CENSORED DATA AND MEASUREMENT ERROR

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. When considering different practical situations, two different models can be written to describe such situations in which the measurement errors affect only the variable of interest or also the censoring variable. Different estimation strategies can be proposed to estimate the density or hazard rate of the underlying variables of interest. We explain these models and strategies and provide L^2-risk bounds for the data driven resulting estimators. Simulations illustrate the performances of the estimators. Lastly, the method is applied to a real data set.

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

1. Introduction

In many clinical situations, time-to-event may be only partly observed. For example, the timing of spontaneous delivery among pregnant women may be censored because of medical intervention whenever delivery is deemed necessary before its natural occurrence. Hence survival, or life times, may only be observed up to a censoring event, and are then considered as randomly right-censored data. 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.

In this paper, the censored data are measured with an error. Measurement error can affect censored data in two ways for which we give here two corresponding examples pertaining to women's health. In the first example introduced above, regarding the time between conception and spontaneous delivery, the date of pregnancy is unknown in spontaneously conceived pregnancies and can only be estimated up to an error using the last menstrual period or fetal ultrasound. 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. This is referred in the sequel as the first model and was the main motivation for this work with an application to real data. In the second case, the measurement error affects only the variable X. For example, the true spontaneous age of menopause X is always unknown although it may be estimated with an error. Furthermore, the age at menopause may be censored because of medical intervention. This censor C is not affected by the noise because the age of the female is exactly known. This is referred in the sequel as the second model.

Let us define these two models more precisely. Let ε denote the random error variable assumed to be independent of X and C. For both models, we assume that the observations are properly classified as censored or uncensored. In the first model, both the censored and uncensored observations are measured with error:

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

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

In the second model, only the variables X are measured with an additive error. Let us denote $Z_j = X_j + \varepsilon_j$. The observations are then

$$(2) \qquad \begin{aligned} W_j &= (X_j + \varepsilon_j) \wedge C_j = Z_j \wedge C_j, \quad j = 1, \dots, n \\ \Delta_j &= \mathbf{1}_{Z_j \leq C_j}, \end{aligned}$$

Note that if $\varepsilon_j \equiv 0$ (no noise), then both models reduce to the usual right-censoring model, and if $C_j \equiv +\infty$ (no censoring), then both models correspond to the convolution model.

The purpose of this work is to provide non-parametric estimators of functions of the distribution of X, based on the observations (Y_j, δ_j) for Model (1) or (W_j, Δ_j) for Model (2). In the context of censored data, it is standard to estimate either the density f_X of the variable X, or its survival function S_X, or the hazard function $h_X = f_X/S_X$. Here, we focus on the estimation of either h_X or f_X.

Nonparametric methods have already been proposed in related areas. Here, we are concerned by both the censoring and the deconvolution frameworks. Regarding censoring, Antoniadis et al. [1999] consider a wavelet hazard estimator which is not adaptive, 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 advocated. A pointwise estimation method for S_X has been proposed by Dattner et al. [2011] when the data are noisy but not censored.

Given that censoring error and additive measurement error are of very different nature, it is quite difficult to bring these two types of literature together. In Model (1), we propose a ratio-strategy to estimate the hazard rate h_X, using a deconvolution estimator of $f_X S_C$ in the numerator and a deconvolution estimator of $S_X S_C$ in the denominator, as proposed by Dattner et al. [2011]. In Model (2), we divide a deconvolution estimator of $f_X S_C$ by a Kaplan-Meier estimator of S_C to estimate the density f_X. In both cases, data-driven procedures and risk bounds are provided.

The paper is organized as follows. Section 2 deals with the first model and a quotient estimator of the hazard rate is proposed. Section 3 studies the second model and an estimator of the density is presented. Estimators are illustrated with a simulation study in Section 4 and are compared to results obtained when either no measurement error or no censored variables are considered. The motivating application of estimation of length of pregnancy is illustrated by an analysis of real data in Section 5. 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. Model (1)

2.1. **Setting.** In Model (1), 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 its characteristic function is such that

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

The following assumption, which is verified by exponential or Gamma distributions for examples, 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

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

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

$$S_m := \{t \in \mathbb{L}^2(\mathbb{R}), \text{supp}(t^*) \subset [-\pi m, \pi m]\}. \tag{3}$$

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:

$$(4) \qquad \widehat{(f_X S_C)}_m(x) = \frac{1}{2\pi n}\sum_{j=1}^n \int_{-\pi m}^{\pi m} \frac{e^{-iux}\delta_j e^{iuY_j}}{f_\varepsilon^*(u)} du.$$

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 obviously increases. The compromise between bias and variance is classically performed by choosing

$$(5) \qquad \hat{m}_1 = \arg\min_{m\in\{1,\ldots,m_{n,1}\}} (-\|\widehat{(f_X S_C)}_m\|^2 + \text{pen}_1(m)),$$

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

$$(6) \qquad \text{pen}_1(m) = \frac{\kappa_1}{n}\left(\frac{1}{n}\sum_{k=1}^n \delta_k\right)\log(J(m))J(m), \text{ with } J(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{du}{|f_\varepsilon^*(u)|^2}.$$

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

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

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

for C and C' two constants which do not depend on 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 can be square-integrable over $\mathbb{R}^+$ (thus over $\mathbb{R}$ if they are extended by 0) contrary to the 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]:

$$(\widehat{S_{X\wedge C}})_m(x) = \frac{1}{2} + \frac{1}{\pi n}\sum_{j=1}^{n} \mathrm{Re} \int_0^{\pi m} \frac{1}{iu}\left(\frac{e^{iu(Y_j - x)}}{f_\varepsilon^*(u)}\right) du. \tag{7}$$

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 (7) 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:

$$(\widehat{S_{X\wedge C}})_m(x) \quad = \quad \mathrm{Re}\left(\frac{1}{2\pi n}\sum_{j=1}^{n}\int_{-\pi m}^{\pi m} \frac{e^{-iux}}{iu}\left(\frac{e^{iuY_j}}{f_\varepsilon^*(u)} - 1\right) du\right) + \psi_m(x) \tag{8}$$

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 $\int_0^{+\infty} |\psi_m(x)|^2 dx = (1/2\pi^2)m^{-1}$.

Then, in order to compute the integrated $\mathbb{L}^2$-risk of $(\widehat{S_{X\wedge C}})_m$, we can see (8) 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,

$$\hat{S}_{X\wedge C}^*(u) = \frac{1}{n}\frac{1}{iu}\sum_{j=1}^{n}\left(\frac{e^{iuY_j}}{f_\varepsilon^*(u)} - 1\right). \tag{9}$$

Lemma 1. *The estimator $\hat{S}_{X\wedge C}^*$ defined by (9) 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 (8) can be seen as the Fourier inversion of (9). 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 on the non random part which has a known value. This allows us to write

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

We emphasize that 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 (7). Under assumptions* **(A1)** *and* **(A2)**, *we have*

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

where c *is a positive constant.*

The first two terms 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 decrease rate of the bias is slow. 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

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

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 noise we have in mind in those models must also have lower bounded supports. For example, an exponential distribution for ε yields a variance term of order m/n. The bias-variance compromise yields to choose an optimal value m_{opt} for the cutoff 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

$$J_2(m) := \frac{1}{\pi}\int_1^{\pi m} \frac{du}{u^2|f^*_\varepsilon(u)|^2}, \text{ and } \quad \text{pen}_2(m) = \kappa_2 \log(n)\frac{J_2(m)}{n}, \tag{10}$$

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

$$\hat{m}_2 = \arg\min_{m\in\{1,\dots,m_{n,2}\}} (-\|\widehat{(S_{X\wedge C})}_m\|^2 + \frac{3}{2\pi^2 m} + \text{pen}_2(m)), \tag{11}$$

for $m_{n,2}$ such that $m_{n,2} \le n$ and $J_2(m_{n,2}) \le n$, 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 (7) and* $\hat{m}_2$ *by (11). Then there exists a numerical constant* κ_0, *such that for* $\kappa_2 \ge \kappa_0$, *we have*

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

where c *is a numerical constant depending on* f^*_ε.

From the proof we find that $\kappa_2 = 12$ suits, but this theoretical value is too large in practice (see Section 4).

Our adaptive procedure $(\widehat{S_{X\wedge C}})_{\hat{m}_2}$ has the advantage of choosing a unique global cutoff $\hat{m}_2$ for m, instead of the pointwise selection procedure described in [Dattner et al., 2011]. The theoretical global rate is not as good as the pointwise one, which avoids the point $x = 0$ where a discontinuity occurs. In particular, integrating the pointwise estimator of Dattner et al. [2011] on a compact subset $[a, b]$ with $a > 0$ would restore a better bias order depending on the pointwise regularity of $S_{X\wedge C}$. Thus, we may expect that the finite sample properties of the estimator remain globally numerically satisfactory (see Section 4.2).

2.4. **Construction of the estimator of h_X.** The two proposed 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

$$\hat{h}_{\hat{m}_1,\hat{m}_2}(x) = \frac{(\widehat{f_X S_C})_{\hat{m}_1}(x)}{(\widehat{S_{X\wedge C}})_{\hat{m}_2}(x)} \mathbf{1}_{(\widehat{S_{X\wedge C}})_{\hat{m}_2}(x)\geq \lambda/\sqrt{n}}, \tag{12}$$

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. Model (2)

In this section, we discuss the alternative model. Assume now that we observe

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

for $j = 1, \dots, n$ and where $Z_j = X_j + \varepsilon_j$. We want to estimate the density f_X of X.

3.1. **Construction of the estimator.** In this case, we estimate the density 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

$$\hat{f}_Z^*(u) = \frac{1}{n} \sum_{j=1}^{n} \frac{\Delta_j}{S_C(W_j)} e^{iuW_j}. \tag{13}$$

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. 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. Note that $\hat{S}_C$ is such that:

$$\forall y \in \mathbb{R},\ \ \hat{S}_C(y) \geq \frac{1}{n+1}. \tag{14}$$

Moreover, it is known that it is a good estimator of S_C on any interval $[0,\tau]$ provided $[0,\tau] \subsetneq [0,\tau_C]$ where $\tau_C = \sup\{y, 1-S_C(y)<1\}$, see [Lo et al., 1989].

Then this estimator can be plugged into (13) to obtain the estimator

$$\tilde{f}_{X,m}(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-iux}\frac{\tilde{f}_Z^*(u)}{f_\varepsilon^*(u)}du, \quad \text{with} \quad \tilde{f}_Z^*(u) = \frac{1}{n}\sum_{j=1}^n \frac{\Delta_j}{\hat{S}_C(W_j)}e^{iuW_j}. \tag{15}$$

3.2. **Upper bound of the $\mathbb{L}^2$ risk.** Let f_m define by $f_m^* = f_X^*\mathbf{1}_{[-\pi m,\pi m]}$ such that f_m is the orthogonal projection of f on S_m defined by (3). As previously, f_m is the function which is estimated by $\tilde{f}_{X,m}$. This implies a nonparametric bias equal to the distance between f_X and f_m. To bound the mean integrated squared error (MISE) defined as $\mathbb{E}\|\tilde{f}_{X,m}-f_X\|^2$, we remark that

$$\begin{aligned}\mathbb{E}\|\tilde{f}_{X,m}-f_X\|^2 &= \mathbb{E}\left(\|f_X-f_m\|^2+\|f_m-\tilde{f}_{X,m}\|^2\right)\\ &\le \|f_X-f_m\|^2+2\mathbb{E}\|f_m-\hat{f}_{X,m}\|^2+2\mathbb{E}\|\hat{f}_{X,m}-\tilde{f}_{X,m}\|^2.\end{aligned}$$

The first term is the standard bias. We have to study the two other terms.

In the context of estimation with censored data, it is usually not possible to estimate the density over the whole domain, but only on a compact set (see the discussion in [Gross and Lai, 1996]). Therefore following Gross and Lai [1996], we consider this assumption regarding the Z's:

Assumption (A2) We assume that the Z_j's are in a compact set $[0,\tau]$ such that $a := S_C(\tau) > 0$ and $b := S_Z(\tau) > 0$.

Now, using $J(m)$ defined by (6), we can bound the MISE:

Proposition 2. *Consider Model 2 under* **(A2)**, *then the estimate defined by (15) satisfies:*

$$\mathbb{E}\|\tilde{f}_{X,m}-f_X\|^2 \le \|f_X-f_m\|^2 + A\frac{J(m)}{n} \quad \text{with} \quad A = 2\int_0^\tau \frac{f_Z(u)}{S_C(u)}du + \frac{4}{b^2a^4}\left(c_1+16\frac{c_3}{a^2b^4}\right)$$

with $J(m)$ defined by (6) and c_1, c_3 defined in Lemma 3 (see Appendix).

3.3. **Cut-off selection.** Now, the cut-off m has to be relevantly chosen from the data. Let us define

$$\text{pen}_3(m) = \frac{J(m)}{n}\left(\kappa_{3,1}\mathbb{E}\left(\frac{\Delta_1}{S_C(W_1)}\right)^2\log\left(J^3(m)\right)+\kappa_{3,2}\frac{4}{a^4b^2}\log n\right)$$

with $\kappa_{3,1}$ and $\kappa_{3,2}$ two constants to be calibrated on simulations. We also define

$$\tilde{m} = \arg\min_{m\in\{1,\dots,m_{n,3}\}}\left(-\|\tilde{f}_{X,m}\|^2+\text{pen}_3(m)\right), \tag{16}$$

where $m_{n,3} \le n$ is an integer such that $\text{pen}_3(m_{n,3}) \le C$. The following theorem yields a bound of the L^2 risk of the estimator $\tilde{f}_{X,\tilde{m}}$.

Theorem 2. *Assume that **(A1)** hold and $\tilde{f}_{X,\tilde{m}}$ is defined by (15) with $\tilde{m}$ as in (16). Then there exist a constant κ_4 such that*

$$\mathbb{E}(\|f_X-\tilde{f}_{X,\tilde{m}}\|^2) \le C\inf_{m\in\{1,\dots,m_{n,3}\}}\left(\|f_X-f_m\|^2+\text{pen}_3(m)\right)+\frac{C'}{n},$$

where C and C' are two constants.

The penalty $\mathrm{pen}_3(m)$ can not be exactly computed as the values a and $\mathbb{E}\left(\frac{\Delta_1}{S_C(W_1)}\right)^2$ are unknown. We propose to estimate them with empirical moment estimators and to plug them in the penalty function.

4. Simulation

4.1. **Design of simulation.** Simulations are used to evaluate the performances of the estimators of both models. For each design of simulations, 100 datasets are simulated. We consider samples of size $n = 400, 1000$. Data are simulated with a Laplace noise with variance σ^2 as follows:

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

with $\sigma = 1/(2\sqrt{5})$ or $\sigma = 1/(\sqrt{5})$. 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 $s2n = 5$ and $s2n = 10$ in our simulations ($\sigma = 1/(2\sqrt{5})$ or $\sigma = 1/(\sqrt{5})$).

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

4.2. **Estimator implementation for model (1).** We first describe the implementation of the numerator $(\widehat{f_X S_C})_{\hat{m}_1}$. The penalty depends on $J(m)$, which is computed by discretization of the integral. Then we compute $\mathrm{pen}_1(m)$ defined by (6) with the choice $\kappa_1 = 2$, obtained after a set of simulation experiments to calibrate it. We consider $m_{n,1} = \mathrm{argmax}(m \in \mathbb{N}, J(m)/n \leq 1)$. Following, we have the final estimation of $\hat{m}_1$ defined by (5). By plugging (5) in (4) we obtain $(\widehat{f_X S_C})_{\hat{m}_1}$ which is our numerator estimator.

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 $\mathrm{pen}_2(m)$ as defined by (10) with $\kappa_2 = 5$, after a set of simulation experiments to calibrate it. We define $m_{n,2} = \mathrm{argmax}(m \in \mathbb{N}, \hat{J}_2(m)/n \leq 1)$. Following, we have the final estimation of $\hat{m}_2$ defined by (11). By plugging this in (7) 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. Model (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 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	0.710	0.292	0.903	0.386	1.376	0.784
	without noise	0.730	0.299	1.108	0.353	1.747	0.734
f_X Beta	with noise	1.511	0.856	2.004	1.147	2.623	1.506
	without noise	1.430	0.618	1.824	0.766	2.370	0.924
f_X Gaussian	with noise	0.572	0.239	1.306	0.611	7.177	6.148
	without noise	0.613	0.215	1.838	0.431	8.482	5.646
f_X Gamma	with noise	0.785	0.344	0.847	0.351	0.955	0.412
	without noise	0.639	0.231	0.861	0.218	1.112	0.306
$s2n = 5$		0% censoring		20% censoring		40% censoring	
		$n = 400$	$n = 1000$	$n = 400$	$n = 1000$	$n = 400$	$n = 1000$
f_X Mixed Gamma	with noise	1.040	0.493	1.141	0.659	1.657	0.864
	without noise	0.810	0.290	0.951	0.390	2.019	0.602
f_X Beta	with noise	2.201	1.093	4.030	1.760	5.081	2.359
	without noise	1.387	0.611	1.634	0.732	2.100	0.942
f_X Gaussian	with noise	0.793	0.369	1.937	1.110	7.477	6.548
	without noise	0.476	0.203	2.005	0.606	8.667	5.912
f_X Gamma	with noise	1.044	0.650	1.503	0.832	2.094	0.873
	without noise	0.557	0.310	0.768	0.281	0.985	0.315

4.3. **Estimator implementation for model (2).** The implementation of the estimator $\tilde{f}_{X,\hat{m}}(x)$ is sensitive to the estimator $\hat{S}_C$, and especially to the constant $a = S_C(\tau)$, which both appear as denominators either in the estimator or in the penalty function. To avoid problems in 0, we decide to consider only the 95% first data of the ordered sample $(W_1, \ldots, W_n)$ (with both the censored and uncensored observations). Then τ is defined as the 95% quantile of the sample $(W_1, \ldots, W_n)$, the constant a is estimated as $\hat{a}$ by the value of $\hat{S}_C$ evaluated in τ and the moment $\mathbb{E}\left(\frac{\Delta_1}{S_C(W_1)}\right)^2$ is also estimated empirically on this 95% sample. Similarly, $\hat{b}$ is estimated with the Kaplan-Meier estimator of S_Z at τ. Finally $J(m)$ is computed by discretization of the integral. We define $\widehat{\text{pen}}_3(m)$ the estimator of $\text{pen}_3(m)$ as:

$$\widehat{\text{pen}}_3(m) = \frac{\hat{J}(m)}{n}\left(\kappa_{3,1}\widehat{\mathbb{E}\left(\frac{\Delta_1}{S_C(W_1)}\right)^2}\log\left(J^3(m)\right) + \kappa_{3,2}\frac{4}{\hat{a}^4\hat{b}^2}\log n\right).$$

TABLE 2. Model (2). 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
$s2n = 5$		0% censoring		20% censoring		40% censoring	
		$n = 400$	$n = 1000$	$n = 400$	$n = 1000$	$n = 400$	$n = 1000$
f_X Mixed Gamma	with noise	0.350	0.147	0.940	0.398	1.169	0.491
	without noise	0.196	0.081	0.214	0.079	0.251	0.123
f_X Beta	with noise	0.458	0.274	0.919	0.483	1.152	0.505
	without noise	0.916	0.642	0.306	0.183	0.413	0.179
f_X Gaussian	with noise	0.233	0.112	1.396	0.662	2.268	1.453
	without noise	0.526	0.202	0.128	0.104	0.392	0.132
f_X Gamma	with noise	0.444	0.250	0.790	0.372	0.818	0.430
	without noise	0.510	0.196	0.193	0.082	0.223	0.099

Throughout numerical estimations we will consider $\kappa_{3,1} = 0.005$ and $\tilde{\kappa}_{3,2} = 0.0003$, after a set of simulation experiments to calibrate them. Note that $\kappa_{3,2}$ is chosen small, which amounts to almost "kill" the associated term.

The computation of $\|\hat{f}_m\|$ is performed following [Comte et al., 2011] and [Comte et al., 2006]. We consider an estimation $\mathcal{M}_n$ defined by

$$\mathcal{M}_n = \{k/K,\ k = 1, \dots, m_{n,3}K\}$$

for a constant K, and by defining an integer $m_{n,3}$ such that $m_{n,3} = \operatorname{argmax}(m \in \mathbb{N}, \hat{J}(m)/n \leq 1)$. Following, we have the final estimation of $\tilde{m}$ defined by:

$$\widehat{\tilde{m}} = \underset{m=k/K, k\in\{1,\dots,m_{n,3}K\}}{\operatorname{argmin}} \Big(- \|\tilde{f}_m\|^2 + \widehat{\operatorname{pen}}(m) \Big) \tag{17}$$

Finally, by plugging (17) in (15) we obtain $\tilde{f}_{X,\widehat{\tilde{m}}}$ which is our final estimator.

4.4. **Results.** The values of the MISE are computed from 100 simulated data sets, for each density and simulation scenario and are given (multiplied by 100) in Tables 1 and 2 for models (1) and

(2), respectively. 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, f_X is estimated with a projection estimator with trigonometric polynomials [Massart, 2007] and h_X is estimated as a quotient of the former estimator of f_X and a Kaplan-Meier estimator of S_X. For case 2, f_X is estimated with a projection estimator with trigonometric polynomials as in Brunel and Comte [2005] and 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 $1/|f_\varepsilon^*|$). Note that trigonometric polynomials are easy to implement but are sometimes subject to bad side-effects. For case 3, f_X is estimated by deconvolution [Comte et al., 2006] and h_X is estimated as the quotient of the former estimator of f_X and an estimator of S_X directly deduced from $(\widehat{S_{X\wedge C}})_m$.

Tables 1 and 2 show that the MISE obtained with the new estimators are close to the MISE obtained with the more standard estimators without noise or without censoring. The results for both model (1) and (2) are satisfactory for the four distributions of X, even in the less favorable Gaussian case. 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 $\hat{h}_{\hat{m}_1,\hat{m}_2}$ (and $\tilde{f}_{X,\widehat{m}}$, respectively) 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 3 (and 4) for data with 20% of censoring, small noise ($s2n = 10$) and $n = 400$. In Table 3, 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. In Table 3, the two first columns are very close, suggesting that the most important point is to correct for censoring. The two last columns show that neglecting censoring can increase the MISE from 0.5 to 3.7 in the Gaussian case.

5. Application to length of pregnancy, using model 1

This work was motivated by the problem of estimating the physiological length of pregnancy, e.g. the time between conception and spontaneous delivery for which model 1 was developed. Although many estimates have been reported, usually of around 40 weeks last menstrual period (about 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, by biochemical tests and also by fetal ultrasound, which is in many cases the preferred method [Stirnemann et al., 2013]. The prediction error using ultrasonographic measurement of fetal crown-rump translates into a Gaussian distribution with mean=0 and standard deviation of 0.3 weeks. Since this error affects the time origin it will impact both censoring times and the variable of interest which is the occurrence of a spontaneous delivery. This situation refers to model 1 described in Section 2. The data we consider here is a sample of 9082 deliveries of live born babies followed in the department of obstetrics, Necker University Hospital in Paris. Dating of conception was performed by ultrasonographic measurement of crown-rump length in all cases. In such data, censoring may

Table 3. Model (1). 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 100 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	$n = 400$	0.779	1.080	1.659	1.846
	$n = 1000$	0.539	0.483	1.450	1.360
f_X Beta	$n = 400$	2.185	1.926	3.504	2.191
	$n = 1000$	1.028	1.012	1.810	1.095
f_X Gaussian	$n = 400$	1.560	2.603	5.340	5.398
	$n = 1000$	0.745	1.192	5.111	1.268
f_X Gamma	$n = 400$	0.985	1.373	1.326	1.086
	$n = 1000$	0.417	0.509	1.058	0.753

Table 4. Model (2). MISE×100 of the estimation of f_X, compared with the MISE obtained 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 100 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	$n = 400$	0.344	0.351	0.305	0.299
	$n = 1000$	0.157	0.115	0.196	0.194
f_X Beta	$n = 400$	0.428	0.341	0.518	1.095
	$n = 1000$	0.224	0.221	0.414	0.664
f_X Gaussian	$n = 400$	0.553	0.190	3.355	3.751
	$n = 1000$	0.301	0.124	3.247	3.675
f_X Gamma	$n = 400$	0.287	0.343	0.566	0.930
	$n = 1000$	0.186	0.202	0.414	0.659

occur because of medically planned deliveries because of maternal or fetal conditions requiring delivery prior to spontaneous labor. In this dataset, this happened in 3463/9082 (38%) cases. Using the estimator (12), 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 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.

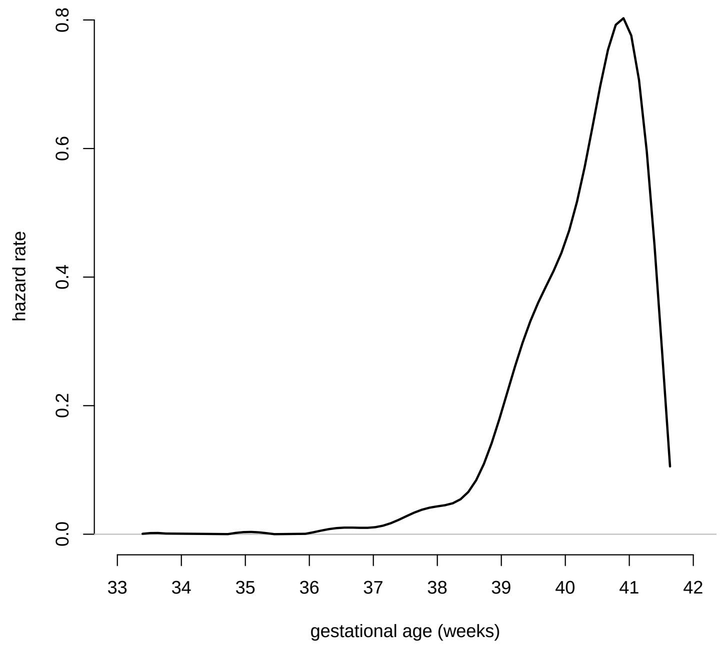

FIGURE 1. Hazard rate for spontaneous delivery estimated from the noisy and censored dataset of 9082 pregnancies with live born deliveries in Necker-Enfants Malades University Hospital

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 $\epsilon > 0$,*

$$\mathbb{E}[\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - (1+2\epsilon)H^2]_+ \leq C\left(\frac{v}{n}e^{-K_1\epsilon\frac{nH^2}{v}} + \frac{M^2}{n^2C^2(\epsilon)}e^{-K_2C(\epsilon)\sqrt{\epsilon}\frac{nH}{M}}\right) \tag{18}$$

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

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

Inequality (18) 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:

$$\forall u \in \mathbb{R}, \forall a \in \mathbb{R}, \quad \left|\frac{\sin(u)}{u}\right| \leq 1 \text{ and } \left|\frac{e^{iua}-1}{u}\right| \leq |a|. \tag{19}$$

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 $\int_0^{+\infty}\int_0^{+\infty} |e^{iuv} f_{X\wedge C}(z)\mathbf{1}_{v\leq z}|dvdz \leq \mathbb{E}(X\wedge C) < \infty$, 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,

$$\|S_{X\wedge C} - \widehat{(S_{X\wedge C})}_m\|^2 = \|S_{X\wedge C} - (S_{X\wedge C})_m + \psi_m\|^2 + \|\widetilde{(S_{X\wedge C})}_m - (S_{X\wedge C})_m\|^2$$

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

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

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 $Z_j(u) = e^{iuY_j} - 1$. Indeed $e^{iuY_j} - f^*_\varepsilon(u) - \mathbb{E}(e^{iuY_j} - f^*_\varepsilon(u)) = Z_j(u) - \mathbb{E}(Z_j(u))$. Then

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

Thanks to inequality (19), we bound this term by $(4/n)\mathbb{E}(Y_1^2)$ for $|u| \in [0,1]$ and by $4/(nu^2)$ for $|u| > 1$. We get

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

Moreover $\|S_{X\wedge C} - (S_{X\wedge C})_m + \psi_m\|^2 = \|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 + \|\psi_m\|^2$, since the support of the Fourier transforms of the functions in the norms are disjoint. By Parseval formula,

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

and as ψ_m is the Fourier transform of $\mathbf{1}_{|x|\geq\pi m}/(2x)$, $\psi_m^*(u) = (\pi/u)\mathbf{1}_{|u|\geq \pi m}$, $\|\psi_m\|^2 = (1/2\pi)\|\psi_m^*\|^2 = 1/(\pi^2 m)$. Gathering the three terms gives the result of Proposition 1.

6.3. **Proof of Theorem 1.** Let $S_m = \{t \in \mathbb{L}_2(\mathbb{R}), \quad \text{Supp}(t^*) \subset [-\pi m, \pi m]\}$. Then the estimator $\widetilde{(S_{X\wedge C})}_m = \widehat{(S_{X\wedge C})}_m - \psi_m(x)$ can be defined as

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

with $\hat{S}^*_{X\wedge C}$ given by (9). Now, as $\gamma_n(\widetilde{(S_{X\wedge C})}_m) = -\|\widetilde{(S_{X\wedge C})}_m\|^2$, and $\gamma_n(\widehat{(S_{X\wedge C})}_m) = -\|\widetilde{(S_{X\wedge C})}_m\|^2 + \|\psi_m\|^2$, we have

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

We notice that

$$\gamma_n(t) - \gamma_n(s) = \|t - S_{X\wedge C}\|^2 - \|s - S_{X\wedge C}\|^2 - \frac{2}{2\pi}\langle t^* - s^*, \hat{S}^*_{X\wedge C} - S^*_{X\wedge C}\rangle. \tag{20}$$

The definitions of $\hat{m}_2$ and $\widehat{(S_{X\wedge C})}_m$ imply that, $\forall m \in \{1,\dots,m_{n,2}\}$,

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

Using (20), this can be rewritten

$$\begin{aligned}\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C}\|^2 + \frac{1}{2}\|\psi_{\hat{m}_2}\|^2 &\leq \|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 + \frac{1}{2}\|\psi_m\|^2 + \text{pen}_2(m) \\ &\quad + \frac{2}{2\pi}\langle (\hat{S}^*_{X\wedge C})_{\hat{m}_2} - (S^*_{X\wedge C})_m, \hat{S}^*_{X\wedge C} - S^*_{X\wedge C}\rangle - \text{pen}_2(\hat{m}_2). \qquad (21)\end{aligned}$$

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

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

Then

$$\begin{aligned}&\frac{2}{2\pi}\langle (\hat{S}^*_{X\wedge C})_{\hat{m}_2} - (S^*_{X\wedge C})_m, \hat{S}^*_{X\wedge C} - S^*_{X\wedge C}\rangle \leq 2\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - (S_{X\wedge C})_m\| \sup_{t\in S_{m\vee\hat{m}}, \|t\|=1} |\nu_n(t)| \\ &\leq \frac{1}{4}\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - (S_{X\wedge C})_m\|^2 + 4\sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} |\nu_n(t)|^2 \\ &\leq \frac{1}{2}\|\widetilde{(S_{X\wedge C})}_{\hat{m}_2} - S_{X\wedge C}\|^2 + \frac{1}{2}\|S_{X\wedge C} - (S_{X\wedge C})_m\|^2 + 4\sup_{t\in S_{m\vee\hat{m}_2}, \|t\|=1} |\nu_n(t)|^2 \qquad (22)\end{aligned}$$

Plugging (22) into (21) yields

$$\begin{aligned}\frac{1}{2}\|\widetilde{(S_{X\wedge C})}_{\hat m_2}-S_{X\wedge C}\|^2+\frac{1}{2}\|\psi_{\hat m_2}\|^2 &\le \frac{3}{2}\|S_{X\wedge C}-(S_{X\wedge C})_m\|^2+\frac{1}{2}\|\psi_m\|^2+\mathrm{pen}_2(m)\\ &\quad +4\sup_{t\in S_{m\vee\hat m_2},\|t\|=1}|\nu_n(t)|^2-\mathrm{pen}_2(\hat m_2).\end{aligned}\tag{23}$$

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

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

We have

$$\sup_{t\in S_{m\vee\hat m_2},\|t\|=1}|\nu_n(t)|^2\le 2\sup_{t\in S_{m\vee\hat m_2},\|t\|=1}R_n^2(t)+2\sup_{t\in S_{m\vee\hat m_2},\|t\|=1}|\nu_n(t)|^2$$

and from the proof of Proposition 1, we easily get

$$\mathbb{E}\left(\sup_{t\in S_{m\vee\hat m_2},\|t\|=1}R_n^2(t)\right)\le \mathbb{E}\left(\int_{|u|\le 1}|\hat S^*_{X\wedge C}(u)-S^*_{X\wedge C}(u)|^2du\right)\le 2\frac{\mathbb{E}(Y_1^2)}{n}\int_0^1\frac{du}{|f_\varepsilon^*(u)|^2}.\tag{24}$$

For the other term we use the following Proposition.

Proposition 3. *Let $p(m,m')=n^{-1}\log(n^2)J_2(m\vee m')$, then*

$$\mathbb{E}\left(\sup_{t\in S_{m\vee\hat m_2},\|t\|=1}|\nu_n(t)|^2-3p(m,\hat m_2)\right)_+\le\frac{c'}{n}.$$

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

$$\begin{aligned}4\mathbb{E}\left[\sup_{t\in S_{m\vee\hat m_2},\|t\|=1}|\nu_n(t)|^2-\mathrm{pen}_2(\hat m_2)/4\right] &\le 4\mathbb{E}\left(\sup_{t\in S_{m\vee\hat m_2},\|t\|=1}|\nu_n(t)|^2-3p(m,\hat m_2)\right)_+\\ &\quad +(\frac{12}{\kappa_2}-1)\mathbb{E}(\mathrm{pen}_2(\hat m_2))+\frac{12}{\kappa_2}\mathrm{pen}_2(m)\\ &\le \frac{c'}{n}+\frac{12}{\kappa_2}\mathrm{pen}_2(m),\end{aligned}\tag{25}$$

for $12/\kappa_2-1\le 0$ i.e. $\kappa_2\ge 12$. Plugging (24) and (25) in (23), we obtain, $\forall m\in\{1,\dots,m_{n,2}\}$,

$$\mathbb{E}(\|\widetilde{(S_{X\wedge C})}_{\hat m_2}-S_{X\wedge C}\|^2+\|\psi_{\hat m_2}\|^2)\le 3\|S_{X\wedge C}-(S_{X\wedge C})_m\|^2+\|\psi_m\|^2+2(1+6/\kappa_2)\mathrm{pen}_2(m)+\frac{c'}{n}.$$

To conclude, we notice that

$$\|\widehat{(S_{X\wedge C})}_{\hat m_2}-S_{X\wedge C}\|^2=\|\widetilde{(S_{X\wedge C})}_{\hat m_2}-S_{X\wedge C}+\psi_{\hat m_2}\|^2\le 2(\|\widetilde{(S_{X\wedge C})}_{\hat m_2}-S_{X\wedge C}\|^2+\|\psi_{\hat m_2}\|^2)$$

which implies

$$\mathbb{E}(\|\widehat{(S_{X\wedge C})}_{\hat m_2}-S_{X\wedge C}\|^2)\le\inf_{m\in\{1,\dots,m_{n,2}\}}\big(6\|S_{X\wedge C}-(S_{X\wedge C})_m\|^2+2\|\psi_m\|^2+6\mathrm{pen}_2(m)\big)+\frac{c}{n},$$

which is the announced result. □

Proof of Proposition 3. Classically we write

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

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 $S_{m\vee m'}$.

Clearly we have $H^2=J_2(m\vee m')/n$, $v=J_2(m\vee m')$ and $M=\sqrt{J_2(m\vee m')}$. Moreover we take $\epsilon=6\log(n^2)\vee 1$, and we get

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

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

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

We notice that $\text{card}(\mathcal{M}_n)\leq n$ and we get the result.

6.4. **Proof of Proposition 2.** Note that, $\tilde{f}_{X,m}=\arg\min_{t\in S_m}\gamma_{n,3}(t)$ with

$$\gamma_{n,3}(t)=\|t\|^2-\frac{2}{n}\sum_{j=1}^n\frac{\Delta_j}{\hat{S}_C(W_j)}\frac{1}{2\pi}\int t^*(u)\frac{e^{iuW_j}}{f_\varepsilon^*(u)}du.$$

Thus $\gamma_{n,3}(\tilde{f}_{X,m})=-\|\tilde{f}_{X,m}\|^2$. To prove Proposition 2, the following Lemma is needed.

Lemma 3. *For all $k\in\mathbb{N}^*$, there exists a constant c_k depending on k such that*

$$\mathbb{E}\left(\sup_{y\in[0,\tau]}|\hat{S}_C(y)-S_C(y)|^{2k}\right)\leq\frac{c_k}{b^{2k}n^k},$$

where $b=S_Z(\tau)$ is defined in **(A2)**.

Recall that the MISE is bounded as

$$\mathbb{E}\|\tilde{f}_{X,m}-f_X\|^2\quad\leq\quad\|f_X-f_m\|^2+2\mathbb{E}\|f_m-\hat{f}_{X,m}\|^2+2\mathbb{E}\|\hat{f}_{X,m}-\tilde{f}_{X,m}\|^2.$$

The first term is the usual bias. Under assumption **(A2)**, the second term of the bound of the MISE can be studied as follows. We have

$$\mathbb{E}\|f_m-\hat{f}_{X,m}\|^2=\frac{1}{2\pi}\mathbb{E}\int_{-\pi m}^{\pi m}\frac{|\hat{f}_Z^*(u)-f_Z^*(u)|^2}{|f_\varepsilon^*(u)|^2}du\leq\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{\mathbb{E}|\hat{f}_Z^*(u)-f_Z^*(u)|^2}{|f_\varepsilon^*(u)|^2}du$$

$$=\quad\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n\left(\frac{\Delta_j}{S_C(W_j)}e^{iuW_j}-f_Z^*(u)\right)\right|^2}{|f_\varepsilon^*(u)|^2}du=\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{\frac{1}{n}\text{Var}\left(\frac{\Delta_j}{S_C(W_j)}e^{iuW_j}\right)}{|f_\varepsilon^*(u)|^2}du$$

$$\leq\quad\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{\frac{1}{n}\mathbb{E}\left(\left(\frac{\Delta_j}{S_C(W_j)}\right)^2\right)}{|f_\varepsilon^*(u)|^2}du=\frac{\mathbb{E}\left(\frac{\Delta_1}{S_C^2(W_1)}\right)}{2\pi n}\int_{-\pi m}^{\pi m}\frac{1}{|f_\varepsilon^*(u)|^2}du\leq\frac{1}{n}\mathbb{E}\left(\frac{1}{S_C(Z_1)}\right)J(m).$$

Assumption **(A2)** ensures that $\mathbb{E}\left(\frac{1}{S_C(Z_1)}\right)$ is well defined and finite

$$\mathbb{E}\left(\frac{1}{S_C(Z_1)}\right) \leq \int_0^\tau \frac{f_Z(u)}{S_C(u)}du \leq \frac{1}{a}\int_0^\tau f_Z(u)du \leq \frac{1}{a} < +\infty.$$

Thus

$$\mathbb{E}\|\hat{f}_{X,m} - f_X\|^2 \leq \frac{J(m)}{n}\int_0^\tau \frac{f_Z(u)}{S_C(u)}du. \tag{26}$$

The third term $\mathbb{E}\|\hat{f}_{X,m} - \tilde{f}_{X,m}\|^2$ is due to the estimation of the survival function. We have

$$\begin{aligned}
\mathbb{E}\|\hat{f}_{X,m} - \tilde{f}_{X,m}\|^2 &= \frac{1}{2\pi}\mathbb{E}\left(\int_{-\pi m}^{\pi m}\left|\frac{1}{n}\sum_{j=1}^n \frac{\Delta_j e^{iuW_j}}{f_\varepsilon^*(u)}\left(\frac{1}{S_C(W_j)} - \frac{1}{\hat{S}_C(W_j)}\right)\right|^2 du\right) \\
&\leq \frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{1}{n}\sum_{j=1}^n \mathbb{E}\left|\frac{\Delta_j e^{iuW_j}}{f_\varepsilon^*(u)}\left(\frac{1}{S_C(W_j)} - \frac{1}{\hat{S}_C(W_j)}\right)\right|^2 du \\
&= J(m)\frac{1}{n}\sum_{j=1}^n \mathbb{E}\left|\Delta_j\frac{\hat{S}_C(Z_j) - S_C(Z_j)}{S_C(Z_j)\hat{S}_C(Z_j)}\right|^2
\end{aligned}$$

Under assumption **(A2)**, we have $S_C(Z_j) > S_C(\tau) = a > 0$. Thus, using Lemma 3,

$$\begin{aligned}
&\mathbb{E}\left|\Delta_j\frac{\hat{S}_C(Z_j) - S_C(Z_j)}{S_C(Z_j)\hat{S}_C(Z_j)}\right|^2 \leq \frac{1}{a^2}\mathbb{E}\left|\Delta_j\frac{\hat{S}_C(Z_j) - S_C(Z_j)}{\hat{S}_C(Z_j)}\right|^2 \\
\leq\ & \frac{1}{a^2}\left[\mathbb{E}\left(\left|\frac{\hat{S}_C(Z_j) - S_C(Z_j)}{\hat{S}_C(Z_j)}\right|^2 \mathbf{1}_{\hat{S}_C(Z_j)>a/2}\right) + \mathbb{E}\left(\left|\frac{\hat{S}_C(Z_j) - S_C(Z_j)}{\hat{S}_C(Z_j)}\right|^2 \mathbf{1}_{\hat{S}_C(Z_j)\leq a/2}\right)\right] \\
\leq\ & \frac{1}{a^2}\left(\frac{\mathbb{E}\left(\sup_{x\in[0,\tau]}|\hat{S}_C(x) - S_C(x)|^2\right)}{(a/2)^2} + \mathbb{E}\left(\left|\frac{\hat{S}_C(Z_j) - S_C(Z_j)}{\hat{S}_C(Z_j)}\right|^2 \mathbf{1}_{|\hat{S}_C(Z_j)-S_C(Z_j)|\geq a/2}\right)\right) \\
\leq\ & \frac{1}{a^2}\left(\frac{c_1}{n}\frac{1}{(a/2)^2 b^2} + (n+1)^2\mathbb{E}\left(\sup_{x\in[0,\tau]}|\hat{S}_C(x) - S_C(x)|^2 \mathbf{1}_{\sup_{x\in[0,\tau]}|\hat{S}_C(x)-S_C(x)|\geq a/2}\right)\right) \\
\leq\ & \frac{1}{a^2}\left(\frac{c_1}{b^2 n}\frac{1}{(a/2)^2} + \frac{(n+1)^2}{(a/2)^4}\mathbb{E}\left(\sup_{x\in[0,\tau]}\left|\hat{S}_C(x) - S_C(x)\right|^6\right)\right)
\end{aligned}$$

where $(n+1)^2$ is due to (14) and we use $\mathbb{E}(|X|^2\mathbf{1}_{|X|>c}) \leq \mathbb{E}(|X|^6/c^4)$. Consequently we get

$$\mathbb{E}\|\hat{f}_{X,m} - \tilde{f}_{X,m}\|^2 \leq \frac{4}{a^4}\frac{1}{n}\left(\frac{c_1}{b^2} + 16\frac{c_3}{b^6 a^2}\right)J(m). \tag{27}$$

Gathering (26) and (27) implies the result. □.

6.5. **Proof of Theorem 2.** By definition of $\tilde{m}$, we have that $\forall m \in \{1,\dots,m_{n,3}\}$,

$$\gamma_{n,3}(\tilde{f}_{X,\tilde{m}}) + \mathrm{pen}(\tilde{m}) \leq \gamma_{n,3}(f_m) + \mathrm{pen}(m)$$

For any $m, m' \leq m_{n,3}$, $\forall t \in S_m$ and $\forall s \in S_{m'}$, we have the decomposition

$$\begin{aligned}\gamma_{n,3}(t) - \gamma_{n,3}(s) &= \|t\|^2 - \|s\|^2 - 2\langle t, \tilde{f}_{X,m}\rangle + 2\langle t, \tilde{f}_{X,m'}\rangle \\ &= \|t - f_X\|^2 - \|s - f_X\|^2 + 2\langle t-s, f_{m^\star}\rangle - 2\langle t-s, \tilde{f}_{X,m^\star}\rangle \\ &= \|t - f_X\|^2 - \|s - f_X\|^2 - 2\langle t-s, \tilde{f}_{X,m^\star} - \hat{f}_{X,m^\star}\rangle - 2\langle t-s, \hat{f}_{X,m^\star} - f_m\rangle\end{aligned}$$

where $m^\star = m \vee m'$. With $t = \tilde{f}_{X,\tilde{m}}$, $s = f_m$, $m^* = m \vee \tilde{m}$, we deduce that

$$\begin{aligned}\|\tilde{f}_{X,\tilde{m}} - f_X\|^2 \leq\ & \|f_m - f_X\|^2 + \mathrm{pen}_3(m) + 2\langle \tilde{f}_{X,\tilde{m}} - f_m, \tilde{f}_{X,m^*} - \hat{f}_{X,m^*}\rangle \\ & +2\langle \tilde{f}_{X,\tilde{m}} - f_m, \hat{f}_{X,m^*} - f_{m^*}\rangle - \mathrm{pen}_3(\tilde{m})\end{aligned}$$

We now detail the collection S_m. We introduce the usual sinus cardinal function $\phi(x) = \sin(x)/x$ and the corresponding normalized functions $\phi_{m,\ell}(x) = \sqrt{m}\phi(mx - \ell)$ for $\ell \in \mathbb{Z}$. The collection $(\phi_{m,\ell})_{\ell\in\mathbb{Z}}$ is an orthonormalized base which generates S_m. In the following, we work with the ball $B_m = \{t \in S_m, \|t\| = 1\}$.

The term $2\langle \tilde{f}_{X,\tilde{m}} - f_m, \hat{f}_{X,m^*} - f_{m^*}\rangle$ can be studied as follows.

$$\begin{aligned}&2\langle \tilde{f}_{X,\tilde{m}} - f_m, \hat{f}_{X,m^*} - f_{m^*}\rangle \leq 2\|\tilde{f}_{X,\tilde{m}} - f_m\| \sup_{t\in B_{m^*}} |\langle t, \hat{f}_{X,m^*} - f_{m^*}\rangle| \\ &\leq \frac{1}{8}\|\tilde{f}_{X,\tilde{m}} - f_m\|^2 + 8\sup_{t\in B_{m^*}} |\nu_n(t)|^2 \leq \frac{1}{4}\|\tilde{f}_{X,\tilde{m}} - f_X\|^2 + \frac{1}{4}\|f_X - f_m\|^2 + 8\sup_{t\in B_{m^*}} |\nu_n(t)|^2\end{aligned}$$

with $\nu_n(t) = \langle t, \hat{f}_{X,m^*} - f_{m^*}\rangle$. We proceed similarly for the term $2\langle \tilde{f}_{X,\tilde{m}} - f_m, \tilde{f}_{X,m^*} - \hat{f}_{X,m^*}\rangle$:

$$\begin{aligned}&2\langle \tilde{f}_{X,\tilde{m}} - f_m, \tilde{f}_{X,m^*} - \hat{f}_{X,m^*}\rangle \leq 2\|\tilde{f}_{X,\tilde{m}} - f_m\| \sup_{t\in B_{m^*}} |\langle t, \tilde{f}_{X,m^*} - \hat{f}_{X,m^*}\rangle| \\ &\leq \frac{1}{8}\|\tilde{f}_{X,\tilde{m}} - f_m\|^2 + 8\sup_{t\in B_{m^*}} |R_n(t)|^2 \leq \frac{1}{4}\|\tilde{f}_{X,\tilde{m}} - f_X\|^2 + \frac{1}{4}\|f_X - f_m\|^2 + 8\sup_{t\in B_{m^*}} |R_n(t)|^2\end{aligned}$$

with $R_n(t) = \langle t, \tilde{f}_{X,m^*} - \hat{f}_{X,m^*}\rangle$.

Let us introduce two functions p_1 and p_2 such that $8p_1(m,\tilde{m}) + 8p_2(m,\tilde{m}) \leq \mathrm{pen}(m) + \mathrm{pen}(\tilde{m})$. These two functions are detailed later on. This yields, $\forall m \leq m_{n,3}$,

$$\begin{aligned}\frac{1}{2}\|\tilde{f}_{X,\tilde{m}} - f_X\|^2 \leq\ & \frac{3}{2}\|f_m - f_X\|^2 + \mathrm{pen}_3(m) + 8\left(\sup_{t\in B_{m^*}} |\nu_n(t)|^2 - p_1(m,\tilde{m})\right)_+ \\ &+\ 8\left(\sup_{t\in B_{m^*}} |R_n(t)|^2 - p_2(m,\tilde{m})\right)_+ + 8p_1(m,\tilde{m}) + 8p_2(m,\tilde{m}) - \mathrm{pen}_3(\tilde{m})\end{aligned}$$

Let us start with the term $\left(\sup_{t\in B_{m^*}} |\nu_n(t)|^2 - p_1(m,\tilde{m})\right)_+$. First remark that $\nu_n(t) = \frac{1}{n}\sum_{j=1}^n \psi_t(W_j,\Delta_j) - \mathbb{E}(\psi_t(W_j,\Delta_j))$ where

$$\psi_t(W_j,\Delta_j) = \frac{\Delta_j}{S_C(W_j)}\frac{1}{2\pi}\int t^*(u)\frac{e^{iuW_j}}{f_\varepsilon^*(u)}du.$$

Thus $\nu_n(t)$ is a centered empirical process and Talagrand's inequality can be applied (Lemma 2).

Lemma 4. *Let $m^\star = m \vee m'$ and define*

$$p_1(m, m') = \kappa\, \mathbb{E}\left(\frac{\delta_W}{S_C(W)}\right)^2 \log(J^3(m^\star))\frac{J(m^\star)}{n}.$$

Then, there exists a numerical constant κ such that, for any $m, m' \in \mathcal{M}_n$, we have

$$\mathbb{E}\left(\left(\sup_{t\in B_{m^\star}} |\nu_n(t)|^2 - p_1(m, m')\right)_+\right) \le C\left(\frac{1}{an(m^\star)^2} + \frac{1}{n}e^{-c\sqrt{n}}\right)$$

where C and c are constants.

Now we bound $\mathbb{E}\left(\sup_{t\in B_{m^*}} |\nu_n(t)|^2 - p_1(m, \tilde{m})\right)_+$ where $m^* = m \vee \tilde{m}$. Indeed, Lemma 4 yields

$$\mathbb{E}\left(\sup_{t\in B_{m^*}} |\nu_n(t)|^2 - p_1(m, \tilde{m})\right)_+ \le \sum_{m'\in\mathcal{M}_N} \mathbb{E}\left(\left(\sup_{t\in B_{m^\star}} |\nu_n(t)|^2 - p_1(m, m')\right)_+\right)$$
$$\le C \sum_{m'\in\mathcal{M}_n}\left(\frac{1}{an(m^\star)^2} + \frac{1}{n}e^{-c\sqrt{n}}\right) \le \frac{C}{n}, \quad \text{if } m_{n,3} \le n^\alpha.$$

Now, we have to bound the term $\left(\sup_{t\in B_{m^*}} |R_n(t)|^2 - p_2(m, \tilde{m})\right)_+$. Remark that

$$R_n(t) = \frac{1}{n}\sum_{j=1}^n \Delta_j\left(\frac{1}{\hat{S}_C(W_j)} - \frac{1}{S_C(W_j)}\right)\frac{1}{2\pi}\int t^*(u)\frac{e^{iuW_j}}{f_\varepsilon^*(u)}du,$$

which is not an empirical process. The result is obtained with the following lemma.

Lemma 5. *Let $p_2(m, m') = \kappa\frac{4}{a^4b^2}\frac{J(m^\star)}{n}\log n$ where $m^\star = m \vee m'$. Then there exists a constant κ such that for any $m, m' \in\le m_{n,3}$,*

$$\mathbb{E}\left(\left(\sup_{t\in B_{m^\star}} |R_n(t)|^2 - p_2(m, m')\right)_+\right) \le C\frac{4}{a^4}\frac{1}{n^3},$$

for some constant C.

As for the previous term, we deduce that $\mathbb{E}\left(\sup_{t\in B_{m^*}} |R_n(t)|^2 - p_2(m, m')\right)_+ \le \frac{C}{n}$. Hence the result of the theorem. □.

Proof of Lemma 4 Talagrand's inequality requires to control the following quantities

$$\mathbb{E}\left[\sup_{t\in B_{m^\star}} |\nu_n(t)|\right] \le H, \quad \sup_{t\in B_{m^\star}} \frac{1}{n}\sum_{j=1}^n Var(\psi_t(W_j, \Delta_j)) \le v, \quad \sup_{t\in B_{m^\star}} \|\psi_t\|_\infty \le M$$

Start with the first term. We study $\mathbb{E}\left[\sup_{t\in B_{m^\star}}\nu_n^2(t)\right]$. Any $t\in B_{m^\star}$ can be written as $t(x)=\sum_{\ell\in\mathbb{Z}}a_{m^\star,\ell}\phi_{m^\star,\ell}(x)$ with $\sum_{\ell\in\mathbb{Z}}a^2_{m^\star,\ell}\leq 1$. Then we have

$$\mathbb{E}\left[\sup_{t\in B_{m^\star}}\nu_n^2(t)\right]\leq\sum_{\ell\in\mathbb{Z}}\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^n\psi_{\phi_{m^\star,\ell}}(W_j,\Delta_j)-\mathbb{E}(\psi_{\phi_{m^\star,\ell}}(W_j,\Delta_j))\right)^2$$

$$\begin{aligned}
&= \sum_{\ell\in\mathbb{Z}}\mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^n\psi_{\phi_{m^\star,\ell}}(W_j,\Delta_j)\right)=\frac{1}{n}\sum_{\ell\in\mathbb{Z}}\mathrm{Var}\left(\psi_{\phi_{m^\star,\ell}}(W_1,\Delta_1)\right)\leq\frac{1}{n}\sum_{\ell\in\mathbb{Z}}\mathbb{E}\;\psi^2_{\phi_{m^\star,\ell}}(W_1,\Delta_1)\\
&\leq \frac{1}{n}\sum_{\ell\in\mathbb{Z}}\mathbb{E}\left(\frac{1}{(2\pi)^2}\left|\int\phi^*_{m^\star,\ell}(u)\frac{\Delta_j}{S_C(W_j)}\frac{e^{iuW_j}}{f^*_\varepsilon(u)}du\right|^2\right)=\frac{1}{n}\frac{1}{2\pi}\mathbb{E}\left(\int_{-\pi m^\star}^{\pi m^\star}\left(\frac{\Delta_j}{S_C(W_j)}\right)^2\frac{du}{|f^*_\varepsilon(u)|^2}\right)\\
&\leq \frac{1}{n}\frac{1}{2\pi}\int_{-\pi m^\star}^{\pi m^\star}\mathbb{E}\left(\frac{\Delta_j}{S_C(W_j)}\right)^2\frac{du}{|f^*_\varepsilon(u)|^2}=\frac{1}{n}\mathbb{E}\left(\frac{\Delta_j}{S_C(W_j)}\right)^2J(m^\star)=:H^2
\end{aligned}$$

Then we study the second term.

$$\begin{aligned}
&\sup_{t\in B_{m^\star}}\frac{1}{n}\sum_{j=1}^n\mathrm{Var}(\psi_t(W_j,\Delta_j))\leq\sup_{t\in B_{m^\star}}\mathbb{E}\left(\psi_t(W_j,\Delta_j)\right)^2\\
&\leq\sup_{t\in B_{m^\star}}\mathbb{E}\left(\left(\frac{\Delta_j}{S_C(W_j)}\right)^2\|t^*\|^2\frac{1}{(2\pi)^2}\int_{-\pi m^\star}^{\pi m^\star}\frac{du}{|f_\varepsilon(u)|^2}\right)\leq\mathbb{E}\left(\frac{\Delta_j}{S_C(W_j)}\right)^2J(m^\star)=:v
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\sup_{t\in B_{m^\star}}\|\psi_t\|_\infty &= \sup_{t\in B_{m^\star}}\sup_{x,c}|\psi_t(x\vee c,\delta_{x\leq c})|=\sup_{t\in B_{m^\star}}\sup_{x,c}\left|\frac{\delta_{x\leq c}}{S_C(x\vee c)}\frac{1}{2\pi}\int t^*(u)\frac{e^{iux\vee c}}{f^*_\varepsilon(u)}du\right|\\
&\leq \frac{1}{a}\frac{1}{2\pi}\sup_{t\in B_{m^\star}}\left(\int|t^*(u)|^2du\right)^{1/2}\left(\int_{-\pi m^\star}^{\pi m^\star}\frac{du}{|f^*_\varepsilon(u)|^2}\right)^{1/2}=\frac{1}{a}\sqrt{J(m^\star)}=M
\end{aligned}$$

By choosing the constant $\epsilon^2=\frac{1}{K_1}\log(J^3(m^\star))$, we get

$$\frac{v}{n}e^{-K_1\epsilon^2\frac{nH^2}{v}}=\frac{1}{n}\mathbb{E}\left(\frac{\Delta_j}{S_C(W_j)}\right)^2J(m^\star)e^{-\log(J^3(m^\star))}\leq\frac{1}{a^2nm^{\star 2}}$$

and

$$\begin{aligned}
\frac{98b^2}{K_1n^2C^2(\epsilon^2)}e^{-\frac{2K_1}{7\sqrt{2}}C(\epsilon^2)\epsilon\frac{nH}{b}} &= \frac{98J(m^\star)}{K_1n^2a^2/K_1\log(J^3(m^\star))}e^{-\frac{6K_1}{7\sqrt{2}K_1}\log(J(m^\star))\frac{n\sqrt{J(m^\star)}a}{\sqrt{na}\sqrt{J(m^\star)}}}\\
&= \frac{98J(m^\star)}{n^2a^23\log(J(m^\star))}e^{-\frac{6}{7\sqrt{2}}\log(J(m^\star))\sqrt{na}}
\end{aligned}$$

The Talagrand Inequality ensures the lemma. □.

Proof of Lemma 5 We want to bound $\mathbb{E}\left(\left(\sup_{t\in B_{m^\star}}|R_n(t)|^2 - p_2(m,m')\right)_+\right)$ with

$$R_n(t) = \frac{1}{n}\sum_{j=1}^n \Delta_j\left(\frac{1}{\hat S_C(W_j)} - \frac{1}{S_C(W_j)}\right)\frac{1}{2\pi}\int t^*(u)\frac{e^{iuW_j}}{f_\varepsilon^*(u)}du.$$

We have

$$\sup_{t\in B_{m^\star}}|R_n(t)|^2 \quad\le\quad \frac{1}{n}\sum_{j=1}^n\left|\frac{S_C(Z_j)-\hat S_C(Z_j)}{S_C(Z_j)\hat S_C(Z_j)}\right|^2\frac{1}{2\pi}\int_{-\pi m^\star}^{\pi m^\star}\frac{1}{|f_\varepsilon^*(u)|^2}du = \frac{1}{n}\sum_{j=1}^n\left|\frac{S_C(Z_j)-\hat S_C(Z_j)}{S_C(Z_j)\hat S_C(Z_j)}\right|^2 J(m^\star).$$

We consider different random domains depending on the levels of $\hat S_C$: $\Omega_1 = \{x,\ \hat S_C(x)\le a/2\}$ and

$$\Omega_2 = \{\|S_C - \hat S_C\|_\infty \ge d\sqrt{\log n/n}\} = \{\sup_{x\in[0,\tau]}|S_C-\hat S_C|_\infty \ge d\sqrt{\log n/n}\}.$$

First, let us consider the domain Ω_1. Remark that

$$\begin{aligned}
&\mathbb{E}\left(\left(\sup_{t\in B_{m^\star}}|R_n(t)|^2\mathbf{1}_{\hat S_C\le a/2} - p_2(m,m')\right)_+\right) \le \mathbb{E}\left(\sup_{t\in B_{m^\star}}|R_n(t)|^2\mathbf{1}_{\hat S_C\le a/2}\right)\\
\le\ &\mathbb{E}\left(\sup_{t\in B_{m^\star}}\left|\frac{S_C(Z_1)-\hat S_C(Z_1)}{S_C(Z_1)\hat S_C(Z_1)}\right|^2\mathbf{1}_{\hat S_C\le a/2}J(m^\star)\right) \le n\mathbb{E}\left(\left|\frac{S_C(Z_1)-\hat S_C(Z_1)}{S_C(Z_j)\hat S_C(Z_1)}\right|^2\mathbf{1}_{\hat S_C\le a/2}\right)
\end{aligned}$$

because m is such that $J(m)\le n$. Recall that for any w, $S_C(w) > a$ and $\hat S_C(w) > 1/(n+1)$. Then for any $p>0$, we have

$$\begin{aligned}
\mathbb{E}\left(\frac{|S_C(Z_1)-\hat S_C(Z_1)|}{S_C(Z_1)\hat S_C(Z_1)}\mathbf{1}_{\hat S_C\le a/2}\right)^p &\le \left(\frac{n+1}{a}\right)^p\mathbb{E}\left(\|S_C-\hat S_C\|_\infty^p\mathbf{1}_{\hat S_C\le a/2}\right)\\
&\le \left(\frac{n+1}{a}\right)^p\left(\frac{2}{a}\right)^q\mathbb{E}\left(\|S_C-\hat S_C\|_\infty^{p+q}\right)
\end{aligned}$$

for any $q>0$. Lemma 3 yields with $q=p+6$, $p=2$, $\mathbb{E}\left(\frac{|S_C(W)-\hat S_C(W)|}{S_C(W)\hat S_C(W)}\mathbf{1}_{\hat S_C\ge a/2}\right)^2 \le c_5 2^8 a^{-10}n^{-3}$. Finally,

$$\mathbb{E}\left(\left(\sup_{t\in B_{m^\star}}|R_n(t)|^2\mathbf{1}_{\hat S_C\le a/2} - p_2(m,m')\right)_+\right) \quad\le\quad \frac{2^8}{a^{10}b^{10}}\frac{c_5}{n^2}$$

Then we consider the domain $\Omega_1^c\cap\Omega_2$.

$$\begin{aligned}
&\mathbb{E}\left(\left(\sup_{t\in B_{m^\star}}|R_n(t)|^2 - p_2(m,m')\right)_+\mathbf{1}_{\hat S_C>\frac{a}{2}}\mathbf{1}_{\|S_C-\hat S_C\|_\infty\le d\sqrt{\log n/n}}\right)\\
&\le \mathbb{E}\left(\left(\frac{4}{a^4}\|S_C-\hat S_C\|_\infty^2 J(m^\star) - p_2(m,m')\right)_+\mathbf{1}_{\hat S_C>\frac{a}{2}}\mathbf{1}_{\|S_C-\hat S_C\|_\infty\le d\sqrt{\log n/n}}\right)\\
&\le \mathbb{E}\left(\left(\frac{4}{a^4}d^2\frac{\log n}{n}J(m^\star) - p_2(m,m')\right)_+\right) = 0
\end{aligned}$$

because $p_2(m,m') > \frac{4}{a^4} d^2 \frac{\log n}{n} J(m^\star)$.

Lastly, let us finish with the domain $\Omega_1^c \cap \Omega_2^c$.

$$\begin{aligned}
&\mathbb{E}\left(\left(\sup |R_n(t)|^2 - p_2(m,m')\right)_+ \mathbf{1}_{\hat{S}_C > \frac{a}{2}} \mathbf{1}_{\|S_C - \hat{S}_C\|_\infty \geq d\sqrt{\log n/n}} \right) \\
&\leq \mathbb{E}\left(\frac{4}{a^4} \|S_C - \hat{S}_C\|_\infty^2 J(m^\star) \mathbf{1}_{\|S_C - \hat{S}_C\|_\infty \geq d\sqrt{\log n/n}} \right) \\
&\leq \frac{16}{a^4} n\, \mathbb{E}\left(\mathbf{1}_{\|S_C - \hat{S}_C\|_\infty \geq d\sqrt{\log n/n}} \right) \leq \frac{16}{a^4} n\, e^{-2(bd)^2 \log n + Abd\sqrt{\log n}}
\end{aligned}$$

where the last inequalities hold by deviation inequality (28). Therefore d is chosen as $cste/b$. □.

6.6. **Proof of lemma 3.** We use a non asymptotic exponential bound for the Kaplan-Meier estimator shown by [Földes and Rejto, 1981] which can be formulated as follows (see [Bitouzé et al., 1999])

$$\mathbb{P}\left(\sqrt{n} \|S_Z\, (\hat{S}_C - S_C)\|_\infty > \lambda \right) \leq 2.5\, e^{-2\lambda^2 + A\lambda}. \tag{28}$$

This inequality implies the Lemma 3. Indeed, we have

$$\begin{aligned}
&\mathbb{E}\left(\sup_{x\in[0,\tau]} |\hat{S}_C(x) - S_C(x)| \right)^{2k} \leq 2k \int_0^{+\infty} u^{2k-1}\, \mathbb{P}(\sup_{x\in[0,\tau]} |\hat{S}_C(x) - S_C(x)| > u)\, du \\
= \quad & 2k \int_0^{+\infty} u^{2k-1}\, \mathbb{P}(b^{-1} \sup_{x\in[0,\tau]} |S_Z\, (\hat{S}_C - S_C)(x)| > u)\, du \\
\leq \quad & 2k \int_0^{+\infty} u^{2k-1}\, \mathbb{P}(\sqrt{n} \|S_Z\, (\hat{S}_C - S_C)\|_\infty > b\sqrt{n}\, u)\, du \leq 5ke^{A^2/8} \int_0^\infty u^{2k-1} \exp\left(-2b^2 n \left[u - \frac{A}{4\sqrt{n}b} \right]^2 \right) du \\
\leq \quad & \frac{5e^{A^2/8}k}{2^k b^{2k}} \int_{-A/(2\sqrt{2})}^{+\infty} \left(z + \frac{A}{2\sqrt{2}} \right)^{2k-1} e^{-z^2} dz n^{-k} = c_k n^{-k} b^{-2k}.
\end{aligned}$$

Acknowledgements

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.

D. Bitouzé, B. Laurent, and P. Massart. A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimators. *Ann. Inst. Henri Poincaré*, 35:735–763, 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.

F. Comte, S. Gaiffas, and A. Guilloux. Adaptive estimation of the conditional intensity of marker-dependent counting processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47:1171–1196, 2011.

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.

A. Földes and L. Rejto. A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete 56*, 1:75–86, 1981.

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.

P. Massart. *Concentration inequalities and model selection. Ecole d'été de Probabilités de Saint-Flour 2003.* Lecture Notes in Mathematics 1896, 2007.

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.