

NONPARAMETRIC ESTIMATION OF THE INTENSITY FUNCTION OF A RECURRENT EVENT PROCESS

OLIVIER BOUAZIZ^(*), FABIENNE COMTE^(**) AND AGATHE GUILLOUX^(***)

ABSTRACT. In this paper, we consider the problem of estimating the intensity of a recurrent event process observed under a standard censoring scheme. We first propose a collection of kernel estimators for which we provide MSE and MISE bounds. Then, we describe and study an adaptive procedure of bandwidth selection, in the spirit of Goldenshluger and Lepski (2010) and we prove an oracle type bound for both the MSE and the MISE of the final estimator. The method is illustrated by simulation experiments.

May 18, 2012

AMS (2000) subject classification. 62N02, 62G05.

Keywords. Adaptive estimation. Censoring. Kernel estimator. Intensity. Non-parametric estimation. Recurrent event process.

1. INTRODUCTION

Recurrent event data arise in many fields such as medicine, insurance, economics and reliability. Medical examples include infections in HIV-infected subjects, tumor recurrences in cancer patients or epileptic seizures of patients. Such repeated events impact on the quality of life of the patients and also increase their risk of death. Therefore it becomes of natural interest to study the rate function of the recurrent event process which represents the instantaneous probability of experiencing a recurrent event at a given time. In this paper, we propose a new kernel estimator of the rate function when the recurrent event process is subject to right censoring and a terminal event is present. Then, we study the finite sample properties of this nonparametric estimator and develop a method to choose the bandwidth using data-driven techniques.

Regression methods have been widely studied to estimate the cumulative mean function or the rate function of the recurrent event process. For instance, Andersen

^(*): MAP5, UMR CNRS 8145, University and IUT Paris Descartes

^(**): MAP5, UMR CNRS 8145 and University Paris Descartes,

^(***): LSTA, University Pierre et Marie Curie.

and Gill [2] considered a Cox model in presence of right censoring and they studied the intensity of the recurrent process under a Poisson assumption. In the absence of terminal events, Pepe and Cai [18] and Lin *et al.* [15] performed estimation of the regression parameters in a more general model, taking into account time dependent covariates. Ghosh and Lin [10, 11] extended these results to the presence of terminal events and derived asymptotic properties of the regression parameter estimates. Finally, Bouaziz *et al.* [5] studied the cumulative mean function through a single-index assumption which can be seen as a generalization of the previous models. Asymptotic results on the parameter estimates were derived and data-driven techniques were used.

However, all these approaches rely on a modelisation assumption on the mean or the rate functions which may not hold in practice. In a more flexible way, non-parametric procedures were considered by several authors. In presence of censored data and without the Poisson assumption, Nelson [17] and Lawless and Nadeau [14] introduced an estimator of the cumulative mean function and derived a robust estimator of its variance. They also obtained confidence intervals which enable them to compare mean functions in a two-sample testing. Then, the theoretical properties of this estimator were derived in Ghosh and Lin [9]. In their main result, the cumulative mean function is proved to converge weakly to a zero mean Gaussian process. More recently, Dauxois and Sencey [8] studied a model of recurrent events with competing risks and a terminal event. They performed two-sample tests on the rate function although their estimation procedure did not need estimation of this function.

Few works using smoothing approach were also introduced in this framework. Bartoszyński *et al.* [3] briefly presented a kernel estimator of the rate function when the recurrent events were supposed to be distributed according to a Poisson process and the censored times constant. Then, Chiang *et al.* [6] extended their results to a more general setting where no Poisson assumption is made, no terminal events are considered and the censoring variables are random, but observed. They studied two types of kernel estimator of the rate function and gave asymptotic results for both estimators. Mainly, the asymptotic normality is proved and confidence intervals are derived using a bootstrap method, where theoretical arguments are provided to validate their procedures. Another kind of smoothing estimator was also introduced in Bouaziz *et al.* [5] to estimate the cumulative mean function when covariables and terminal events are present.

In this paper, we propose a new kernel estimator of the rate function, the derivative of the cumulative mean function, in a nonparametric context, with unobserved random censoring variables and terminal events. For this estimator, we develop an adaptive procedure to select the bandwidth, based on the recent work of Golden-shluger and Lepski [12]. We establish oracle inequalities, for the \mathbb{L}_2 -risk and the integrated \mathbb{L}_2 -risk of our estimator with a data-driven choice of the bandwidth. This is the first non-asymptotic result in this setting. In addition, the data-driven procedure is easily implementable.

The paper is structured as follows. After presenting the recurrent event model in the next section, we introduce our estimation procedure and infer a kernel-type estimator of the rate function in Section 3.1. In Sections 3.2 and 3.3 we give Mean Squared Error (MSE) and Mean Integrated Squared Error (MISE) bounds of the estimator for a fixed bandwidth. An adaptive procedure of bandwidth selection is then presented in Section 4. In particular, we derive our main result, an oracle bound for both the MSE and the MISE of our rate function estimator. A simulation study is conducted in Section 5 in order to assess the practical properties of the method. We also provide a comparison with a bootstrap method adapted from [6]. Finally, a few concluding remarks gathered in Section 6 end our presentation. The main proofs are detailed in Section 7 and some technical results are postponed to the appendix in Section 8.

2. NOTATIONS AND FIRST ASSUMPTIONS

2.1. Notations. In the following, for a real $q \geq 1$ and for a function $f : \mathbb{R} \mapsto \mathbb{R}$ such that $|f|^q$ is integrable or bounded, we denote by

$$\|f\|_q = \left(\int |f(x)|^q dx \right)^{1/q} \quad \text{and} \quad \|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|.$$

For simplicity, we also set $\|f\| = \|f\|_2$. The integrals and the supremum are restricted to the support of f and for τ a positive real number, we set $\|f\|_{\infty, \tau} = \sup_{x \in [0, \tau]} |f(x)|$.

We denote by $x^* = \arg \min_{x \in \mathcal{X}} f(x)$ the point x^* such that $f(x^*)$ realizes the minimum of the function f over the set \mathcal{X} , if it exists.

For k a positive integer, $f^{(k)}$ represents the derivative of order k of the function f , and we set by convention $f^{(0)} \equiv f$.

For h a positive real number, f_h represents the function $f_h(\cdot) = f(\cdot/h)/h$.

For two square-integrable functions f and g from \mathbb{R} to \mathbb{R} , we denote the convolution product of f and g by $f * g$, where

$$f * g(t) = \int f(t-x)g(x)dx = \int f(t)g(x-t)dt.$$

Finally, for two quantities $\alpha(n)$ and $\gamma(n)$, the notations $\alpha(n) \lesssim \gamma(n)$ and $\alpha(n) \propto \gamma(n)$ will be used to say that there exists a positive constant c such that $\alpha(n) \leq c\gamma(n)$ and $\alpha(n) = c\gamma(n)$ respectively.

2.2. Process assumptions. Let D be the terminal event (e.g. death) and $N^*(t)$ be the number of recurrent events experienced up to time t . As no recurrent event can occur after the terminal event, the process $N^*(\cdot)$ has jumps of size $+1$ on $[0, D]$.

Let C be the censoring time, assumed to be independent of both $N^*(\cdot)$ and D . The i.i.d. observations are then given by:

$$\begin{cases} T_i = D_i \wedge C_i \\ \delta_i = I(D_i \leq C_i) \\ N_i(t) = N_i^*(t \wedge C_i), \end{cases}$$

for $i = 1, \dots, n$. The distribution functions of D , C and $T = D \wedge C$ are respectively denoted by:

$$(1) \quad F(t) = \mathbb{P}[D \leq t], \quad G(t) = \mathbb{P}[C \leq t] \quad \text{and} \quad H(t) = \mathbb{P}[T \leq t], \quad t \geq 0.$$

The mean function of N^* is defined as $\mathbb{E}[N^*(t)] = \mu(t)$ for all $t \geq 0$. We assume that N^* has an intensity, in the sense that there exists a non-negative function λ such that, for all $t \geq 0$:

$$\mathbb{E}[N^*(t)] = \mu(t) = \int_0^t \lambda(s)ds.$$

Note that this definition is different from the conventional definition of the intensity of a recurrent event process. In our context $\lambda(t)$ refers to the occurrence probability of recurrent events at time t unconditionally to the history of the recurrent events process. In addition, $\lambda(t)$ is defined unconditionally to $t \leq D$ or $t \leq C$ contrary to the usual model assumptions in a recurrent events framework. This function is defined in Cook and Lawless [7] and is referred to as the rate function, and denoted by ρ . It is also introduced in Dauxois and Sencey [8] as the frequency function. On the opposite, the definition of the rate function in Chiang *et al.* [6] is different from ours. But notice that, multiplying their rate function by $1 - G(\cdot -)$, gives the intensity function as defined in our context.

Our aim here is to infer on this intensity function λ . To this purpose we first introduce some assumptions.

Assumption 1. Assume that:

- (i) C is independent of D and of the process $(N(t))_{t \geq 0}$,
- (ii) $\mathbb{P}[dN^*(C) \neq 0] = 0$,
- (iii) $\mathbb{P}[D = C] = 0$.

Assumption (i) is common in the context of recurrent events when censored data are present (see e.g. [8],[9]). Assumptions (ii) and (iii) are technical assumptions used to prevent us from ties between death, censoring and the apparition of recurrent event. Notice that in practical situations, if such ties exist, they can be dealt with by assigning to censored events values just slightly larger than their actual values.

The next assumption is introduced to circumvent problems arising in the tails of the distributions of C and N .

Assumption 2. Assume that:

- (i) for F, G, H defined by (1), there exist three positive constants τ, c_F and c_G such that $\tau < \inf\{t : H(t) = 1\}$ and, for all $t \in [0, \tau]$,

$$1 - G(t) \geq c_G, \quad 1 - F(t) \geq c_F.$$

- (ii) there exists $c_\tau > 0$, such that $N(t) \leq c_\tau$ almost surely for every $t \in [0, \tau]$.
- (iii) $\|\lambda\|_{\infty, \tau} := \sup_{t \in [0, \tau]} \lambda(t) < \infty$.

The first assumption is common in the context of estimation with censored observations (cf. [1]) while the second can be found e.g. in [8]. The last one is an additional condition only required for the pointwise setting.

2.3. Kernel and functional assumptions. In this paper, our goal is to perform non-parametric estimation of the function λ using a kernel-type estimator. Very classical regularity conditions are required for the intensity function and the kernel K . We first impose λ to belong to a Hölder space (see [20]).

Assumption 3. Let $\beta > 0$ and $L > 0$. Assume $\lambda^{(l)}$ exists for $l = \lfloor \beta \rfloor$ and

$$|\lambda^{(l)}(t+z) - \lambda^{(l)}(t)| \leq L|z|^{\beta-l}, \quad \forall z \in [-h, h], t \in [h, \tau - h].$$

We also need to impose some conditions on the kernel K and the bandwidth h . Note that the following set of assumptions are fulfilled by many standard kernel functions.

Assumption 4. Assume that

(i) K has a compact support $[-1, 1]$, $\int_{-1}^1 K(u)du = 1$ and $\|K\|^2 = \int_{-1}^1 K^2(u)du < \infty$,

(ii) $\|K\|_\infty = \sup_{u \in [-1, 1]} |K(u)| < \infty$,

(iii) K is a $l = \lfloor \beta \rfloor$ order kernel, in the sense that

$$\int_{-1}^1 u^j K(u)du = 0, \text{ for } j = 1, \dots, l, \quad \int_{-1}^1 u^\beta K(u)du < \infty,$$

(iv) $nh \geq 1$ and $0 < h < 1$.

Considering these four assumptions, it is now possible to perform estimation of λ . Our kernel estimator is introduced in the next section.

3. STUDY OF THE MSE AND THE MISE OF $\hat{\lambda}_h$

3.1. Kernel estimator. One of the difficulties for estimating the intensity function comes from the fact that the process $N^*(t)$ is not directly observable. Therefore, our estimation procedure is based on the next equality which provides a new expression of λ relying on the process N instead of $N^*(t)$.

Under Assumption 1 and since N^* does not jump after D , we have:

$$(2) \quad \mathbb{E}[dN(t)] = \mathbb{E}[dN^*(t \wedge C)] = \mathbb{E}[dN^*(t)\mathbb{E}[I(t \leq C)|N^*]] = \lambda(t)(1 - G(t-))dt.$$

The distribution function G is estimated by \hat{G} , the Lo *et al.* [16] modified Kaplan-Meier estimator,

$$\hat{G}(t) = \begin{cases} 1 - \prod_{i: T_{(i)} \leq t} \left(1 - \frac{1}{n - i + 2}\right)^{1 - \delta_{(i)}} & \text{if } t \leq T_{(n)}, \\ \hat{G}(T_{(n)}), & \text{if } t > T_{(n)}, \end{cases}$$

where $T_{(i)}$ denotes the i th order statistic associated to the sample T_1, \dots, T_n (that is $T_{(1)} \leq \dots \leq T_{(n)}$ and the $(\delta_{(i)})$'s are the δ_i 's associated to the new indexes). Notice that, from this definition, for all $t \geq 0$:

$$(3) \quad 1 - \hat{G}(t) \geq (n + 1)^{-1}.$$

Then, we can propose the following kernel estimator to estimate λ :

$$(4) \quad \hat{\lambda}_h(t) = \frac{1}{nh} \sum_{i=1}^n \int K\left(\frac{t-s}{h}\right) \frac{dN_i(s)}{1 - \hat{G}(s-)},$$

where K is a kernel function and h a bandwidth, satisfying Assumption 4. It is important to notice that the kernel is bounded with compact support on $[-1, 1]$

and consequently the integral in (4) will vanish outside the interval $[t - h, t + h]$. Therefore, given a bandwidth h , we will in the following only discuss estimation of λ for t such that $t \pm h \in [0, \tau]$.

Let us also introduce the following pseudo-estimator:

$$\tilde{\lambda}_h(t) = \frac{1}{nh} \sum_{i=1}^n \int K\left(\frac{t-s}{h}\right) \frac{dN_i(s)}{1-G(s-)},$$

which is the kernel estimator of λ in the case where G is known. In the following, the study of the quadratic error of $\hat{\lambda}_h - \lambda$ is divided into two steps. We first study the error of $\tilde{\lambda}_h - \lambda$, then the one of $\tilde{\lambda}_h - \hat{\lambda}_h$. The final results, a bound for the Mean Squared Error (MSE) at a fixed point and the Mean Integrated Squared Error (MISE) of $\hat{\lambda}_h - \lambda$ are given in Theorem 1.

3.2. Study of the pseudo estimator $\tilde{\lambda}_h$. We obtain with rather classical tools the following results for the risk of the pseudo-estimator. We state successively the pointwise error and the integrated error as the sum of a bias term and a variance term.

Proposition 1. *Under Assumptions 1 to 4 we have:*

(a) *for all $t \in [h, \tau - h]$:*

$$\mathbb{E} \left[(\tilde{\lambda}_h(t) - \lambda(t))^2 \right] \leq c_1^2 h^{2\beta} + \frac{c_\tau \|\lambda\|_{\infty, \tau} \|K\|^2}{nhc_G},$$

where

$$c_1 = \frac{L}{l!} \int_{-1}^1 |u|^\beta K(u) du.$$

(b) $\int_h^{\tau-h} \mathbb{E} \left[(\tilde{\lambda}_h(t) - \lambda(t))^2 \right] dt \leq \tau c_1^2 h^{2\beta} + \frac{c_\tau \Lambda(\tau)}{nh} \|K\|^2$, where

$$\Lambda(\tau) = \int_0^\tau \frac{\lambda(s) ds}{1-G(s-)}.$$

Proof. For the bias terms, observe that, from Equation (2)

$$\mathbb{E}[\tilde{\lambda}_h(t)] = \frac{1}{h} \int K\left(\frac{t-s}{h}\right) \lambda(s) ds$$

and using a change of variables, this leads to

$$\left(\mathbb{E}[\tilde{\lambda}_h(t)] - \lambda(t) \right)^2 \leq \left(\int_{-1}^1 K(u) (\lambda(t+uh) - \lambda(t)) du \right)^2.$$

Now write $\lambda(t + uh) = \lambda(t) + \lambda'(t)uh + \dots + \frac{(uh)^l}{l!}\lambda^{(l)}(t + \xi uh)$, for $0 \leq \xi \leq 1$, and use Assumptions 3 and 4 to obtain the required square bias bounds $c_1^2 h^{2\beta}$ in (a) and $\tau c_1 h^{2\beta}$ in (b).

Let us denote by $\mathbb{V}[X]$ the variance of X . For the variance terms, recalling that $K_h(\cdot) = (1/h)K(\cdot/h)$, we write

$$\begin{aligned} \mathbb{V}[\tilde{\lambda}_h(t)] &= \frac{1}{n} \mathbb{V} \left[\int \frac{K_h(t-s)}{1-G(s-)} dN(s) \right] \\ &\leq \frac{1}{n} \mathbb{E} \left[\left(\int \frac{K_h(t-s)}{1-G(s-)} dN(s) \right)^2 \right]. \end{aligned}$$

Then apply Lemma 9 (see Section 8), under Assumption 2 (ii):

$$\begin{aligned} \mathbb{V}[\tilde{\lambda}_h(t)] &\leq \frac{c_\tau}{n} \mathbb{E} \left[\int \frac{K_h^2(t-s)}{(1-G(s-))^2} dN(s) \right] \\ &\leq \frac{c_\tau}{n} \int \frac{K_h^2(t-s)}{1-G(s-)} \lambda(s) ds. \end{aligned}$$

From this point, Assumption 2 (i) and (iii) and the equality $\int K_h^2(t-s) ds = h^{-1} \|K\|^2$ give the pointwise variance bound of (a) while a change of variables gives the integrated variance term of (b).

Gathering the bias and the variance bounds gives the MSE and the MISE stated in (a) and (b) and thus the result of Proposition 1 follows. \square

3.3. Study of the estimator $\hat{\lambda}_h$. The most difficult part concerns the study of the difference between $\hat{\lambda}_h$ and $\tilde{\lambda}_h$. We give our final conclusion here and postpone the proof until Section 7.

Lemma 1. *Under Assumptions 1 to 4, for all $t \in [h, \tau - h]$, we have*

$$\mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 \right] \leq c \frac{\log(n)}{n},$$

and

$$\mathbb{E} \left[\int_h^{\tau-h} (\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 dt \right] \leq c' \frac{\log(n)}{n},$$

where c is a constant depending on $\|K\|_\infty, \|\lambda\|_{\infty, \tau}, c_\tau$ and c' is a constant depending on $\Lambda(\tau), \|K\|$ and c_τ .

Now, gathering the results of Proposition 1 (a) – (b) and Lemma 1 gives the following global bounds for the estimator.

Theorem 1. *Under Assumptions 1 to 4 we have:*

(a) for all $t \in [h, \tau - h]$,

$$\mathbb{E} \left[(\hat{\lambda}_h(t) - \lambda(t))^2 \right] \leq 2c_1^2 h^{2\beta} + 2 \frac{c_\tau \|\lambda\|_{\infty, \tau}}{nhc_G} \|K\|^2 + c \frac{\log(n)}{n},$$

(b)

$$\int_h^{\tau-h} \mathbb{E} \left[(\hat{\lambda}_h(t) - \lambda(t))^2 \right] dt \leq 2\tau c_1^2 h^{2\beta} + 2 \frac{c_\tau \Lambda(\tau)}{nh} \|K\|^2 + c' \frac{\log(n)}{n},$$

where c_1 is the constant defined in Proposition 1 and c and c' are the two constants introduced in Lemma 1.

Note that the inequalities stated in Theorem 1 are nonasymptotic. In both cases, they provide a bound which contains firstly a squared-bias term of order $h^{2\beta}$, secondly a variance term of order $1/(nh)$ and lastly a residual term which is negligible. If one wants to obtain an asymptotic convergence rate from these results, one has to optimize with respect to h to obtain the smallest possible order of the risk bounds. Classically, it appears that we should choose $h \propto n^{-1/(2\beta+1)}$ and obtain a rate proportional to $n^{-2\beta/(2\beta+1)}$. Nevertheless, to reach such a rate, we would need to know β , the regularity index of the unknown function: however, in practical situation this knowledge is usually unavailable. In the following, we provide a data-driven way of selecting the bandwidth which allows to reach almost or exactly the optimal rate without requiring the knowledge of β .

4. ADAPTIVE ESTIMATION OF λ

4.1. Pointwise bandwidth selection. In this part we want to select automatically a relevant bandwidth for our estimator using Goldenshluger and Lepski's [12] method. Let $t = t_0$ be the point of interest and define for any t :

$$\hat{\lambda}_{h,h'}(t) = K_{h'} * \hat{\lambda}_h(t),$$

where we recall that $f * g$ denotes the convolution product of the functions f and g . Note that, from the definition of $\hat{\lambda}_{h,h'}$,

$$\hat{\lambda}_{h,h'}(t) = \frac{1}{n} \sum_{i=1}^n \int K_{h'} * K_h(t-s) \frac{dN_i(s)}{1 - \hat{G}(s-)} = \frac{1}{n} \sum_{i=1}^n \int K_h * K_{h'}(t-s) \frac{dN_i(s)}{1 - \hat{G}(s-)},$$

so that $\hat{\lambda}_{h,h'}(t) = K_h * \hat{\lambda}_{h'}(t) = \hat{\lambda}_{h',h}(t)$. Then, for some $\kappa_0 > 0$, define

$$(5) \quad V_0(h) = \kappa_0 \frac{c_\tau \|\lambda\|_{\infty, \tau} \|K\|^2 \log(n)}{nhc_G}$$

and consider, for \mathcal{H}_n a discrete set of bandwidths specified in the following,

$$(6) \quad A_0(h, t_0) = \sup_{h' \in \mathcal{H}_n} \left\{ (\hat{\lambda}_{h'} - \hat{\lambda}_{h, h'})^2(t_0) - V_0(h') \right\}_+.$$

Lastly, we define our adaptive estimator in the following way:

$$(7) \quad \hat{h}(t_0) = \operatorname{argmin}_{h \in \mathcal{H}_n} (A_0(h, t_0) + V_0(h)) \quad \text{and} \quad \check{\lambda}(t_0) = \hat{\lambda}_{\hat{h}(t_0)}(t_0).$$

Theorem 2. *Under Assumptions 1 to 4, and if \mathcal{H}_n is a finite discrete set of bandwidths such that $\operatorname{Card}(\mathcal{H}_n) \leq n$,*

$$(8) \quad \forall h \in \mathcal{H}_n, \quad nh \geq \kappa_1 \log(n), \quad \text{for some } \kappa_1 \geq 0,$$

and

$$(9) \quad \sum_{k: h_k \in \mathcal{H}_n} \frac{1}{nh_k} \lesssim \log^a(n), \quad \text{for some } a \geq 0,$$

then there exists a constant κ_0 such that $\check{\lambda}$ defined by (5), (6) and (7) satisfies:

$$(10) \quad \forall h \in \mathcal{H}_n, \quad \mathbb{E} \left[(\check{\lambda}(t_0) - \lambda(t_0))^2 \right] \leq c(c_1^2 h^{2\beta} + V_0(h)) + c' \frac{\log^{(1+a)}(n)}{n},$$

where c is a real (absolute) number and c' a real constant depending on c_τ , $\|\lambda\|_{\infty, \tau}$ and c_G .

Remark 1. Note that $V_0(h)$ contains several types of terms:

- κ_0 , a real number. The proof below shows that κ_0 taken equal to 80 would give the theoretical result but a much smaller value works, in practice (see Section 5).
- $\log(n)/(nh)$ which gives the asymptotic order of the term and is known.
- $\|K\|$, a known constant, as the kernel is user chosen.
- c_τ and $\|\lambda\|_{\infty, \tau}$ which are unknown quantities that can respectively be estimated by

$$(11) \quad \hat{c}_\tau = \max_{1 \leq i \leq n} N_i(\tau), \quad \widehat{\|\lambda\|}_{\infty, \tau} = \sup_{x \in [h_n, \tau - h_n]} \hat{\lambda}_{h_n}(x).$$

Here h_n is an arbitrary bandwidth (it can be taken equal to $n^{-1/5}$ for instance). Note that if we replace in $V_0(h)$ the unknown terms by their estimates given in (11), we get an estimate $\hat{V}_0(h)$. Inserting this in theoretical part would imply several additional steps to the study of the estimate. For the sake of simplicity, we do not provide this part of the study.

The bound (10) holds for all $h \in \mathcal{H}_n$ and therefore reaches automatically the rate $(n/\log(n))^{-2\beta/(2\beta+1)}$ provided that an optimal value for h of order $(n/\log(n))^{-1/(2\beta+1)}$ belongs to \mathcal{H}_n . We can note that a logarithmic loss occurs here with respect to the optimal non-adaptive rate. This is also what happens for classical density estimation and we can thus conjecture that the procedure is nevertheless adaptive optimal.

Example of \mathcal{H}_n . Considering constraints (8) and (9) on \mathcal{H}_n , we can propose

$$\mathcal{H}_n = \left\{ \frac{k}{n} : k = \lfloor \log^2(n) \rfloor, \lfloor \log^2(n) \rfloor + 1, \dots, n \right\}$$

so that $\text{Card}(\mathcal{H}_n) \leq n$ and $\forall k = \lfloor \log^2(n) \rfloor, \dots, n$, we have $h_k \in [n^{-1}, 1]$ and $h_k \geq \log(n)/n$ which gives (8). Moreover, $k_0 = \lfloor n^{2\beta/(2\beta+1)}(\log(n))^{1/(2\beta+1)} \rfloor$ is guaranteed to be such that $k_0/n \propto (n/\log(n))^{-1/(2\beta+1)}$ belongs to \mathcal{H}_n . Besides, $\sum_k 1/(nh_k) = O(\log(n))$ and condition (9) holds with $a = 1$.

4.2. Global bandwidth selection. In the global risk setting, we set, for some $\kappa > 0$,

$$(12) \quad V(h) = \kappa \frac{c_\tau \Lambda(\tau) \|K\|^2}{nh}$$

and we consider for \mathcal{H}_n a discrete set of bandwidths specified in the following,

$$(13) \quad A(h) = \sup_{h' \in \mathcal{H}_n} \left\{ \|\hat{\lambda}_{h'} - \hat{\lambda}_{h,h'}\|^2 - V(h') \right\}_+.$$

Finally we define:

$$(14) \quad \hat{h} = \underset{h \in \mathcal{H}_n}{\text{argmin}} (A(h) + V(h)) \quad \text{and} \quad \lambda^* = \hat{\lambda}_{\hat{h}}.$$

Theorem 3. *Under Assumptions 1 to 4, and if \mathcal{H}_n is a finite discrete set of bandwidths such that $\text{Card}(\mathcal{H}_n) \leq n$, condition (9) is fulfilled and*

$$(15) \quad \sum_{k : h_k \in \mathcal{H}_n} \exp(-b/h_k) < +\infty, \quad \forall b > 0,$$

then there exists a constant κ such that λ^ defined by (12), (13) and (14) satisfies:*

$$(16) \quad \forall h \in \mathcal{H}_n, \quad \int_1^{\tau^{-1}} \mathbb{E} \left[(\lambda^*(t) - \lambda(t))^2 \right] dt \leq c(\tau c_1^2 h^{2\beta} + V(h)) + c' \frac{\log^{1+a}(n)}{n},$$

where c is a numerical constant and c' a constant depending on c_τ , $\Lambda(\tau)$ and c_G .

Remark 2. Note that all the points in Remark 1 can be transposed to $V(h)$. The additional term $\Lambda(\tau)$ is also unknown and can be estimated by:

$$\hat{\Lambda}(\tau) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{dN_i(s)}{(1 - \hat{G}(s-))^2}.$$

It is worth emphasizing here that, if \mathcal{H}_n is large enough to contain bandwidths of order $h_{opt} \propto n^{-1/(2\beta+1)}$, then the adaptive estimator automatically reaches the optimal rate $n^{-2\beta/(2\beta+1)}$, without requiring the knowledge of β . Compared to the pointwise setting, no logarithmic loss occurs here.

Let us now give two examples of \mathcal{H}_n fulfilling Assumption 4 (iv), conditions (9) and (15).

Example 1. Take

$$\mathcal{H}_n = \left\{ h_k = \frac{1}{k} : k = 1, 2, \dots, \lfloor \sqrt{n} \rfloor \right\}.$$

Then $\text{Card}(\mathcal{H}_n) \leq \sqrt{n} \leq n$ and $\forall k = 1, \dots, \lfloor \sqrt{n} \rfloor$, we have $h_k \in [n^{-1}, 1]$. Moreover

$$\sum_{k : h_k \in \mathcal{H}_n} (1/(nh_k)) = \frac{1}{n} \sum_{k=1}^{\lfloor \sqrt{n} \rfloor} k = O(1)$$

which ensures condition (9). Lastly

$$\sum_{k : h_k \in \mathcal{H}_n} \exp(-b/h_k) = \sum_{k=1}^{\lfloor \sqrt{n} \rfloor} e^{-bk} = O(1)$$

and (15) is ensured.

Let us emphasize that since $h_{opt} \propto n^{-1/(2\beta+1)}$, the condition $n^{-1/2} \leq n^{-1/(2\beta+1)} \leq 1$ is required, that is $\beta \geq 1/2$. This means that there is a minimal regularity condition to impose on the function of interest for (16) to hold.

Example 2. Take

$$\mathcal{H}_n = \left\{ h_k = \frac{1}{2^k}, k = 1, 2, \dots, \lfloor \log(n)/\log(2) \rfloor \right\}.$$

Then $\text{Card}(\mathcal{H}_n) \leq \log(n)/\log(2) \leq n$ and $\forall k = 1, 2, \dots, \lfloor \log(n)/\log(2) \rfloor$, we have $h_k \in [n^{-1}, 1]$. Moreover

$$\sum_{k : h_k \in \mathcal{H}_n} (1/(nh_k)) = \frac{1}{n} \sum_{k=1}^{\lfloor \log(n)/\log(2) \rfloor} 2^k = O(1),$$

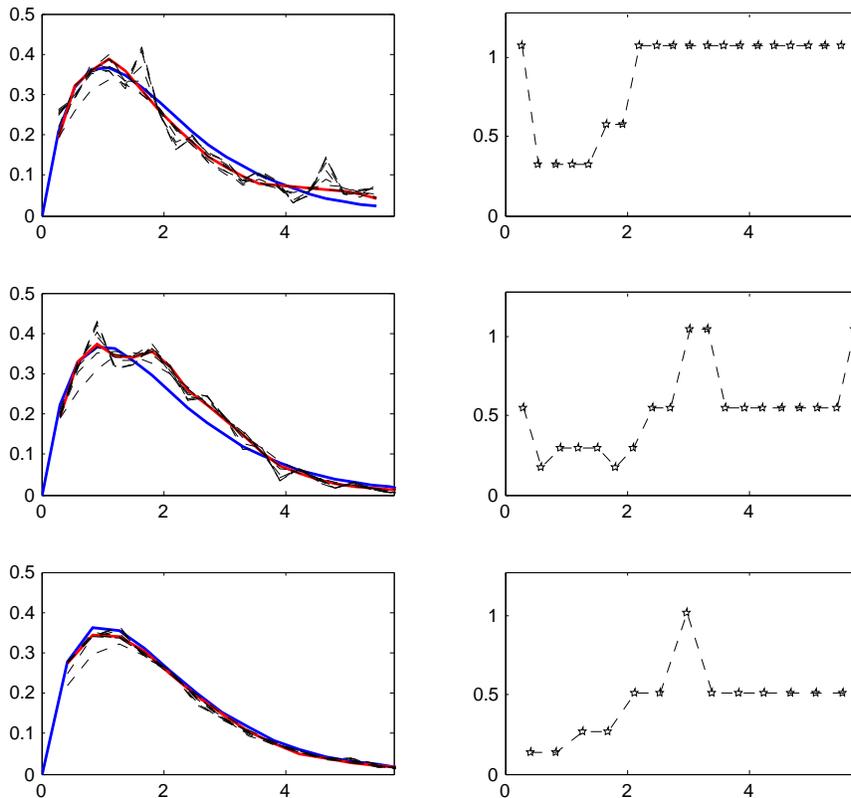


FIGURE 1. Scenario 1 with $\beta = 1$ and $n = 500$, $\bar{r}e = 1.02$, $pc = 0\%$ (top), $n = 1000$, $\bar{r}e = 1.04$, $pc = 0\%$ (middle), $n = 5000$, $\bar{r}e = 0.97$, $pc = 0\%$ (bottom)

which ensures condition (9). Lastly

$$\sum_{k: h_k \in \mathcal{H}_n} \exp(-b/h_k) = \sum_{k=1}^{\lfloor \log(n)/\log(2) \rfloor} e^{-b2^k} = O(1)$$

and (15) is verified.

Here, no minimum regularity condition of the function to estimate is needed.

5. SIMULATIONS

We illustrate the behavior of estimator $\check{\lambda}$, constructed with the pointwise bandwidth selection of Section 4.1 and conduct a Monte Carlo study to compare our adaptive procedure for the selection of the bandwidth to a bootstrap-based selection.

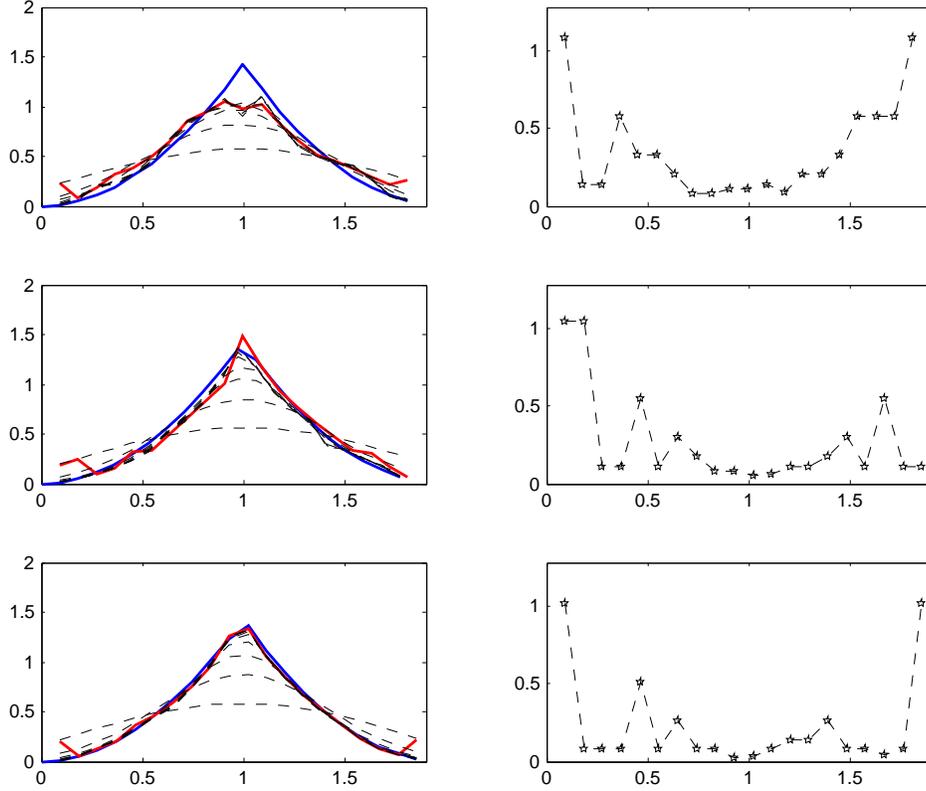


FIGURE 2. Scenario 2 with $\beta = 0.05$ and $n = 500$, $\bar{r}e = 0.92$, $pc = 0\%$ (top), $n = 1000$, $\bar{r}e = 0.89$, $pc = 0\%$ (middle), $n = 5000$, $\bar{r}e = 0.91$, $pc = 0\%$ (bottom)

5.1. Description of the simulation scheme. Recurrent events data are simulated as follows. For individuals $i = 1, \dots, n$, the terminal event D_i is simulated according to the distribution F , the censoring time C_i according to G . Conditionally on D_i , the number $n(i)$ of recurrent events experienced by individual i on time interval $[0, D_i]$ are simulated according to a Poisson distribution $\mathcal{P}(\int_0^{D_i} \varphi(u) du)$. Finally the recurrent times for individual i are simulated as $n(i)$ i.i.d. random variables with common probability density function $\varphi / \int_0^D \varphi(u) du$. The intensity of the process N^* is then given by:

$$\lambda(t) = \varphi(t)(1 - F(t)).$$

We consider two scenarios for the simulated data:

1. $\varphi(t) = t$ and $1 - F(t) = \exp(-\beta t)$.
2. $\varphi(t) = (3/2)(1 - |t - 1|)^2$ on $[0, 2]$ and $1 - F(t) = \exp(-\beta t)$ on $[0, 2]$.

The estimators of Section 4.1 are constructed with Epanechnikov kernels:

$$K_{E,2}(t) = (3/4)(1 - t^2), \text{ if } |t| \leq 1.$$

We use a data-driven criterion for the selection of the bandwidth, by replacing $V_0(h)$ in Definition (5) by:

$$\hat{V}_0(h) = \kappa_0 \frac{\hat{c}_\tau \|\hat{\lambda}\|_{\infty, \tau} \|K\|^2 \log(n)}{nh\hat{c}_G},$$

with

$$\begin{aligned} \hat{c}_\tau &= \max_{i=1, \dots, n} \left(\sup_{t \in [0, T_{\max}]} N^i(t) \right) + 2 \\ \|\hat{\lambda}\|_{\infty, \tau} &= \sup_{t \in [0, T_{\max}]} |\hat{\lambda}_{0.5}(t)| \text{ and} \\ \hat{c}_G &= 1 - \hat{G}(E_{\max} -) \text{ and} \\ \kappa_0 &= 10^{-2}, \end{aligned}$$

where E_{\max} is the greatest observed recurrent event. We set the universal value of κ_0 at 10^{-2} after an extensive simulation study: we compare the MSE for several candidate values in the range $10^{-5} - 10^2$ and in different scenarios.

The finite set of bandwidths (\mathcal{H}_n) considered in the algorithm is given by:

$$\mathcal{H}_n = \{\log^2(n)/n + 1/2^k, k = 0, 1, \dots, \lfloor \log(n)/\log(2) \rfloor\}.$$

5.2. Illustration of the behavior of the adaptive estimator. In figures 1-3, the intensity functions are estimated on a 20-point grid, regularly spaced on $[0, E_{\max}]$. The number of observations n , the mean number of recurrent $\bar{r}e$ and the level of censoring pc are reported in the captions. In each figure, the left plots show the true intensity functions in red, the estimators in blue, and the set of all the estimators proposed to the selection algorithm is dashed black. The right plots show the value of the selected windows for all points on the grid.

In Figures 1 and 2, we investigate the behavior of our estimators, when the sample size n grows. In scenario 1, where the intensity λ to recover is smooth, as in scenario 2, where λ has a singularity, the estimator behaves as expected: it improves with the sample size.

In Figure 3, we illustrate the behavior of our estimator when the censoring level grows. In this case, the censoring time has an exponential distribution, with $1 - G(t) = \exp(-\gamma t)$, where the parameter γ takes the values $\gamma = 1/30$ (top), $\gamma = 1/3$

(middle) and $\gamma = 1$ (bottom). The resulting levels of censoring and mean numbers of recurrent events are indicated in the caption. Note that, as the level of the censoring grows, the numbers of observed recurrent events vanishes (from $\bar{r}e = 1.12$, when $pc = 4\%$, to $\bar{r}e = 0.25$, when $pc = 50\%$) as does the time intervals, on which they are observed (from $[0, 9]$, when $pc = 4\%$, to $[0, 2.5]$, when $pc = 50\%$).

From a general point of view, we can see in Figures 1, 2 and 3 that the algorithm makes very different bandwidth choices, depending on the point of time. Therefore, the pointwise strategy is very useful. In particular, we can see in Figures 1 and 2 that the minimal bandwidth choice occurs at time 1 which in both cases is the location of the maximum; moreover, the selected bandwidth is smaller when the function is less smooth (look at the value of \hat{h} for the peaks of the functions). Lastly, Figure 3 shows that the pointwise strategy is relevant: indeed, it is obviously a good strategy to change the bandwidth in function of the time since none of the proposed curves would globally give a better estimate.

5.3. Monte Carlo study. In this section, we aim at comparing our adaptive strategy for the selection of the bandwidth to a bootstrap-based strategy described below, in the spirit of Chiang *et al.* [6]. Note that no theoretical results are available for the bootstrap method in our present setting. In addition Chiang *et al.* [6] in their context presents only asymptotic results and for fixed bandwidth.

Towards that end, we conduct a Monte Carlo study (with $M = 100$ replications) and calculate the following squared error on the simulated data of replication m (for $m = 1, \dots, M$):

$$SE(\hat{\lambda}^m) = \frac{1}{K - 2\lfloor K\rho \rfloor} \sum_{k=1+\lfloor K\rho \rfloor}^{K-\lfloor K\rho \rfloor} (\hat{\lambda}_{\eta(t_k)}^m(t_k) - \lambda(t_k))^2,$$

where $\hat{\lambda}_{\eta(t_k)}^m$ is the estimator $\hat{\lambda}$ calculated on the m th dataset in the Monte Carlo experiment, at point t_k (on a K point grid) and for the selected pointwise bandwidth $\eta(t_k)$. In the previous equation, $0 \leq \rho < 1$ represents the proportion of the smallest and largest observations withdrawn in the computation of SE to avoid the boundary effects.

We consider

- the adaptive bandwidth selection procedure described in the previous subsection and will denote by $SE(\hat{\lambda}_{Adapt}^m)$ its squared error: in this case $\hat{\lambda}_{\eta(t_k)}^m = \hat{\lambda}_{\hat{h}(t_k)}^m$, where $\hat{h}(t_k)$ is defined in Equation 7.

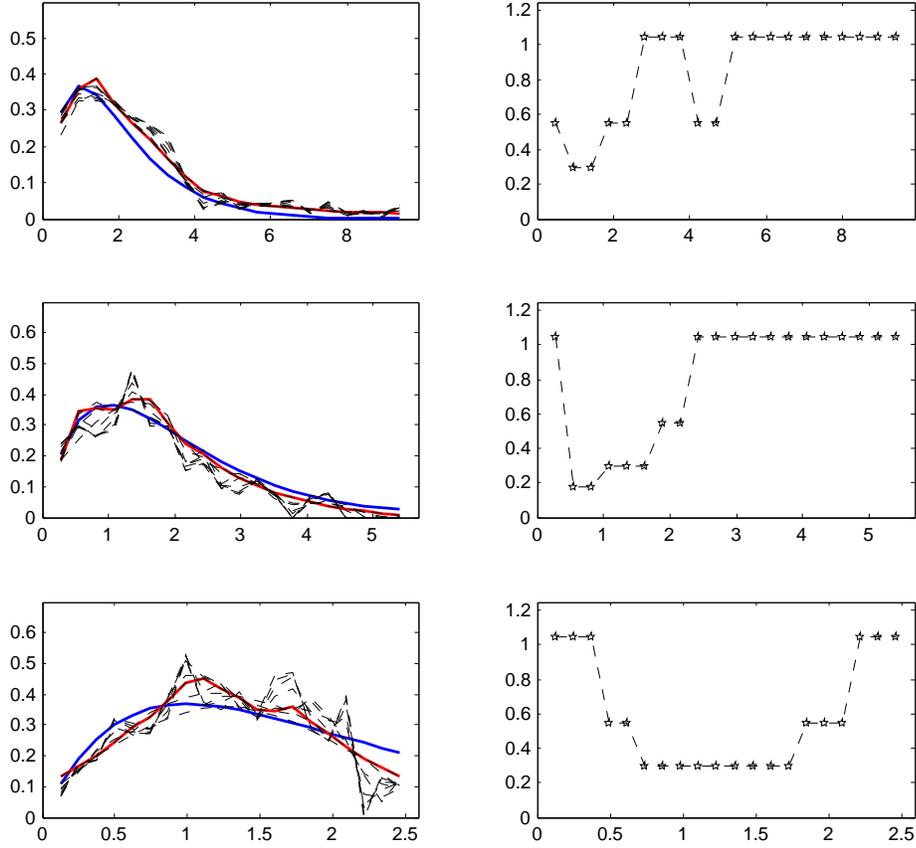


FIGURE 3. Scenario 1 with $\beta = 1$ and $n = 1000$, $\bar{r}e = 1.12$, $pc = 4\%$ (top), $n = 1000$, $\bar{r}e = 0.55$, $pc = 25\%$ (middle), $n = 1000$, $\bar{r}e = 0.25$, $pc = 50\%$ (bottom)

- a bootstrap-based procedure and will denote by $SE(\hat{\lambda}_{Boot}^m)$ its squared error: in this case $\hat{\lambda}_{\eta(t_k)}^m = \hat{\lambda}_{\hat{h}^{m,*}(t_k)}^m$, where $\hat{h}^{m,*}(t_k)$ is defined by:

$$\hat{h}^{m,*}(t_k) = \underset{h \in \mathcal{H}_n}{\operatorname{argmin}} \widehat{MSE}^*(\hat{\lambda}_h^m(t_k))$$

$\times 10^3$	$n = 200$		$n = 500$		$n = 1000$	
	mean	median	mean	median	mean	median
Adaptive	1.72	1.44	0.71	0.50	0.31	0.21
Bootstrap	1.76	1.57	0.73	0.55	0.37	0.28

TABLE 1. Scenario 1 with 0% of censoring and $\beta = 1$

$\times 10^3$	$n = 200$		$n = 500$		$n = 1000$	
	mean	median	mean	median	mean	median
Adaptive	4.25	3.24	2.05	1.56	1.26	1.10
Bootstrap	4.30	3.24	1.95	1.56	0.95	0.73

TABLE 2. Scenario 1 with $\sim 33\%$ of censoring and $\beta = 1$

with

$$\widehat{MSE}^*(\hat{\lambda}_h^m(t_k)) = \widehat{Var}^*(\hat{\lambda}_h^m(t_k)) + \left(\widehat{Bias}^*(\hat{\lambda}_h^m(t_k))\right)^2.$$

The term $\widehat{Var}^*(\hat{\lambda}_h^m(t_k))$ is the estimated variance calculated on B samples bootstrapped from the m th dataset in the Monte Carlo experiment and

$$\widehat{Bias}^*(\hat{\lambda}_h^m(t)) = \frac{1}{nh} \sum_{i=1}^n \int \left\{ K_{E,2} \left(\frac{t-s}{h} \right) - K_{E,4} \left(\frac{t-s}{h} \right) \right\} \frac{dN_i^m(s)}{1 - \hat{G}^m(s-)}$$

where N_i^m and \hat{G}^m are calculated on the m th Monte Carlo experiment and $K_{E,4} = (15/8) \times (1 - (7/2)u^2)K_{E,2}(u)$, see Chiang *et al.* [6] and Schucany [19] for the estimation of the bias and Hansen [13] for the definition of $K_{E,4}$.

In tables 1 to 3, we display the mean and the median of $SE(\hat{\lambda}_{Adapt}^m)$ and $SE(\hat{\lambda}_{Boot}^m)$ obtained on $M = 100$ Monte Carlo experiments. The number of bootstrap sample $B = 100$ and $\tau = 0.1$. The variances and interquartile ranges are roughly the same for both methods. Note that the grid of bandwidths proposed in the bootstrap algorithm is borrowed from our theoretical proposal.

We can see from tables 1-3 that the performances of the two bandwidth selection methods are roughly similar: our proposal is slightly better in table 1 while the bootstrap method performs slightly better in table 3. This is certainly partly due to the choice of the set of bandwidths \mathcal{H}_n which sets the bootstrap under control. Let also emphasize that the number of kernel estimators that are computed are $|\mathcal{H}_n|^2$ for

$\times 10^2$	$n = 200$		$n = 500$		$n = 1000$	
	mean	median	mean	median	mean	median
Adaptive	2.25	2.28	1.55	1.56	0.03	0.03
Bootstrap	1.69	1.60	0.99	0.95	0.02	0.03

TABLE 3. Scenario 2 with 0% of censoring and $\beta = 0.05$

our method versus $B|\mathcal{H}_n|$ for the bootstrap method, where $B = 100$ and $|\mathcal{H}_n| \simeq 10$ for $n = 1000$, which makes our method approximately 10 times faster.

6. CONCLUDING REMARKS

In this work, we not only provide a kernel estimator for the intensity function of a recurrent event process, but we also prove oracle type inequalities for the risk of an adaptive estimator with data-driven selected bandwidth. We have studied both cases of pointwise risk for pointwise chosen bandwidth and integrated global risk with a globally selected bandwidth. Our bandwidth selection proposal is original and slightly different from standard cross-validation methods. This is because it is based on recent ideas developed by Goldenshluger and Lepski [12]: in this sense, our results are new and the way of proving the results is of interest. We also assess the practical feasibility and the good performances of our proposal through a short simulation study: we found it more challenging to evaluate the pointwise selection and we illustrated the different bandwidths choices performed by the algorithm.

7. PROOFS

7.1. Proof of Lemma 1. The proof relies on four additional lemmas which are presented below. First, write:

$$\hat{\lambda}_h(t) - \tilde{\lambda}_h(t) = \frac{1}{nh} \sum_{i=1}^n \int \frac{\hat{G}(s-) - G(s-)}{(1 - \hat{G}(s-))(1 - G(s-))} K\left(\frac{t-s}{h}\right) dN_i(s).$$

Then introduce the sets

$$\begin{aligned} \Omega_G &= \left\{ \omega : \forall t \in [0, \tau], G(t) - \hat{G}(t) \geq -c_G/2 \right\}, \\ \Omega_G^* &= \left\{ \omega : \forall t \in [0, \tau], |G(t) - \hat{G}(t)| \leq c_0 \sqrt{n^{-1} \log n} \right\}, \end{aligned}$$

and

$$(17) \quad \Omega_{c_0} = \Omega_G \cap \Omega_G^*.$$

Our idea is to study the difference process $\hat{\lambda}_h - \tilde{\lambda}_h$ on Ω_{c_0} and its complementary. The next lemma gives a useful bound of $\mathbb{P}[\Omega_{c_0}^c]$. The proof is postponed to Section 8.

Lemma 2. *For all $p \in \mathbb{N}$, there exists a choice of the constant $c_0 = c_0(p)$ such that,*

$$(18) \quad \mathbb{P}[\Omega_{c_0(p)}^c] \leq c_2 n^{-p},$$

where c_2 is a constant depending on p , c_F and c_G , and $c_0(p)$ also depends on c_F .

In the following, we denote by $\Omega_p = \Omega_{c_0(p)}$ such that Equation (18) in Lemma 2 holds. We now start the proof of Lemma 1 by studying the difference process $\hat{\lambda}_h - \tilde{\lambda}_h$ on the set Ω_p^c .

Lemma 3. *Under Assumptions 1 to 4, for all $p \in \mathbb{N}$, $t \in [h, \tau - h]$, we have:*

$$\mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p^c) \right] \leq (n+1)^2 n^{2-p/2} c_3 (\|K\|_\infty)^2,$$

where

$$c_3 = c_\tau^{3/2} \sqrt{c_2} \left(\int_0^\tau \frac{\lambda(s) ds}{(1 - G(s-))^3} \right)^{1/2}.$$

Consequently, choosing $p \geq 10$ yields $\mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p^c) \right] \leq c/n$ for a positive constant c .

Lemma 4. *Under Assumptions 1 to 4, for all $p \in \mathbb{N}$, we have:*

$$\int_h^{\tau-h} \mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p^c) \right] dt \leq (n+1)^2 n^{1-p/2} c_3 \|K\|^2.$$

Consequently, choosing $p \geq 8$ yields $\int_h^{\tau-h} \mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p^c) \right] dt \leq c/n$ for a positive constant c .

Proof of Lemmas 3 and 4. From the facts that $1 - \hat{G}(t) \geq (n+1)^{-1}$ (see Equation (3)) and $\|\hat{G} - G\|_\infty < 1$, we have for all $t \in [h, \tau - h]$:

$$(19) \quad \begin{aligned} \mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p^c) \right] &\leq \frac{(n+1)^2}{n^2} \mathbb{E} \left[\left(\sum_{i=1}^n \int \frac{K_h(t-s)}{1-G(s-)} dN_i(s) \right)^2 I(\Omega_p^c) \right] \\ &\leq (n+1)^2 \mathbb{E} \left[\left(\int \frac{K_h(t-s)}{1-G(s-)} dN(s) \right)^2 I(\Omega_p^c) \right] \\ &\leq (n+1)^2 c_\tau \mathbb{E} \left[\int \frac{K_h^2(t-s) I(\Omega_p^c)}{(1-G(s-))^2} dN(s) \right], \end{aligned}$$

where the last inequality is obtained from Lemma 9. Now, for the proof of Lemma 3, use consecutively the Cauchy-Schwarz inequality and Lemma 9 to obtain:

$$\begin{aligned} \mathbb{E} \left[\int \frac{K_h^2(t-s) I(\Omega_p^c)}{(1-G(s-))^2} dN(s) \right] &\leq \mathbb{E}^{1/2} \left[\left(\int \frac{K_h^2(t-s)}{(1-G(s-))^2} dN(s) \right)^2 \right] \sqrt{\mathbb{P}[\Omega_p^c]} \\ &\leq (\|K\|_\infty)^2 h^{-2} \sqrt{c_\tau} \mathbb{E}^{1/2} \left[\int_0^\tau \frac{dN(s)}{(1-G(s-))^4} \right] \sqrt{\mathbb{P}[\Omega_p^c]} \\ &\leq (\|K\|_\infty)^2 h^{-2} n^{-p/2} \sqrt{c_2 c_\tau} \left(\int_0^\tau \frac{\lambda(s) ds}{(1-G(s-))^3} \right)^{1/2}, \end{aligned}$$

and conclude the proof using the fact that $h^{-1} \leq n$. To prove Lemma 4 write,

$$\int_h^{\tau-h} \int \frac{K_h^2(t-s)}{(1-G(s-))^2} dN(s) dt \leq h^{-1} \|K\|^2 \int_0^\tau \frac{dN(s)}{(1-G(s-))^2}.$$

Then, using Cauchy-Schwarz inequality, we get from inequality (19):

$$\begin{aligned} \int_h^{\tau-h} \mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p^c) \right] dt &\leq \frac{(n+1)^2 c_\tau}{h} \|K\|^2 \mathbb{E} \left[\int_0^\tau \frac{I(\Omega_p^c) dN(s)}{(1-G(s-))^2} \right] \\ &\leq \frac{(n+1)^2}{h} c_\tau \|K\|^2 \mathbb{E}^{1/2} \left[\left(\int_0^\tau \frac{dN(s)}{(1-G(s-))^2} \right)^2 \right] \sqrt{\mathbb{P}[\Omega_p^c]} \\ &\leq \frac{(n+1)^2 n^{-p/2}}{h} c_\tau^{3/2} \sqrt{c_2} \|K\|^2 \left(\int_0^\tau \frac{\lambda(s) ds}{(1-G(s-))^3} \right)^{1/2}, \end{aligned}$$

and again, we conclude the proof using the fact that $h^{-1} \leq n$. \square

We now study the difference process of $\hat{\lambda}_h - \tilde{\lambda}_h$ on Ω_p .

Lemma 5. *Under Assumptions 1 to 4, we have for all $t \in [h, \tau - h]$ and any $p \in \mathbb{N}$,*

$$\mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p) \right] \leq \frac{c_4 \log n}{n} \|\lambda\|_{\infty, \tau} \left\{ (\|K\|_1)^2 \|\lambda\|_{\infty, \tau} + \frac{c_\tau \|K\|^2}{c_G n h} \right\},$$

where $c_4 = 4c_0^2 c_G^{-2}$ and $c_0 = c_0(p)$.

Consequently, for $t \in [h, \tau - h]$, we have

$$\mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p) \right] \leq \frac{c \log(n)}{n},$$

where c is a positive constant.

Lemma 6. *Under Assumptions 1 to 4, we have, for any $p \in \mathbb{N}$*

$$\int_h^{\tau-h} \mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p) \right] dt \leq \frac{c_4 \log n}{n} \|K\|^2 \left\{ 2 \int_0^\tau \lambda^2(t) dt + \frac{c_\tau \Lambda(\tau)}{n h} \right\},$$

where $\Lambda(\tau)$ is defined in Theorem 1. Consequently, we have

$$\int_h^{\tau-h} \mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p) \right] dt \leq \frac{c \log(n)}{n},$$

where c is a positive constant.

Proof of Lemmas 5 and 6. First, use the facts that

- $1 - \hat{G}(t) = 1 - G(t) + G(t) - \hat{G}(t) \geq c_G/2$ on Ω_G ,
- $\|G(t) - \hat{G}(t)\|_\infty \leq c_0 \sqrt{n^{-1} \log n}$ on Ω_G^* ,

to write:

$$\mathbb{E} \left[(\hat{\lambda}_h(t) - \tilde{\lambda}_h(t))^2 I(\Omega_p) \right] \leq \frac{4c_0^2 \log n}{nc_G^2} \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n \int \frac{|K_h(t-s)|}{1-G(s-)} dN_i(s) \right)^2 \right].$$

Then, we have:

$$\begin{aligned} \left(\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \int \frac{|K_h(t-s)|}{1-G(s-)} dN_i(s) \right] \right)^2 &= \left(\int |K_h(t-s)| \lambda(s) ds \right)^2 \\ &\leq (\|K\|_1 \|\lambda\|_{\infty, \tau})^2, \end{aligned}$$

and

$$\mathbb{V} \left[\frac{1}{n} \sum_{i=1}^n \int \frac{|K_h(t-s)|}{1-G(s-)} dN_i(s) \right] \leq \frac{c_\tau \|K\|^2 \|\lambda\|_{\infty, \tau}}{c_G n h},$$

which follows from Proposition 1. Combining these two bounds gives the final result of Lemma 5.

The proof of Lemma 6 follows the same line. From a change of variables and the Cauchy-Schwarz inequality we have:

$$\begin{aligned} \int_h^{\tau-h} \left(\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \int \frac{|K_h(t-s)|}{1-G(s-)} dN_i(s) \right] \right)^2 dt &= \int_h^{\tau-h} \left(\int |K_h(t-s)| \lambda(s) ds \right)^2 dt \\ &\leq \left(\int_{-1}^1 K^2(u) du \right) \int_h^{\tau-h} \int_{-1}^1 \lambda^2(t-uh) du dt \\ &\leq 2 \|K\|^2 \int_0^\tau \lambda^2(t) dt, \end{aligned}$$

where the last inequality is obtained with an other change of variables. On the other hand, from similar arguments as in the proof of Proposition 1, we have

$$\int_h^{\tau-h} \mathbb{V} \left[\frac{1}{n} \sum_{i=1}^n \int \frac{|K_h(t-s)|}{1-G(s-)} dN_i(s) \right] dt \leq \frac{c_\tau \Lambda(\tau)}{nh} \|K\|^2,$$

and the result follows. \square

Gathering the results of Lemmas 3 to 6 imply the result of Lemma 1. \square

7.2. Proof of Theorem 2. First, for all $h \in \mathcal{H}_n$, the following sequence of inequalities holds:

$$\begin{aligned} (\check{\lambda}(t_0) - \lambda(t_0))^2 &\leq 3(\hat{\lambda}_{\hat{h}(t_0)}(t_0) - \hat{\lambda}_{h, \hat{h}(t_0)}(t_0))^2 + 3(\hat{\lambda}_{h, \hat{h}(t_0)}(t_0) - \hat{\lambda}_h(t_0))^2 + 3(\hat{\lambda}_h(t_0) - \lambda(t_0))^2 \\ &\leq 3(A_0(h, t_0) + V_0(\hat{h}(t_0))) + 3(A_0(\hat{h}(t_0), t_0) + V_0(h)) + 3(\hat{\lambda}_h(t_0) - \lambda(t_0))^2 \\ &\leq 6A_0(h, t_0) + 6V_0(h) + 3(\hat{\lambda}_h(t_0) - \lambda(t_0))^2. \end{aligned}$$

Since $V_0(h)$, see (5), and $(\hat{\lambda}_h(t_0) - \lambda(t_0))^2$, see Theorem 1, (a), have the adequate order (with additional $\log(n)$ for V_0), we only study $A_0(h, t_0)$. With obvious definition of $\tilde{\lambda}_{h, h'} = K_{h'} * \tilde{\lambda}_h$, $\lambda_h(t_0) = \mathbb{E}[\tilde{\lambda}_h(t_0)]$ and $\lambda_{h, h'}(t_0) = \mathbb{E}[\tilde{\lambda}_{h, h'}(t_0)]$, $A_0(h, t_0)$ can be decomposed into five terms:

$$\begin{aligned} A_0(h, t_0) &= \sup_{h' \in \mathcal{H}_n} \left\{ (\hat{\lambda}_{h'}(t_0) - \hat{\lambda}_{h, h'}(t_0))^2 - V_0(h') \right\}_+ \\ &\leq 5 \sup_{h' \in \mathcal{H}_n} \left\{ (\tilde{\lambda}_{h'}(t_0) - \lambda_{h'}(t_0))^2 - V_0(h')/10 \right\}_+ \\ &\quad + 5 \sup_{h' \in \mathcal{H}_n} \left\{ (\tilde{\lambda}_{h, h'}(t_0) - \lambda_{h, h'}(t_0))^2 - V_0(h')/10 \right\}_+ \\ &\quad + 5 \sup_{h' \in \mathcal{H}_n} (\hat{\lambda}_{h'}(t_0) - \tilde{\lambda}_{h'}(t_0))^2 + 5 \sup_{h' \in \mathcal{H}_n} (\hat{\lambda}_{h, h'}(t_0) - \tilde{\lambda}_{h, h'}(t_0))^2 \\ &\quad + 5 \sup_{h' \in \mathcal{H}_n} (\lambda_{h'}(t_0) - \lambda_{h, h'}(t_0))^2 \\ &:= 5(T_{0,1} + T_{0,2} + T_{0,3} + T_{0,4} + T_{0,5}). \end{aligned}$$

We start with the last one:

$$\begin{aligned} |\lambda_{h'}(t_0) - \lambda_{h, h'}(t_0)| &= |K_{h'} * \lambda(t_0) - K_{h'} * K_h * \lambda(t_0)| = |K_{h'} * (\lambda - K_h * \lambda)(t_0)| \\ &\leq \|K\|_1 \sup_{t \in [0, \tau]} |(\lambda - K_h * \lambda)(t)|. \end{aligned}$$

This yields to say

$$T_{0,5} \leq \|K\|_1^2 \|\lambda - K_h * \lambda\|_{\infty, \tau}^2 \leq (\|K\|_1)^2 c_1^2 h^{2\beta},$$

since $\lambda - K_h * \lambda$ corresponds to the bias term in Proposition 1.

Then we decompose $T_{0,3}$ into two terms corresponding to $I(\Omega_p)$ and $I(\Omega_p^c)$ where Ω_p is defined by (17). First, from Lemma 3, we have

$$\begin{aligned} \mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} (\hat{\lambda}_{h'} - \tilde{\lambda}_{h'})^2(t_0) I(\Omega_p^c) \right] &\leq \sum_{k, h_k \in \mathcal{H}_n} \mathbb{E} \left[(\hat{\lambda}_{h_k} - \tilde{\lambda}_{h_k})^2(t_0) I(\Omega_p^c) \right] \\ &\leq \sum_{k, h_k \in \mathcal{H}_n} 4c_3 (\|K\|_\infty)^2 n^{4-p/2} \leq 4c_3 (\|K\|_\infty)^2 n^{5-p/2}, \end{aligned}$$

using the fact that $\text{Card}(\mathcal{H}_n) \leq n$. Consequently, this term is of order $1/n$ as soon as $p \geq 12$. On the other hand, the following sequence of inequalities holds:

$$\begin{aligned} &\mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} (\hat{\lambda}_{h'} - \tilde{\lambda}_{h'})^2(t_0) I(\Omega_p) \right] \\ &\leq \frac{4c_0^2 \log(n)}{c_G^2 n} \mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} \left(\int \frac{|K_{h'}(t_0 - s)|}{1 - G(s-)} \left(\frac{1}{n} \sum_{i=1}^n dN_i(s) \right) \right)^2 \right] \\ &\leq \frac{8c_0^2 \log(n)}{c_G^2 n} \mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} \left(\int \frac{|K_{h'}(t_0 - s)|}{1 - G(s-)} \left(\frac{1}{n} \sum_{i=1}^n dN_i(s) - \lambda(s)(1 - G(s-)) ds \right) \right)^2 \right] \\ &\quad + \frac{8c_0^2 \log(n)}{c_G^2 n} \sup_{h' \in \mathcal{H}_n} \left(\int |K_{h'}(t_0 - s)| \lambda(s) ds \right)^2 \\ &\leq \frac{8c_0^2 \log(n)}{c_G^2 n} \sum_{k, h_k \in \mathcal{H}_n} \mathbb{V} \left[\frac{1}{n} \sum_{i=1}^n \int \frac{|K_{h_k}(t_0 - s)|}{1 - G(s-)} dN_i(s) \right] + \frac{8c_0^2 \|\lambda\|_{\infty, \tau}^2 \log(n)}{c_G^2 n} \|K\|_1^2 \\ (20) \quad &\leq \frac{8c_0^2 \log(n)}{c_G^3 n} \sum_{k, h_k \in \mathcal{H}_n} \frac{c_\tau \|\lambda\|_{\infty, \tau} \|K\|^2}{nh_k} + \frac{8c_0^2 (2\|\lambda\|_{\infty, \tau})^2 \log(n)}{c_G^2 n} \|K\|_1^2, \end{aligned}$$

where the bound on the variance term comes from the proof of Proposition 1. Therefore $\mathbb{E}[T_{0,3}] \lesssim \log^{1+a}(n)/n$ from Condition (9) and this ends the study of $T_{0,3}$.

The term $T_{0,4}$ can be handled in a similar way using the relation $\hat{\lambda}_{h, h'}(t_0) - \tilde{\lambda}_{h, h'}(t_0) = K_{h'} * (\hat{\lambda}_h - \tilde{\lambda}_h)(t_0)$. Indeed,

$$\begin{aligned} \mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} (\hat{\lambda}_{h, h'} - \tilde{\lambda}_{h, h'})^2(t_0) I(\Omega_p^c) \right] &= \mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} \left(K_{h'} * (\hat{\lambda}_h - \tilde{\lambda}_h) \right)^2(t_0) I(\Omega_p^c) \right] \\ &\leq (\|K\|_1)^2 \mathbb{E} \left[\|\hat{\lambda}_h - \tilde{\lambda}_h\|_{\infty, \tau}^2 I(\Omega_p^c) \right] \\ &\leq 4c_3 (\|K\|_1 \|K\|_\infty)^2 n^{4-p/2}, \end{aligned}$$

from Lemma 3 and

$$\begin{aligned} & \mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} (\hat{\lambda}_{h,h'} - \tilde{\lambda}_{h,h'})^2(t_0) I(\Omega_p) \right] \\ & \leq \frac{8c_0^2 \log(n)}{c_G^2 n} \sum_{k, h_k \in \mathcal{H}_n} \mathbb{V} \left[\frac{1}{n} \sum_{i=1}^n \int \frac{|K_{h_k} * K_h(t_0 - s)|}{1 - G(s-)} dN_i(s) \right] \\ & \quad + \frac{8c_0^2}{c_G^2} \|\lambda\|_{\infty, \tau} \frac{\log(n)}{n} \left(\sup_{h' \in \mathcal{H}_n} \|K_{h'} * K_h\|_1 \right)^2. \end{aligned}$$

Then, using the property

$$(21) \quad \|f * g\|_q \leq \|f\|_1 \|g\|_q \quad \text{for } q \geq 1,$$

it is easy to see that

$$\begin{aligned} \mathbb{V} \left[\frac{1}{n} \sum_{i=1}^n \int \frac{|K_{h_k} * K_h(t_0 - s)|}{1 - G(s-)} dN_i(s) \right] & \leq \frac{c_\tau}{nc_G} \|\lambda\|_{\infty, \tau} \|K_h * K_{h_k}\|^2 \\ & \leq \frac{c_\tau}{nc_G} \|\lambda\|_{\infty, \tau} (\|K_h\|_1)^2 \|K_{h_k}\|^2 \\ & \leq \frac{c_\tau \|\lambda\|_{\infty, \tau} (\|K\|_1)^2 \|K\|^2}{nc_G h_k}, \end{aligned}$$

and

$$(\|K_{h'} * K_h\|_1)^2 \leq (\|K_{h'}\|_1 \|K_h\|_1)^2 = \|K\|_1^4.$$

We conclude as previously that $\mathbb{E}[T_{0,4}] \lesssim \log^{1+a}(n)/n$.

Finally, let us study the terms $T_{0,1}$ and $T_{0,2}$. We start by recalling the following concentration inequality.

Lemma 7. *[Bernstein inequality] Let ξ_1, \dots, ξ_n be independent and identically distributed random variables and $S_n(\xi) = \sum_{i=1}^n \xi_i$. Then, for $\eta > 0$,*

$$(22) \quad \mathbb{P}(|S_n(\xi) - \mathbb{E}[S_n(\xi)]| \geq n\eta) \leq 2 \max \left(\exp \left(-\frac{n\eta^2}{4w} \right), \exp \left(-\frac{n\eta}{4b} \right) \right),$$

where w and b are such that $|\xi_1| \leq b$ almost surely and $\mathbb{V}(\xi_1) \leq w$.

Now, we want to apply this result to $\xi_i = \int K_h(t_0 - s) dN_i(s) / (1 - G(s-))$. First, we need to establish the values of the bounds b and w . We have

$$|\xi_1| \leq (c_\tau \|K\|_\infty) / (c_G h) := b \quad \text{and} \quad \mathbb{V}(\xi_1) \leq c_\tau \|\lambda\|_{\infty, \tau} \|K\|^2 / (c_G h) := w.$$

Thus, Inequality (22) can be written in the following way: for some $x > 0$,

$$\begin{aligned} & \mathbb{P} \left[|\tilde{\lambda}_h(t_0) - \lambda_h(t_0)| \geq \sqrt{V_0(h)/10 + x} \right] \\ & \leq 2 \max \left(\exp(-n(V_0(h)/10 + x)/(4w)), \exp(-n\sqrt{V_0(h)/10 + x}/(4b)) \right) \\ & \leq 2 \max \left(\exp(-n(V_0(h)/10 + x)/(4w)), \exp(-n\sqrt{V_0(h)/5}/(8b)) \exp(-n\sqrt{x/2}/(4b)) \right). \end{aligned}$$

Then, we set $\kappa_0 \geq 80$, in order to have

$$\frac{nV_0(h)}{40w} = (\kappa_0/40) \log(n) \geq 2 \log(n).$$

On the other hand,

$$\frac{n\sqrt{V_0(h)}}{8b\sqrt{5}} = \frac{\|K\| \sqrt{c_G \kappa_0} \|\lambda\|_{\infty, \tau}}{8\|K\|_{\infty} \sqrt{5} c_{\tau}} \sqrt{nh \log(n)} := \kappa_2 \sqrt{nh \log(n)}.$$

Then taking $\kappa_1 \geq 4\kappa_2^{-2}$ in Condition (8) gives,

$$\frac{n\sqrt{V_0(h)}}{8b\sqrt{5}} \geq 2 \log(n).$$

Therefore, we have

$$\mathbb{P} \left[|\tilde{\lambda}_h(t_0) - \lambda_h(t_0)| \geq \sqrt{V_0(h)/10 + x} \right] \leq 2n^{-2} \max \left(e^{-\kappa_3 nhx}, e^{-\kappa_4 nh\sqrt{x}} \right),$$

where

$$\kappa_3 = \frac{c_G}{4c_{\tau} \|\lambda\|_{\infty, \tau} \|K\|^2} \quad \text{and} \quad \kappa_4 = \frac{c_G}{4c_{\tau} \|K\|_{\infty} \sqrt{2}}.$$

This yields

$$\begin{aligned} \mathbb{E} \left[\left\{ |\tilde{\lambda}_h(t_0) - \lambda_h(t_0)|^2 - V_0(h)/10 \right\}_+ \right] & \leq \int_0^{+\infty} \mathbb{P} \left[|\tilde{\lambda}_h(t_0) - \lambda_h(t_0)| \geq \sqrt{V_0(h)/10 + x} \right] dx \\ & \leq 2n^{-2} \max \left(\int_0^{+\infty} e^{-\kappa_3 nhx} dx, \int_0^{+\infty} e^{-\kappa_4 nh\sqrt{x}} dx \right) \\ & \leq 2n^{-2} \max \left(\frac{1}{\kappa_3 nh}, \frac{2}{\kappa_4^2 (nh)^2} \right) \leq \kappa_5 n^{-2}, \end{aligned}$$

for some positive constant κ_5 . Finally,

$$\begin{aligned} \mathbb{E}[T_{0,1}] & = \mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} \left\{ (\tilde{\lambda}_{h'} - \lambda_{h'})^2(t_0) - V_0(h')/10 \right\}_+ \right] \\ & \leq \sum_{k, h_k \in \mathcal{H}_n} \mathbb{E} \left[\left\{ (\tilde{\lambda}_{h_k} - \lambda_{h_k})^2(t_0) - V_0(h_k)/10 \right\}_+ \right] \\ & \leq \kappa_5 \text{Card}(\mathcal{H}_n) n^{-2}, \end{aligned}$$

and since $\text{Card}(\mathcal{H}_n) \leq n$, we conclude that $\mathbb{E}[T_{0,1}] \lesssim n^{-1}$.

The last term is $T_{0,2}$ which can be treated in a similar way. Write

$$\begin{aligned} \mathbb{E}[T_{0,2}] &= \mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} \left\{ (\tilde{\lambda}_{h,h'} - \lambda_{h,h'})^2(t_0) - V_0(h')/10 \right\}_+ \right] \\ &\leq \sum_{k, h_k \in \mathcal{H}_n} \mathbb{E} \left[\left\{ (\tilde{\lambda}_{h,h_k} - \lambda_{h,h_k})^2(t_0) - V_0(h_k)/10 \right\}_+ \right]. \end{aligned}$$

Then the sequel is the same as for the proof of $T_{0,1}$ except that all h vanish because $\|K_h * K_{h'}\|_\infty \leq \|K_{h'}\|_\infty \|K\|_1$.

Gathering the bounds of the five terms gives the result of Theorem 2. \square

7.3. Proof of Theorem 3. Following the lines of the proof of Theorem 2, we have, for all $h \in \mathcal{H}_n$,

$$\begin{aligned} \|\lambda^* - \lambda\|^2 &\leq 3\|\hat{\lambda}_{\hat{h}} - \hat{\lambda}_{h,\hat{h}}\|^2 + 3\|\hat{\lambda}_{h,\hat{h}} - \hat{\lambda}_h\|^2 + 3\|\hat{\lambda}_h - \lambda\|^2 \\ &\leq 3(A(h) + V(\hat{h})) + 3(A(\hat{h}) + V(h)) + 3\|\hat{\lambda}_h - \lambda\|^2 \\ &\leq 6A(h) + 6V(h) + 3\|\hat{\lambda}_h - \lambda\|^2. \end{aligned}$$

Here again, $V(h)$ and $\|\hat{\lambda}_h - \lambda\|^2$ (see Theorem 1, (b)) have the adequate order and we only need to study $A(h)$. Recall that $\tilde{\lambda}_{h,h'} = K_{h'} * \tilde{\lambda}_h$, $\lambda_h(t) = \mathbb{E}[\tilde{\lambda}_h(t)]$, $\lambda_{h,h'}(t) = \mathbb{E}[\tilde{\lambda}_{h,h'}(t)]$ and write:

$$\begin{aligned} A(h) &= \sup_{h' \in \mathcal{H}_n} \left\{ \|\hat{\lambda}_{h'} - \hat{\lambda}_{h,h'}\|^2 - V(h') \right\}_+ \\ &\leq 5 \sup_{h' \in \mathcal{H}_n} \left\{ \|\tilde{\lambda}_{h'} - \lambda_{h'}\|^2 - V(h')/10 \right\}_+ + 5 \sup_{h' \in \mathcal{H}_n} \left\{ \|\tilde{\lambda}_{h,h'} - \lambda_{h,h'}\|^2 - V(h')/10 \right\}_+ \\ &\quad + 5 \sup_{h' \in \mathcal{H}_n} \|\hat{\lambda}_{h'} - \tilde{\lambda}_{h'}\|^2 + 5 \sup_{h' \in \mathcal{H}_n} \|\hat{\lambda}_{h,h'} - \tilde{\lambda}_{h,h'}\|^2 + 5 \sup_{h' \in \mathcal{H}_n} \|\lambda_{h'} - \lambda_{h,h'}\|^2 \\ &:= 5(T_1 + T_2 + T_3 + T_4 + T_5). \end{aligned}$$

We start with T_5 :

$$\|\lambda_{h'} - \lambda_{h,h'}\|^2 = \|K_{h'} * (\lambda - K_h * \lambda)\|^2 \leq (\|K_{h'}\|_1)^2 \|\lambda - K_h * \lambda\|^2,$$

where we used the property (21) with $q = 2$. This yields to

$$T_5 \leq (\|K\|_1)^2 \tau c_1^2 h^{2\beta},$$

since $\|\lambda - K_h * \lambda\|$ corresponds to the bias term in Proposition 1.

Now, the same kind of arguments can be applied to T_4 :

$$\hat{\lambda}_{h,h'} - \tilde{\lambda}_{h,h'} = K_{h'} * (\hat{\lambda}_h - \tilde{\lambda}_h),$$

and so,

$$\mathbb{E}[T_4] \leq \|K\|_1^2 \mathbb{E} \left[\|\hat{\lambda}_h - \tilde{\lambda}_h\|^2 \right] \leq c' (\|K\|_1)^2 \log(n)/n,$$

where the last inequality was obtained from Lemma 1.

The term T_3 can be dealt with in the same way as $T_{0,3}$ in the proof of Theorem 2. First, from Lemma 4,

$$\begin{aligned} \mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} \int (\hat{\lambda}_{h'} - \tilde{\lambda}_{h'})^2(t) I(\Omega_p^c) dt \right] &\leq \sum_{j, h_j \in \mathcal{H}_n} \int \mathbb{E}[(\hat{\lambda}_{h_j} - \tilde{\lambda}_{h_j})^2(t) I(\Omega_p^c)] dt \\ &\leq \sum_{j, h_j \in \mathcal{H}_n} 4c_3 \|K\|^2 n^{3-p/2} \leq 4c_3 \|K\|^2 n^{4-k/2}, \end{aligned}$$

and this term is of order $1/n$ as long as $p \geq 10$. Then, using similar inequalities as in (20) yields

$$\begin{aligned} &\mathbb{E} \left[\sup_{h' \in \mathcal{H}_n} \int_h^{\tau-h} (\hat{\lambda}_{h'} - \tilde{\lambda}_{h'})^2(t) I(\Omega_p) dt \right] \\ &\leq \frac{8c_0^2 \log(n)}{c_G^2 n} \sum_{k, h_k \in \mathcal{H}_n} \frac{c_\tau \Lambda(\tau) \|K\|^2}{nh_k} + \frac{16c_0^2 \log(n)}{c_G^2 n} \|K\|^2 \int_0^\tau \lambda^2(t) dt, \end{aligned}$$

and we conclude from Equation (9) that $\mathbb{E}[T_3] \lesssim \log^{a+1}(n)/n$.

We finish the proof with T_1 and T_2 . As in Theorem 2, these two terms can be treated using a concentration inequality. First, we need to express each of them as a centered empirical process. For T_1 , write

$$\mathbb{E} \left[\left\{ \sup_{h' \in \mathcal{H}_n} \|\tilde{\lambda}_{h'} - \lambda_{h'}\|^2 - V(h')/10 \right\}_+ \right] \leq \sum_{k, h_k \in \mathcal{H}_n} \mathbb{E} \left[\left\{ \|\tilde{\lambda}_{h_k} - \lambda_{h_k}\|^2 - V(h_k)/10 \right\}_+ \right],$$

and recall that

$$(23) \quad \|\tilde{\lambda}_{h_k} - \lambda_{h_k}\|^2 = \sup_{f \in \mathbb{L}_2([h_k, \tau-h_k]), \|f\|=1} \langle \tilde{\lambda}_{h_k} - \lambda_{h_k}, f \rangle^2.$$

Now, we introduce the following centered empirical process:

$$\nu_{n, h_k}(f) = \langle \tilde{\lambda}_{h_k} - \lambda_{h_k}, f \rangle = \frac{1}{n} \sum_{i=1}^n \int_{h_k}^{\tau-h_k} f(u) \left(\int K_{h_k}(u-s) \left(\frac{dN_i(s)}{1-G(s-)} - \lambda(s) ds \right) \right) du.$$

As $f \mapsto \nu_{n,h_k}(f)$ is continuous, the supremum in (23) can be taken over a countable dense subset of $\{f \in \mathbb{L}_2([1, \tau - 1]), \|f\| = 1\}$, which we denote by $\mathcal{B}_\tau(1)$. Therefore,

$$\mathbb{E}[T_1] \leq \sum_{k, h_k \in \mathcal{H}_n} \mathbb{E} \left[\left\{ \sup_{f \in \mathcal{B}_\tau(1)} \nu_{n, h_k}^2(f) - V(h_k)/10 \right\}_+ \right]$$

and the expectation here can be bounded using the following concentration inequality.

Theorem 4. (Talagrand Inequality) *Let ξ_1, \dots, ξ_n be independent random values, and let $\nu_{n,\xi}(f) = (1/n) \sum_{i=1}^n \{f(\xi_i) - \mathbb{E}[f(\xi_i)]\}$. Then, for a countable class of functions \mathcal{F} uniformly bounded and $\varepsilon > 0$, we have*

$$\mathbb{E} \left[\left\{ \sup_{f \in \mathcal{F}} \nu_{n,\xi}^2(f) - 2(1 + 2\varepsilon^2)H^2 \right\}_+ \right] \leq \frac{4}{d} \left(\frac{W}{n} e^{-d\varepsilon^2 \frac{nH^2}{W}} + \frac{98M^2}{dn^2\varphi^2(\varepsilon)} e^{-\frac{2d\varphi(\varepsilon)\varepsilon}{7\sqrt{2}} \frac{nH}{M}} \right),$$

with $\varphi(\varepsilon) = \sqrt{1 + \varepsilon^2} - 1$, $d = 1/6$ and

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M, \quad \mathbb{E} \left[\sup_{f \in \mathcal{F}} |\nu_{n,\xi}(f)| \right] \leq H, \quad \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \mathbb{V}[f(\xi_i)] \leq W.$$

To apply this result, we first need to compute appropriate values of the bounds H , M , W and the constant ε . Clearly,

$$\mathbb{E} \left[\sup_{f \in \mathcal{B}_\tau(1)} \nu_{n, h_k}^2(f) \right] \leq \mathbb{E} \left[\|\tilde{\lambda}_{h_k} - \lambda_{h_k}\|^2 \right] = \int_{h_k}^{\tau - h_k} \mathbb{V} \left[\tilde{\lambda}_{h_k}(t) \right] dt = V(h_k)/\kappa$$

and thus we require $H^2 = V(h_k)/\kappa$. Then we set $\varepsilon^2 = 1/2$ and $\kappa = 40$ in order to have $2(1 + 2\varepsilon^2)H^2 = V(h_k)/10$.

Now to find the bound M , use the Cauchy-Schwarz inequality and the fact that $\|f\| = 1$ on $\mathcal{B}_\tau(1)$ to write:

$$\begin{aligned} \left| \int_{h_k}^{\tau - h_k} f(u) \int K_{h_k}(u - s) \frac{dN(s)}{1 - G(s-)} du \right| &= \left| \int \left(\int_{h_k}^{\tau - h_k} f(u) K_{h_k}(u - s) du \right) \frac{dN(s)}{1 - G(s-)} \right| \\ &\leq \|f\| \int \left(\int_{h_k}^{\tau - h_k} K_{h_k}^2(u - s) du \right)^{1/2} \frac{dN(s)}{1 - G(s-)} \leq \frac{c_\tau \|K\|}{c_G} \frac{1}{\sqrt{h_k}} := M. \end{aligned}$$

Lastly, we need to determine the adequate bound W . Introduce the notation $K_{h_k}^-(s) = K_{h_k}(-s)$ and write:

$$\begin{aligned}
& \mathbb{V} \left[\int_{h_k}^{\tau-h_k} f(u) \int K_{h_k}(u-s) \frac{dN(s)}{1-G(s-)} du \right] \\
& \leq \mathbb{E} \left[\left(\iint_{h_k}^{\tau-h_k} K_{h_k}(u-s) f(u) du \frac{dN(s)}{1-G(s-)} \right)^2 \right] \\
& \leq \mathbb{E} \left[\left(\int K_{h_k}^- * f(s) \frac{dN(s)}{1-G(s-)} \right)^2 \right] \\
& \leq c_\tau \left(\int \frac{(K_{h_k}^- * f)^2(s)}{1-G(s-)} \lambda(s) ds \right) \\
& \leq \frac{c_\tau \|\lambda\|_{\infty, \tau} \|K_{h_k}^- * f\|^2}{c_G} \leq \frac{c_\tau \|\lambda\|_{\infty, \tau} (\|K_{h_k}^-\|_1)^2 \|f\|^2}{c_G} = \frac{c_\tau \|\lambda\|_{\infty, \tau} (\|K\|_1)^2}{c_G} := W,
\end{aligned}$$

where we used Lemma 9 and the property (21) for $q = 2$. Therefore, W is a constant and we can now apply Talagrand Inequality:

$$\mathbb{E} \left[\left\{ \sup_{f \in \mathcal{B}_\tau(1)} \nu_{n, h_k}^2(f) - V(h_k)/10 \right\}_+ \right] \leq \frac{\vartheta_1}{n} \left(\exp(-\vartheta_2/h_k) + \frac{1}{nh_k} \exp(-\vartheta_3\sqrt{n}) \right),$$

for some positive constants ϑ_1, ϑ_2 and ϑ_3 . Then, from conditions (9), (15) and the fact that $\text{Card}(\mathcal{H}_n) \leq n$, we conclude:

$$\mathbb{E}[T_1] \leq \frac{\vartheta_1}{n} \sum_{k, h_k \in \mathcal{H}_n} \left(\exp(-\vartheta_2/h_k) + \frac{1}{nh_k} \exp(-\vartheta_3\sqrt{n}) \right) \lesssim \frac{1}{n}.$$

The proof for T_2 follows the same line as for T_1 . First,

$$\mathbb{E}[T_2] \leq \sum_{k, h_k \in \mathcal{H}_n} \mathbb{E} \left[\left\{ \|\tilde{\lambda}_{h, h_k} - \lambda_{h, h_k}\|^2 - V(h_k)/10 \right\}_+ \right]$$

and the Talagrand inequality needs to be applied to the centered process $\langle \hat{\lambda}_{h, h_k} - \lambda_{h, h_k}, f \rangle$, where $f \in \mathcal{B}_\tau(1)$. Since $\tilde{\lambda}_{h, h_k} = K_h * \tilde{\lambda}_{h_k}$ and $\lambda_{h, h_k} = K_h * \lambda_{h_k}$ the same bounds H, M and W can be used, up to a constant. Indeed, using the inequalities

$$\|K_h * K_{h_k}\|_2 \leq \|K\|_1 \|K\|_2 (h_k)^{-1/2} \quad \text{and} \quad \|K_h * K_{h_k}^-\|_1 \leq (\|K\|^2)_1$$

it can be shown that Theorem 4 can be applied with

$$H^2 = \frac{V(h_k) (\|K\|_1)^2}{\kappa}, \quad M = \frac{c_\tau \|K\|_1 \|K\|}{c_G \sqrt{h_k}} \quad \text{and} \quad W = \frac{c_\tau \|\lambda\|_{\infty, \tau} (\|K\|_1)^4}{c_G}.$$

Finally, we obtain again $\mathbb{E}[T_2] \lesssim 1/n$.

Gathering the bounds of the five terms gives the result of Theorem 3.

8. TECHNICAL LEMMAS

In order to give a proof of Lemma 2, we first need to introduce the following result which is a direct consequence of Theorem 1 in [4].

Lemma 8. *For all $k \in \mathbb{N}^*$, there exists a positive constant c_k depending on k such that*

$$\mathbb{E} \left[\|\hat{G} - G\|_{\infty, \tau}^{2k} \right] \leq \frac{c_k}{n^k}.$$

Proof. We use a nonasymptotic exponential bound for the Kaplan-Meier estimator which can be formulated as follows (see Bitouzé *et al.*, [4]): there exists a positive constant η such that for any positive constant ε ,

$$(24) \quad \mathbb{P} \left[\sqrt{n} \|(1 - F)(\hat{G} - G)\|_{\infty, \tau} > \varepsilon \right] \leq 2.5 e^{-2\varepsilon^2 + \eta\varepsilon}$$

and so

$$\begin{aligned} & \mathbb{E} \left[\|\hat{G} - G\|_{\infty, \tau}^{2k} \right] \\ & \leq 2k \int_0^{+\infty} u^{2k-1} \mathbb{P} \left[\|\hat{G} - G\|_{\infty, \tau} > u \right] du \\ & \leq 2k \int_0^{+\infty} u^{2k-1} \mathbb{P} \left[c_F^{-1} \|(1 - F)(\hat{G} - G)\|_{\infty, \tau} > u \right] du \\ & \leq 2k \int_0^{+\infty} u^{2k-1} \mathbb{P} \left[\sqrt{n} \|(1 - F)(\hat{G} - G)\|_{\infty, \tau} > c_F \sqrt{n} u \right] du \\ & \leq 5ke^{\eta^2/8} \int_0^{\infty} u^{2k-1} \exp \left\{ -2c_F^2 n \left(u - \frac{\eta}{4\sqrt{n}c_F} \right)^2 \right\} du \\ & \leq \frac{5e^{\eta^2/8} k}{2^k c_F^{2k}} \int_{-\eta/(2\sqrt{2})}^{+\infty} \left(z + \frac{\eta}{2\sqrt{2}} \right)^{2k-1} e^{-z^2} dz n^{-k} := c_k n^{-k}. \end{aligned}$$

□

Proof of Lemma 2. Since $\mathbb{P}[\Omega^c] \leq \mathbb{P}[\Omega_G^c] + \mathbb{P}[(\Omega_G^*)^c]$, we bound each term separately. For any $k > 0$, we have

$$\mathbb{P}[\Omega_G^c] \leq \mathbb{P} \left[\|G - \hat{G}\|_{\infty, \tau} > c_G/2 \right] \leq \frac{4^k}{c_G^{2k}} \mathbb{E} \left[(\|G - \hat{G}\|_{\infty, \tau})^{2k} \right].$$

Thus, Lemma 8 implies that

$$(25) \quad \mathbb{P} [\Omega_G^c] \leq d_k n^{-k}, \text{ where } d_k > 0.$$

Next, we use (24) and write:

$$\begin{aligned} & \mathbb{P} \left[\|\hat{G} - G\|_{\infty, \tau} > c_0 \sqrt{n^{-1} \log(n)} \right] \\ & \leq \mathbb{P} \left[\|(1 - F)(y)\hat{G} - G\|_{\infty, \tau} > c_0 c_F \sqrt{n^{-1} \log(n)} \right] \\ & \leq 2.5 \exp(-2c_F^2 c_0^2 \log(n) + \eta c_F c_0 \sqrt{\log(n)}) \leq 2.5 \exp((-2c_F c_0 + \eta) c_0 c_F \log(n)). \end{aligned}$$

Thus, for $c_0 \geq (\eta + \sqrt{\eta^2 + 8k})(4c_F)^{-1}$ we have

$$\mathbb{P} [\Omega_G^{*c}] = \mathbb{P} \left[\|G - \hat{G}\|_{\infty, \tau} > c_0 \sqrt{n^{-1} \log n} \right] \leq 2.5 n^{-k}.$$

This result and Equation (25) imply $\mathbb{P}[\Omega^c] \leq (d_k + 2.5)n^{-k}$. \square

We conclude this section with a very useful inequality concerning integrals with respect to the counting process N .

Lemma 9. (*Cauchy-Schwarz*) *For every bounded function h on $[0, \tau]$, we have*

$$N(\tau) \int_{\tau_1}^{\tau_2} h^2(s) dN(s) \geq \left(\int_{\tau_1}^{\tau_2} h(s) dN(s) \right)^2,$$

where $0 \leq \tau_1 \leq \tau_2 \leq \tau$.

Proof. We have

$$\begin{aligned} 0 & \leq \int_{\tau_1}^{\tau_2} \left(h(s) - \int_{\tau_1}^{\tau_2} \frac{h(s) dN(s)}{N(\tau)} \right)^2 \frac{dN(s)}{N(\tau)} \\ 0 & \leq \frac{1}{N(\tau)} \int_{\tau_1}^{\tau_2} h^2(s) dN(s) - 2 \left(\int_{\tau_1}^{\tau_2} h(s) \frac{dN(s)}{N(\tau)} \right)^2 + \left(\int_{\tau_1}^{\tau_2} h(s) \frac{dN(s)}{N(\tau)} \right)^2 \int_{\tau_1}^{\tau_2} \frac{dN(s)}{N(\tau)}. \end{aligned}$$

Then, notice that $\int_{\tau_1}^{\tau_2} dN(s) \leq N(\tau)$ to obtain the desired result. \square

REFERENCES

- [1] P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York, 1993.
- [2] P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, 10(4):1100–1120, 1982.
- [3] R. Bartoszyński, B. W. Brown, C. M. McBride, and J. R. Thompson. Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process. *Ann. Statist.*, 9(5):1050–1060, 1981.

- [4] D. Bitouzé, B. Laurent, and P. Massart. A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator. *Ann. Inst. H. Poincaré Probab. Statist.*, 35(6):735–763, 1999.
- [5] O. Bouaziz, Geffray S., and O. Lopez. Semi-parametric inference for the recurrent event process by means of a single-index model. *arXiv:1005.4553*, 2010.
- [6] C.-T. Chiang, M.-C. Wang, and C.-Y. Huang. Kernel estimation of rate function for recurrent event data. *Scand. J. Statist.*, 32(1):77–91, 2005.
- [7] R. J. Cook and J.F. Lawless. *The statistical analysis of recurrent events*. Statistics for biology and health. Springer, New York, 2007.
- [8] J.-Y. Dauxois and S. Sencey. Non-parametric tests for recurrent events under competing risks. *Scand. J. Stat.*, 36(4):649–670, 2009.
- [9] D. Ghosh and D. Y. Lin. Nonparametric analysis of recurrent events and death. *Biometrics*, 56(2):554–562, 2000.
- [10] D. Ghosh and D. Y. Lin. Marginal regression models for recurrent and terminal events. *Statist. Sinica*, 12(3):663–688, 2002.
- [11] D. Ghosh and D. Y. Lin. Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*, 59(4):877–885, 2003.
- [12] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632, 2011.
- [13] Bruce E. Hansen. Exact mean integrated squared error of higher order kernel estimators. *Econometric Theory*, 21(6):1031–1057, 2005.
- [14] J. F. Lawless and C. Nadeau. Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2):158–168, 1995.
- [15] D. Y. Lin, L. J. Wei, I. Yang, and Z. Ying. Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 62(4):711–730, 2000.
- [16] 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 estimator. *Probab. Theory Related Fields*, 80(3):461–473, 1989.
- [17] W. Nelson. Confidence limits for recurrence data – applied to cost or number of product repairs. *Technometrics*, 37(2):147–157, 1995.
- [18] M. Pepe and J. Cai. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Amer. Statist. Assoc.*, 88(423):811–820, 1993.
- [19] William R. Schucany. Adaptive bandwidth choice for kernel regression. *J. Amer. Statist. Assoc.*, 90(430):535–540, 1995.
- [20] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.