neither censoring nor measurement error is modeled

seeking to optimize obstetric protocols at term by reducing unnecessary inductions and planning follow-up in prolonged pregnancies.

5. Discussion and perspectives

In this section, we discuss an alternative model that could be considered when both noise and censoring contaminate the data. Assume that we observe

$$(W_j = (X_j + \varepsilon_j) \wedge C_j,\ \Delta_j = \mathbf{1}_{Z_j \leq C_j})$$

for $j = 1, \ldots, n$ and where $Z_j = X_j + \varepsilon_j$. Despite its simple presentation, seemingly closely related to model (1), this model is more difficult to work with. In this case, the most natural quantity to estimate is the density f_X of X, rather than the hazard function. However, we were not able to prove the upper bound of the L^2 risk of the estimator of f_X.

More precisely, we consider the estimator by deconvolution of f_X of X, as follows:

$$\hat{f}_{X,m}(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-iux}\frac{\hat{f}_Z^*(u)}{f_\varepsilon^*(u)}du$$

where

$$\hat{f}_Z^*(u) = \frac{1}{n}\sum_{j=1}^{n}\frac{\Delta_j}{S_C(W_j)}e^{iuW_j}. \tag{12}$$

The censoring correction $\Delta_j/S_C(W_j)$ is standard for such data and sometimes called "Inverse Probability Censoring Weight" (IPCW) in the literature. Then setting f_m such that $f_m^* = f^*\mathbf{1}_{[-\pi m,\pi m]}$, we would easily get under the assumption

Assumption (A2): $\tau_Z < \tau_C$ where $\tau_L = \sup\{y, S_L(y) > 0\}$,

$$\mathbb{E}(\|\hat{f}_{X,m} - f\|^2) \leq \|f - f_X\|^2 + 2\mathbb{E}\left(\frac{1}{S_C(Z_1)}\right)\frac{J(m)}{n}.$$

Taking the variance term as a penalty for model selection would give an adaptive estimator with good theoretical properties under **(A2)**.

As S_C is unknown, it can be estimated with the Kaplan-Meier estimator $\hat{S}_C$, with the modification suggested by Lo et al. [1989]:

$$\hat{S}_C(y) = \prod_{W_{(i)}\leq y}\left(\frac{n-i+1}{n-i+2}\right)^{1-\Delta_{(i)}}$$

where $(W_{(i)}, \Delta_{(i)})$ is ordered following the W_j's. Then this estimator can be plugged into (12) to obtain the estimator

$$\tilde{f}_{X,m}(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-iux}\frac{\tilde{f}_Z^*(u)}{f_\varepsilon^*(u)}du, \quad \text{with} \quad \tilde{f}_Z^*(u) = \frac{1}{n}\sum_{j=1}^{n}\frac{\Delta_j}{\hat{S}_C(W_j)}e^{iuW_j}. \tag{13}$$

However, Assumption **(A2)** is no longer adequate to handle the substitution of S_C by its estimator, and contradictory conditions appear. To obtain an upper bound of the $\mathbb{L}^2$ risk of $\tilde{f}_{X,m}$, one would need to restrict the variables to belong a fixed compact set, which is rather standard in the the context of estimation with censored data (see the discussion in Gross and Lai [1996]), but this is not possible due to the additional deconvolution step: it leads to an assumption regarding the Z's that is contradictory. Thus, the estimation of f_X remains an opened theoretical question in this model. Note that we implemented the method and provide in Table 4 the empirical results on simulation: from practical point of view, the procedure seems to work and gives coherent results.

TABLE 4. MISE×100 of the estimation of f_X, compared with the MISE obtained when data are not censored, or not noisy, or neither censored nor noisy. MISE was averaged over 100 samples. Data are simulated with a Laplace noise, and an exponential censoring variable.

$s2n = 10$		0% censoring		20% censoring		40% censoring	
		$n = 400$	$n = 1000$	$n = 400$	$n = 1000$	$n = 400$	$n = 1000$
f_X Mixed Gamma	with noise	0.203	0.097	0.381	0.171	0.445	0.185
	without noise	0.181	0.082	0.258	0.087	0.259	0.102
f_X Beta	with noise	0.271	0.193	0.353	0.258	0.432	0.255
	without noise	0.975	0.579	0.280	0.163	0.349	0.191
f_X Gaussian	with noise	0.139	0.054	0.527	0.255	1.719	0.973
	without noise	0.481	0.237	0.146	0.070	0.452	0.127
f_X Gamma	with noise	0.290	0.138	0.316	0.166	0.371	0.170
	without noise	0.549	0.235	0.196	0.083	0.211	0.114

6. Proofs

We recall the following version of Talagrand inequality.

Lemma 2. *Let $T_1, \dots, T_n$ be independent random variables and $\nu_n(r) = (1/n)\sum_{j=1}^n [r(T_j) - \mathbb{E}(r(T_j))]$, for r belonging to a countable class $\mathcal{R}$ of measurable functions. Then, for $\epsilon > 0$,*

$$\mathbb{E}[\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - (1+2\epsilon)H^2]_+ \le C\left(\frac{v}{n}e^{-K_1\epsilon\frac{nH^2}{v}} + \frac{M^2}{n^2C^2(\epsilon)}e^{-K_2C(\epsilon)\sqrt{\epsilon}\frac{nH}{M}}\right) \tag{14}$$

with $K_1 = 1/6$, $K_2 = 1/(21\sqrt{2})$, $C(\epsilon) = \sqrt{1+\epsilon} - 1$ and C a universal constant and where

$$\sup_{r\in\mathcal{R}} \|r\|_\infty \le M, \quad \mathbb{E}\left(\sup_{r\in\mathcal{R}} |\nu_n(r)|\right) \le H, \quad \sup_{r\in\mathcal{R}} \frac{1}{n}\sum_{j=1}^n \mathrm{Var}(r(T_j)) \le v.$$

Inequality (14) is a straightforward consequence of the Talagrand inequality given in [Klein and Rio, 2005]. Moreover, standard density arguments allow us to apply it to the unit ball of spaces.

The following elementary inequalities will be also used:

$$\forall u \in \mathbb{R}, \forall a \in \mathbb{R}, \quad \left|\frac{\sin(u)}{u}\right| \le 1 \text{ and } \left|\frac{e^{iua} - 1}{u}\right| \le |a|. \tag{15}$$

6.1. **Proof of Lemma 1.** Let us first remark that $\hat{S}^*_{X\wedge C}$ is well defined on $\mathbb{R}$ because

$$\lim_{u\to 0} \frac{e^{iuY_j} - f^*_\varepsilon(u)}{iu} = Y_j - \mathbb{E}(\varepsilon_1).$$

Moreover $\lim_{u\to 0} \hat{S}^*_{X\wedge C}(u) = \frac{1}{n}\sum_{i=1}^n Y_j - \mathbb{E}(\varepsilon_1)$ which tends a.s. when n grows to infinity to $\mathbb{E}(Y_1 - \varepsilon_1) = \mathbb{E}(X_1 \wedge C_1) = S^*_{X\wedge C}(0)$.

Then we prove that $\hat{S}^*_{X\wedge C}$ is an unbiased estimate of $S^*_{X\wedge C}$. We have

$$\mathbb{E}[\hat{S}^*_{X\wedge C}(u)] = \frac{1}{iu}\mathbb{E}[e^{iu(X\wedge C)} - 1] = \frac{1}{iu}\int (e^{iuz} - 1) f_{X\wedge C}(z)dz.$$

Then, noticing that $(e^{iuz}-1)/(iu) = \int_0^z e^{iuv}dv$ and that $\int_0^{+\infty}\int_0^{+\infty} |e^{iuv} f_{X\wedge C}(z)\mathbf{1}_{v\le z}|dvdz \le \mathbb{E}(X\wedge C) < \infty$, the Fubini Theorem implies that

$$\begin{aligned}\mathbb{E}[\hat{S}^*_{X\wedge C}(u)] &= \int_0^{+\infty}\left(\int_0^z e^{iuv}dv\right) f_{X\wedge C}(z)dz = \int_0^{+\infty} e^{iuv}\left(\int_v^{+\infty} f_{X\wedge C}(z)dz\right)dv \\ &= \int_0^{+\infty} e^{iuv} S_{X\wedge C}(v)dv = S^*_{X\wedge C}(u).\end{aligned}$$

6.2. **Proof of Proposition 1.** Let us set $\widetilde{(S_{X\wedge C})}_m = \widehat{(S_{X\wedge C})}_m - \psi_m(x)$. Clearly,

$$\begin{aligned}\mathbb{E}(\|S_{X\wedge C} - \widehat{(S_{X\wedge C})}_m\|^2) &= \|S_{X\wedge C} - (S_{X\wedge C})_m + \psi_m\|^2 + \mathbb{E}(\|\widetilde{(S_{X\wedge C})}_m - (S_{X\wedge C})_m\|^2) \\ &\le 2\|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 + 2\|\psi_m\|^2 \\ &\quad +\mathbb{E}(\|\widetilde{(S_{X\wedge C})}_m - (S_{X\wedge C})_m\|^2), \qquad (16)\end{aligned}$$

where $(S_{X\wedge C})_m$ is such that $(S_{X\wedge C})^*_m = S^*_{X\wedge C}1_{[-\pi m,\pi m]}$.

First we have, by Parseval formula,

$$2\|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 = \frac{1}{\pi}\int_{|u|\ge\pi m} |S^*_{X\wedge C}(u)|^2 du. \qquad (17)$$

Next, as ψ_m is the Fourier transform of $\mathbf{1}_{|x|\ge\pi m}/(2i\pi x)$, we have $\psi^*_m(u) = -1/(iu)\mathbf{1}_{|u|\ge\pi m}$, and

$$\|\psi_m\|^2 = (1/2\pi)\|\psi^*_m\|^2 = 1/(\pi^2 m). \qquad (18)$$

For the last term, we use Parseval equality again:

$$\mathbb{E}(\|\widetilde{(S_{X\wedge C})}_m - (S_{X\wedge C})_m\|^2) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m}\mathbb{E}(|\hat{S}^*_{X\wedge C}(u) - S^*_{X\wedge C}(u)|^2)du.$$

Let us set

$$\hat{S}^*_{X\wedge C}(u) - S^*_{X\wedge C}(u) = \frac{1}{n}\frac{1}{iu}\sum_{j=1}^n \frac{Z_j(u) - \mathbb{E}(Z_j(u))}{f^*_\varepsilon(u)}$$

with $Z_j(u) = e^{iuY_j} - 1$. We notice that $e^{iuY_j} - f^*_\varepsilon(u) - \mathbb{E}(e^{iuY_j} - f^*_\varepsilon(u)) = Z_j(u) - \mathbb{E}(Z_j(u))$. Thus,

$$\mathbb{E}(|\hat{S}^*_{X\wedge C}(u) - S^*_{X\wedge C}(u)|^2) = \frac{1}{nu^2}\frac{\mathrm{Var}(Z_1(u))}{|f^*_\varepsilon(u)|^2} \le \frac{1}{nu^2}\frac{\mathbb{E}(|e^{iuY_1} - 1|^2)}{|f^*_\varepsilon(u)|^2} = \frac{4}{n}\frac{\mathbb{E}(\sin^2(uY_1/2))}{u^2|f^*_\varepsilon(u)|^2}.$$

Thanks to inequality (15), we bound this term by $(1/n)\mathbb{E}(Y_1^2)$ for $|u| \in [0,1]$ and, using $|\sin(z)| \le 1$, $\forall z \in \mathbb{R}$, by $4/(nu^2)$ for $|u| > 1$. We get

$$\mathbb{E}(\|\widetilde{(S_{X\wedge C})}_m - (S_{X\wedge C})_m\|^2) \le \frac{\mathbb{E}(Y_1^2)}{\pi n}\int_0^1 \frac{du}{|f^*_\varepsilon(u)|^2} + \frac{4}{\pi n}\int_1^{\pi m}\frac{du}{u^2|f^*_\varepsilon(u)|^2}. \qquad (19)$$

Plugging (17), (18) and (19) into (16) gives the result of Proposition 1.

6.3. **Proof of Theorem 1.** Let $S_m = \{t \in \mathbb{L}_2(\mathbb{R}), \quad \mathrm{Supp}(t^*) \subset [-\pi m, \pi m]\}$. Note that the estimator $\widetilde{(S_{X\wedge C})}_m = \widetilde{(S_{X\wedge C})}_m - \psi_m(x)$ satisfies

$$\widetilde{(S_{X\wedge C})}_m = \arg\min_{t\in S_m} \gamma_n(t), \quad \gamma_n(t) = \|t\|^2 - \frac{2}{2\pi}\langle t^*, \hat{S}^*_{X\wedge C}\rangle$$

with $\hat{S}^*_{X\wedge C}$ given by (8). One can see this by noticing that

$$\gamma_n(t) = \frac{1}{2\pi}\left(\|t^* - \widetilde{(S_{X\wedge C})}^*_m\|^2 - \|\widetilde{(S_{X\wedge C})}^*_m\|^2\right)$$

is minimal on S_m for $t = \widetilde{(S_{X\wedge C})}_m \in S_m$. Thus $\gamma_n(\widetilde{(S_{X\wedge C})}_m) = -\|\widetilde{(S_{X\wedge C})}_m\|^2 = \min_{t\in S_m}\gamma_n(t)$.

On the other hand we have $\|\widehat{(S_{X\wedge C})}_m\|^2 = \|\widetilde{(S_{X\wedge C})}_m\|^2 + \|\psi_m\|^2$, since the support of the Fourier transforms of the functions in the norms are disjoint. Therefore

$$\begin{aligned}\|\widehat{(S_{X\wedge C})}_m\|^2 &= -\|\widetilde{(S_{X\wedge C})}_m\|^2 - \|\psi_m\|^2 \\ &= \min_{t\in S_m}\gamma_n(t) - \|\psi_m\|^2.\end{aligned}$$

Therefore, definition (10) of $\hat{m}_2$ can be written

$$\begin{aligned}\hat{m}_2 &= \arg\min_{m\in\{1,\dots,m_{n,2}\}}[-\|\widehat{(S_{X\wedge C})}_m\|^2 + \frac{3}{2}\|\psi_m\|^2 + \mathrm{pen}_2(m)] \\ &= \arg\min_{m\in\{1,\dots,m_{n,2}\}}[\min_{t\in S_m}\gamma_n(t) + \frac{1}{2}\|\psi_m\|^2 + \mathrm{pen}_2(m)]. \qquad (20)\end{aligned}$$

We notice that

$$\gamma_n(t) - \gamma_n(s) = \|t - S_{X\wedge C}\|^2 - \|s - S_{X\wedge C}\|^2 - \frac{2}{2\pi}\langle t^* - s^*, \hat{S}^*_{X\wedge C} - S^*_{X\wedge C}\rangle. \qquad (21)$$

The equality (20) for $\hat{m}_2$ implies that, $\forall m \in \{1,\dots,m_{n,2}\}$ and for all $t \in S_m$,

$$\gamma_n(\widetilde{(S_{X\wedge C})}_{\hat{m}_2}) + \frac{1}{2}\|\psi_{\hat{m}_2}\|^2 + \mathrm{pen}_2(\hat{m}_2) \le \gamma_n(t) + \frac{1}{2}\|\psi_m\|^2 + \mathrm{pen}_2(m).$$

Taking $t = (S_{X\wedge C})_m$ and using (21), this can be rewritten

$$\begin{aligned}\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C}\|^2 + \frac{1}{2}\|\psi_{\hat{m}_2}\|^2 &\le \|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 + \frac{1}{2}\|\psi_m\|^2 + \mathrm{pen}_2(m) \\ &\quad + \frac{2}{2\pi}\langle(\hat{S}^*_{X\wedge C})_{\hat{m}_2} - (S^*_{X\wedge C})_m, \hat{S}^*_{X\wedge C} - S^*_{X\wedge C}\rangle \\ &\quad - \mathrm{pen}_2(\hat{m}_2). \qquad (22)\end{aligned}$$

Let us define, for $t \in S_m$,

$$\nu_n(t) = \frac{1}{2\pi}\int t^*(-u)(\hat{S}^*_{X\wedge C}(u) - S^*_{X\wedge C}(u))du.$$

Then considering the term in the second line of (22):

$$T := \frac{2}{2\pi}\langle(\hat{S}^*_{X\wedge C})_{\hat{m}_2} - (S^*_{X\wedge C})_m, \hat{S}^*_{X\wedge C} - S^*_{X\wedge C}\rangle$$

and using that $\forall x, y \geq 0, 2xy \leq 4x^2 + y^2/4$, we get

$$\begin{aligned} T &\leq 2\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - (S_{X\wedge C})_m\| \sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} |\nu_n(t)| \\ &\leq \frac{1}{4}\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - (S_{X\wedge C})_m\|^2 + 4 \sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} |\nu_n(t)|^2 \\ &\leq \frac{1}{2}\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C}\|^2 + \frac{1}{2}\|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 + 4 \sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} |\nu_n(t)|^2. \end{aligned} \tag{23}$$

Plugging (23) into (22) yields

$$\begin{aligned} \frac{1}{2}\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C}\|^2 + \frac{1}{2}\|\psi_{\hat{m}_2}\|^2 &\leq \frac{3}{2}\|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 + \frac{1}{2}\|\psi_m\|^2 + \mathrm{pen}_2(m) \\ &\quad +4 \sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} |\nu_n(t)|^2 - \mathrm{pen}_2(\hat{m}_2). \end{aligned} \tag{24}$$

Now we split $\nu_n(t) = \nu_{n,1}(t) + R_n(t)$ with

$$R_n(t) = \frac{1}{2\pi}\int_{|u|\leq 1} t^*(-u)(\hat{S}^*_{X\wedge C}(u) - S^*_{X\wedge C}(u))du, \quad \nu_{n,1}(t) = \nu_n(t) - R_n(t).$$

We have

$$\sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} |\nu_n(t)|^2 \leq 2 \sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} R_n^2(t) + 2 \sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} |\nu_{n,1}(t)|^2.$$

Moreover by Schwarz Inequality, we have

$$\mathbb{E}\left(\sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} R_n^2(t)\right) \leq \frac{1}{2\pi}\mathbb{E}\left(\int_{|u|\leq 1} |\hat{S}^*_{X\wedge C}(u) - S^*_{X\wedge C}(u)|^2 du\right),$$

and from the proof of Proposition 1, we easily get

$$\mathbb{E}\left(\sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} R_n^2(t)\right) \leq \frac{2}{2\pi}\frac{\mathbb{E}(Y_1^2)}{n}\int_0^1 \frac{du}{|f^*_\varepsilon(u)|^2} = \frac{c}{n} \tag{25}$$

where c is defined in Proposition 1. For the other term we use the following Proposition.

Proposition 2. *Let $p(m, m') = n^{-1}\log(n^2)J_2(m \vee m')$, then*

$$\mathbb{E}\left(\sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} |\nu_{n,1}(t)|^2 - 3p(m, \hat{m}_2)\right)_+ \leq \frac{c'}{n}.$$

The proof of Proposition 2 follows from Talagrand inequality and is proved below. Now we notice that $3\kappa_2 p(m,m') \leq 6\mathrm{pen}_2(m) + 6\mathrm{pen}_2(m')$ so that

$$\begin{aligned} 8\mathbb{E}\left[\sup_{t\in S_{m\vee\hat{m}_2},\|t\|=1} |\nu_{n,1}(t)|^2 - \mathrm{pen}_2(\hat{m}_2)/8\right] &\leq 8\mathbb{E}\left(\sup_{t\in S_{m\vee\hat{m}_2},\|t\|=1} |\nu_{n,1}(t)|^2 - 3p(m,\hat{m}_2)\right)_+ \\ &\quad +(\frac{48}{\kappa_2}-1)\mathbb{E}(\mathrm{pen}_2(\hat{m}_2)) + \frac{48}{\kappa_2}\mathrm{pen}_2(m) \\ &\leq \frac{c'}{n} + \frac{48}{\kappa_2}\mathrm{pen}_2(m), \end{aligned} \tag{26}$$

for $48/\kappa_2 - 1 \leq 0$ i.e. $\kappa_2 \geq 48$. Plugging (25) and (26) in (24), we obtain, $\forall m \in \{1,\dots,m_{n,2}\}$,

$$\begin{aligned} \mathbb{E}(\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C}\|^2 + \|\psi_{\hat{m}_2}\|^2) &\leq 3\|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 + \|\psi_m\|^2 \\ &\quad +2(1+24/\kappa_2)\mathrm{pen}_2(m) + \frac{8c}{n}. \end{aligned}$$

To conclude, we notice that

$$\|\widehat{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C}\|^2 = \|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C} + \psi_{\hat{m}_2}\|^2 \leq 2(\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C}\|^2 + \|\psi_{\hat{m}_2}\|^2)$$

which implies

$$\mathbb{E}(\|\widehat{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C}\|^2) \leq \inf_{m\in\{1,\dots,m_{n,2}\}} \left(6\|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 + 2\|\psi_m\|^2 + 6\mathrm{pen}_2(m)\right) + \frac{c'}{n},$$

$c' = 16c$, which is the announced result. □

6.4. **Proof of Proposition 2.** We set, for $t \in S_m$,

$$r_t(x) = \frac{1}{2\pi}\int t^*(u)\frac{e^{iux}\mathbf{1}_{|u|\geq 1}}{iuf_\varepsilon^*(u)}du$$

, so that

$$\nu_{n,1}(t) = \frac{1}{n}\sum_{j=1}^n [r_t(Y_j) - \mathbb{E}(r_t(Y_j))].$$

Classically we write

$$\mathbb{E}\left(\sup_{t\in S_{m\vee\hat{m}_2},\|t\|=1} |\nu_{n,1}(t)|^2 - 3p(m,\hat{m}_2)\right)_+ \leq \sum_{m'\in\mathcal{M}_n} \mathbb{E}\left(\sup_{t\in S_{m\vee m'},\|t\|=1} |\nu_{n,1}(t)|^2 - 3p(m,m')\right)_+$$

and we apply Inequality of Lemma 2 to $\mathcal{R} = S_{m\vee m'}$, by using standard arguments of continuity of $t \mapsto \nu_{n,1}(t)$ and density of a countable subset of $S_{m\vee m'}$. Next we have to compute H^2, M, v such that

$$\sup_{t\in S_{m\vee m'}} \sup_{x\in\mathbb{R}} |r_t(x)| \leq M, \quad \mathbb{E}\left(\sup_{t\in S_{m\vee m'}} |\nu_n(r_t)|\right) \leq H, \quad \sup_{t\in S_{m\vee m'}} \frac{1}{n}\sum_{j=1}^n \mathrm{Var}(r_t(Y_j)) \leq v.$$

Similarly to previous computation, we get $H^2 = J_2(m \vee m')/n$, $v = J_2(m \vee m')$ and $M = \sqrt{J_2(m \vee m')}$. Moreover we take $\epsilon = 6\log(n^2) \vee 1 = 6\log(n^2)$ for $n \geq 2$, and we get

$$\mathbb{E}\left(\sup_{t\in S_{m\vee m'},\|t\|=1} |\nu_{n,1}(t)|^2 - 3p(m,m')\right)_+ \leq \frac{C}{n}\left(J_2(m\vee m')e^{-\log(n^2)} + \frac{J_2(m\vee m')}{n}e^{-K_2\sqrt{n}}\right)$$

using that $\epsilon \geq 1$. Now we have $J_2(m \vee m') \leq n$, by definition of $\mathcal{M}_{n,2}$ so that

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}\left(\sup_{t\in S_{m\vee m'},\|t\|=1} |\nu_{n,1}(t)|^2 - 3p(m,m')\right)_+ \leq \frac{C}{n}\left(\frac{\mathrm{card}(\mathcal{M}_n)}{n} + \mathrm{card}(\mathcal{M}_n)e^{-K_2\sqrt{n}}\right).$$

We notice that $\mathrm{card}(\mathcal{M}_n) \leq n$ and thus $\mathrm{card}(\mathcal{M}_n)e^{-K_2\sqrt{n}}$ is bounded to get the result.

Acknowledgments

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

References

N. Akakpo and C. Durot. Histogram selection for possibly censored data. *Mathematical Methods of Statistics*, 19:189–218, 2010.

A. Antoniadis, G. Gregoire, and G. Nason. Density and hazard rate estimation for right-censored data by using wavelet methods. *J. R. Stat. Soc., Ser. B*, 61:63–84, 1999.

E. Brunel and F. Comte. Penalized contrast estimation of density and hazard rate with censored data. *Sankhya*, 67:441–475, 2005.

E. Brunel and F. Comte. Adaptive estimation of hazard rate with censored data. *Communications in Statistics, Theory and methods*, 37:1284–1305, 2008.

F. Comte, Y. Rozenholc, and M.-L. Taupin. Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.*, 34(3):431–452, 2006.

I. Dattner, A. Goldenshluger, and A. Juditsky. On deconvolution of distribution functions. *Ann. Statist.*, 39:2477–250, 2011.

A. Delaigle and I. Gijbels. Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.*, 56(1):19–47, 2004.

S. Dohler and L. Ruschendorf. Adaptive estimation of hazard functions. *Probab. Math. Statist.*, 22:355–379, 2002.

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

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

S.T. Gross and T.L. Lai. Nonparametric estimation and regression analysis with left-truncated and right-censored data. *Journal of the American Statistical Association*, 91:1166–1180, 1996.

T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33:1060–1077, 2005.

L. Li. On the minimax optimality of wavelet estimators with censored data. *J. Statist. Plann. Inference*, 137:1138–1150, 2007.

L. Li. On the block thresholding wavelet estimators with censored data. *J. Multivariate Anal.*, 99: 1518–1543, 2008.

S. H. Lo, Y. P. Mack, and J. L. Wang. Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimators. *Probab. Theory Related Fields*, 80: 461–473, 1989.

M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053, 1999.

P. Reynaud-Bouret. Penalized projection estimators of the Aalen multiplicative intensity. *Bernoulli*, 12:633–661, 2006.

L. Stefanski and R.J. Carroll. Deconvoluting kernel density estimators. *Statistics*, 21(2):169–184, 1990.

J.J. Stirnemann, A. Samson, J.P. Bernard, and J.C. Thalabard. Day-specific probabilities of conception in fertile cycles resulting in spontaneous pregnancies. *Human Reproduction*, 28(4):1110–1116, 2013.