Statistica Sinica **23** (2013), 635-665
doi:http://dx.doi.org/10.5705/ss.2011.137

NONPARAMETRIC ESTIMATION OF THE INTENSITY FUNCTION OF A RECURRENT EVENT PROCESS

Olivier Bouaziz[1], Fabienne Comte[2] and Agathe Guilloux[3]

[1]*MAP*5*, UMR CNRS* 8145*, University and IUT Paris Descartes,*
[2]*MAP*5*, UMR CNRS* 8145 *and University Paris Descartes*
and [3]*LSTA, University Pierre et Marie Curie*

Abstract: In this paper, we consider 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. We then describe and study an adaptive procedure of bandwidth selection, in the spirit of Goldenshluger and Lepski (2011), and prove an oracle type bound for both the MSE and the MISE of the final estimator. The method is illustrated by simulation experiments.

Key words and phrases: Adaptive estimation, censoring, intensity, kernel estimator, nonparametric estimation, recurrent event process.

1. Introduction

Recurrent event data arise in such fields as medicine, insurance, economics, and reliability. Medical examples include infections in HIV-infected subjects, tumor recurrences in cancer patients and epileptic seizures of patients. Such repeated events impact on the quality of life of the patients and increase their risk of death. Therefore it is of interest to study the rate function of the recurrent event process that 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 used to estimate the cumulative mean function or the rate function of the recurrent event process. For instance, Andersen and Gill (1982) 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 (1993) and Lin et al. (2000) performed estimation of the regression parameters in a more general model, taking into account time dependent covariates. Ghosh and Lin (2002, 2003) extended these results to the presence of terminal events and derived asymptotic properties of the regression parameter estimates. Finally, Bouaziz, Geffray, and Lopez (2010) studied the cumulative mean function through a single-index assumption that 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 models for the mean or the rate functions that may not hold in practice. In a more flexible way, nonparametric procedures were considered by several authors. In presence of censored data and without the Poisson assumption, Nelson (1995) and Lawless and Nadeau (1995) introduced an estimator of the cumulative mean function and derived a robust estimator of its variance. They also obtained confidence intervals which enabled them to compare mean functions in a two-sample test. The theoretical properties of this estimator were derived in Ghosh and Lin (2000). In their main result, the cumulative mean function is shown to converge weakly to a zero mean Gaussian process. More recently, Dauxois and Sencey (2009) 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 a smoothing approach were introduced in this framework. Bartoszyński et al. (1981) briefly presented a kernel estimator of the rate function when the recurrent events were distributed according to a Poisson process and the censored times constant. Then, Chiang, Wang, and Huang (2005) 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, asymptotic normality was proved and confidence intervals were derived using a bootstrap method, with theoretical arguments provided to validate their procedures. Another kind of smoothing estimator was introduced in Bouaziz, Geffray, and Lopez (2010) 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 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 Goldenshluger and Lepski (2011). 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 on the estimator for a fixed bandwidth. An adaptive procedure of bandwidth selection is 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 to assess the practical properties of the method. We also provide a comparison with a bootstrap method adapted from Chiang, Wang, and Huang (2005). Concluding remarks are made in Section 6. The main proofs are detailed in Section 7, and some technical results are postponed to the Appendix, Section 8.

2. Notation and Assumptions

2.1. Notation

For a real $q \geq 1$ and a function $f : \mathbb{R} \longmapsto \mathbb{R}$ such that $|f|^q$ is integrable or bounded, we take

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

For simplicity, we 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 square-integrable functions f and g from $\mathbb{R}$ to $\mathbb{R}$, we denote the convolution product of f and g by $f * g$. For quantities $\alpha(n)$ and $\gamma(n)$, the notations $\alpha(n) \lesssim \gamma(n)$ and $\alpha(n) \propto \gamma(n)$ mean 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

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

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

$$F(t) = \mathbb{P}[D \leq t], \quad G(t) = \mathbb{P}[C \leq t] \text{ and } H(t) = \mathbb{P}[T \leq t], \qquad t \geq 0. \tag{2.1}$$

The mean function of N^* is $\mathbb{E}[N^*(t)] = \mu(t)$ 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 one. In our context $\lambda(t)$ refers to the occurrence probability of recurrent events at time t unconditioned by 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 was defined in Cook and Lawless (2007) and referred to as the rate function, denoted by ρ. It was also introduced in Dauxois and Sencey (2009) as the frequency function. While the definition of the rate function in Chiang, Wang, and Huang (2005) is different from ours, notice that multiplying their rate function by $1 - G(\cdot-)$ gives the intensity function here.

To make inference about λ, we introduce some assumptions.

Assumption 1.

(i) C is independent of D and of the process $(N(t))_{t \geq 0}$;

(ii) $\mathbb{P}\big[dN^*(C) \neq 0\big] = 0$;

(iii) $\mathbb{P}[D = C] = 0$.

Assumption (i) is common in the context of recurrent events when censored data are present (see e.g., Dauxois and Sencey (2009), Ghosh and Lin (2000)). Assumptions (ii) and (iii) preclude ties between death, censoring, and the apparition of recurrent event. Notice that in practical situations, if such ties exist, one can assign censored events values that are just larger than their actual values.

The next assumption circumvents problems arising in the tails of the distributions of C and N.

Assumption 2.

(i) for F, G, H defined by (2.1), there exist 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., Andersen et al. (1993)) while the second can be found e.g., in Dauxois and Sencey (2009). The last one is an additional condition required only for the pointwise setting.

2.3. Kernel and functional assumptions

Our goal is to perform non-parametric estimation of λ using a kernel-type estimator. Classical regularity conditions are required for the intensity function and the kernel K.

Assumption 3. There exists $\beta > 0$, $L > 0$ such that $\lambda^{(l)}$ exists for $l = \lfloor \beta \rfloor$ and

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

The following set of assumptions is fulfilled by many standard kernel functions.

Assumption 4.
(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

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

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

3. The MSE and MISE of $\hat{\lambda}_h$

3.1. Kernel estimator

One of the difficulties in estimating the intensity function is that the process $N^*(t)$ is not directly observable. Our estimation procedure is based on an equality that provides an 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

$$\mathbb{E}[dN(t)] = \mathbb{E}[dN^*(t \wedge C)] = \mathbb{E}[dN^*(t)\mathbb{E}[I(t \leq C)|N^*]] = \lambda(t)\big(1 - G(t-)\big)dt. \tag{3.1}$$

The distribution function G is estimated by $\hat{G}$, the Lo, Mack, and Wang (1989) 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 ith order statistic in the sample $T_1, \ldots, T_n$ and the $(\delta_{(i)})$'s are the δ_i's associated to these indexes. Notice then that, for all $t \geq 0$

$$1 - \hat{G}(t) \geq (n+1)^{-1}. \tag{3.2}$$

We propose a kernel estimator to estimate λ

$$\hat{\lambda}_h(t) = \frac{1}{nh} \sum_{i=1}^{n} \int K\left(\frac{t-s}{h}\right) \frac{dN_i(s)}{1 - \hat{G}(s-)}, \tag{3.3}$$

where K and h satisfy Assumption 4. As the kernel has support $[-1, 1]$, the integrand in (3.3) vanishes outside $[t-h, t+h]$. Therefore, given a bandwidth h, we only discuss estimation of λ for t such that $t \pm h \in [0, \tau]$.

Consider the pseudo-estimator

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

the kernel estimator of λ if G were known. The study of the quadratic error of $\hat{\lambda}_h - \lambda$ is divided into two parts – the error of $\tilde{\lambda}_h - \lambda$, and that of $\tilde{\lambda}_h - \hat{\lambda}_h$. Bounds on 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 results with rather classical tools, 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*

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

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

where

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

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

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

Proof. For the bias terms, we proceed as in Tsybakov (2009) and observe that, from (3.1)

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

and, using a change of variables, this leads to

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

Now write $\lambda(t+uh) = \lambda(t) + \lambda'(t)uh + \cdots + ((uh)^l/l!)\lambda^{(l)}(t+\xi uh)$ for $0 \leq \xi \leq 1$, and use Assumptions 3 and 4 to obtain the required squared 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) and use 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.

3.3. Study of the estimator $\hat{\lambda}_h$

The difference between $\hat{\lambda}_h$ and $\tilde{\lambda}_h$ is expressed here, and the proof is in Section 7.

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

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

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

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

Gathering the results of Proposition 1 (a)−(b) and Lemma 1 provides global bounds for the estimator.

Theorem 1. *Under Assumptions* 1 *to* 4 *we have:*
(a) *for all $t \in [h, \tau - h]$,*

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

(b)

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

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

The inequalities stated in Theorem 1 are nonasymptotic; in both cases, they provide a bound that contains a squared-bias term of order $h^{2\beta}$, a variance term of order $1/(nh)$ and a residual term that is negligible. For an asymptotic convergence rate, 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 $h \propto n^{-1/(2\beta+1)}$ to 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. To circumvent this impossibility we provide a data-driven way of selecting the bandwidth that allows one to reach almost, or exactly, the optimal rate without knowledge of β.

4. Adaptive Estimation of λ

4.1. Pointwise bandwidth selection

We wish to automatically select a relevant bandwidth for our estimator using the Goldenshluger and Lepski (2011) 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 $*$ denotes the convolution product. 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 $\kappa_0 > 0$, take

$$V_0(h) = \kappa_0 \frac{c_\tau \|\lambda\|_{\infty,\tau}\|K\|^2 \log(n)}{nhc_G} \tag{4.1}$$

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

$$A_0(h,t_0) = \sup_{h'\in\mathcal{H}_n} \left\{(\hat{\lambda}_{h'} - \hat{\lambda}_{h,h'})^2(t_0) - V_0(h')\right\}_+ . \tag{4.2}$$

We define our adaptive estimator as follows:

$$\hat{h}(t_0) = \operatorname*{argmin}_{h\in\mathcal{H}_n} \left(A_0(h,t_0) + V_0(h)\right) \text{ and } \check{\lambda}(t_0) = \hat{\lambda}_{\hat{h}(t_0)}(t_0). \tag{4.3}$$

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

$$\forall h \in \mathcal{H}_n,\ nh \geq \kappa_1 \log(n), \quad \textit{for some } \kappa_1 \geq 0, \tag{4.4}$$

and

$$\sum_{k:\ h_k\in\mathcal{H}_n} \frac{1}{nh_k} \lesssim \log^a(n), \quad \textit{for some } a \geq 0, \tag{4.5}$$

then there exists a constant κ_0 *such that* $\check{\lambda}$ *defined by* (4.1)*,* (4.2)*, and* (4.3) *satisfies:*

$$\forall h \in \mathcal{H}_n, \quad \mathbb{E}\left[\left(\check{\lambda}(t_0) - \lambda(t_0)\right)^2\right] \leq c(c_1^2 h^{2\beta} + V_0(h)) + c'\frac{\log^{(1+a)}(n)}{n}, \tag{4.6}$$

where c *is an absolute constant and* c' *a 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 can be taken as 80 to get the theoretical result, but a smaller value works in practice (see Section 5); $\log(n)/(nh)$ gives the asymptotic order of the term and is known; $\|K\|$ is a known constant, as the kernel is user-chosen; c_τ and $\|\lambda\|_{\infty,\tau}$ are unknown quantities that can be estimated by

$$\hat{c}_\tau = \max_{1\leq i\leq n} N_i(\tau), \quad \widehat{\|\lambda\|}_{\infty,\tau} = \sup_{x\in[h_n,\tau-h_n]} \hat{\lambda}_{h_n}(x). \tag{4.7}$$

Here h_n is an arbitrary bandwidth (it can be taken as $n^{-1/5}$ for instance). Note that if we replace the unknown terms in $V_0(h)$ by their estimates, we get an estimate $\hat{V}_0(h)$.

The bound (4.6) holds for all $h \in \mathcal{H}_n$ and therefore automatically reaches 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 $\mathcal{H}_n$. 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 adaptively optimal.

Example of $\mathcal{H}_n$. Considering the constraints (4.4) and (4.5) on $\mathcal{H}_n$, we can take

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

so that $\mathrm{Card}(\mathcal{H}_n) \leq n$ 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 (4.4). 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 (4.5) holds with $a = 1$.

4.2. Global bandwidth selection

In the global risk setting, we set, for some $\kappa > 0$,

$$V(h) = \kappa \frac{c_\tau \Lambda(\tau) \|K\|^2}{nh}, \tag{4.8}$$

and we consider for $\mathcal{H}_n$ the discrete set of bandwidths specified above. Let

$$A(h) = \sup_{h' \in \mathcal{H}_n} \left\{ \|\hat{\lambda}_{h'} - \hat{\lambda}_{h,h'}\|^2 - V(h') \right\}_+ \tag{4.9}$$

and

$$\hat{h} = \operatorname*{argmin}_{h \in \mathcal{H}_n} (A(h) + V(h)) \ \text{ and } \lambda^* = \hat{\lambda}_{\hat{h}}. \tag{4.10}$$

Theorem 3. *Under Assumptions 1 to 4, if $\mathcal{H}_n$ is a finite discrete set of bandwidths such that Card($\mathcal{H}_n$) $\leq n$, (4.5) holds and*

$$\sum_{k \,:\, h_k \in \mathcal{H}_n} \exp(-\frac{b}{h_k}) < +\infty, \quad \forall b > 0, \tag{4.11}$$

then there exists a constant κ such that λ^* defined by (4.8), (4.9) and (4.10) satisfies:

$$\forall h \in \mathcal{H}_n, \quad \int_1^{\tau-1} \mathbb{E}\left[\left(\lambda^*(t) - \lambda(t)\right)^2 \right] dt \leq c(\tau c_1^2 h^{2\beta} + V(h)) + c' \frac{\log^{1+a}(n)}{n}, \tag{4.12}$$

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 pertain 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.

Here are examples of $\mathcal{H}_n$ that satisfy Assumption 4 (iv), (4.5), and (4.11).

Example 1. Take

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

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

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

which ensures condition (4.5). Lastly,

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

and (4.11) is ensured.

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 (4.12) to hold.

Example 2. Take

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

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

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

which ensures (4.5). Lastly

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

and (4.11) is verified. Here, no minimum regularity condition of the function to estimate is needed.

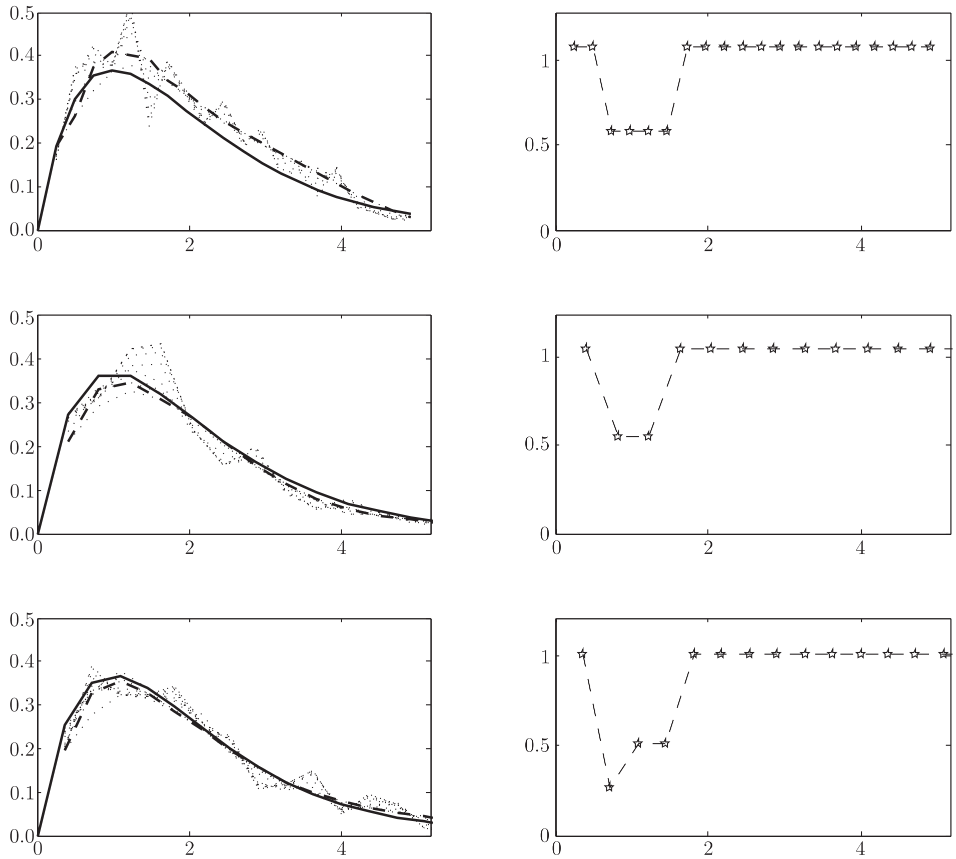

Figure 1. Scenario 1 with $\beta = 1$ and $n = 500$, $\bar{re} = 1.01$, $pc = 0\%$ (top), n =1,000, $\bar{re} = 1.05$, $pc = 0\%$ (middle), $n = 5,000$, $\bar{re} = 0.96$, $pc = 0\%$ (bottom)

5. Simulations

We illustrate the behavior of the 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 boostrap-based selection.

5.1. Description of the simulation scheme

Recurrent events data were simulated as follows. For individuals $i = 1, \ldots, n$, the terminal event D_i was 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 $[0, D_i]$ were simulated according to a Poisson distribution $\mathcal{P}(\int_0^{D_i} \varphi(u)du)$. Finally the recurrent times for

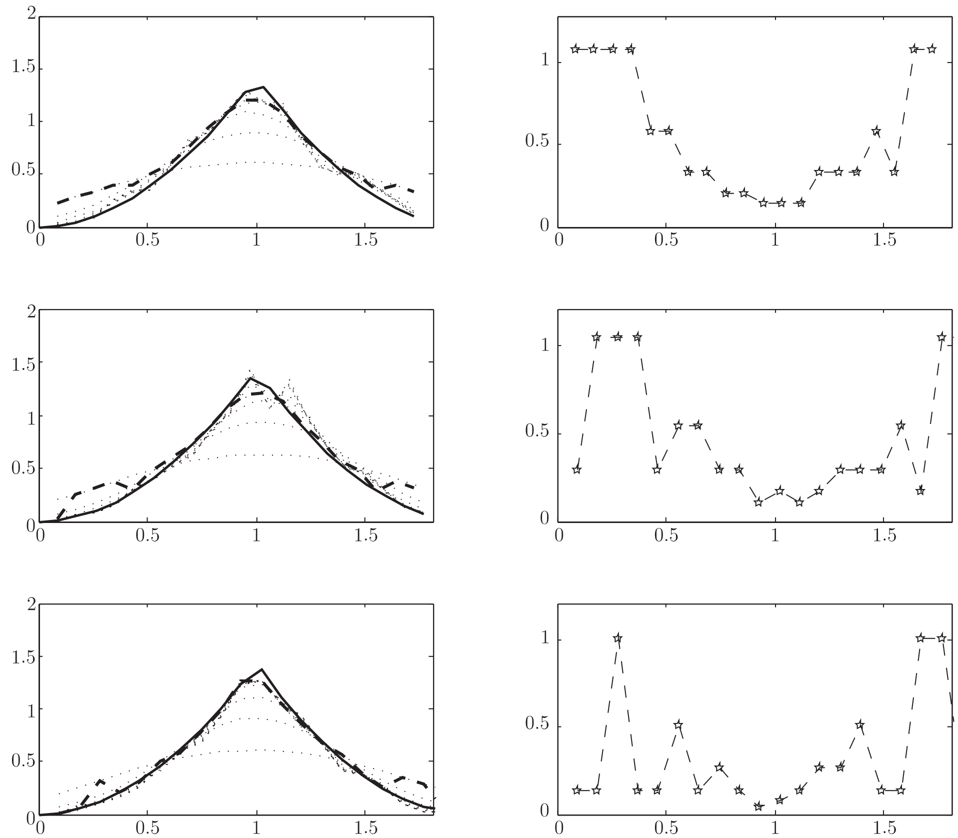

Figure 2. Scenario 2 with $\beta = 0.05$ and $n = 500$, $\bar{re} = 0.87$, $pc = 0\%$ (top), n =1,000, $\bar{re} = 0.95$, $pc = 0\%$ (middle), $n = 5,000$, $\bar{re} = 0.93$, $pc = 0\%$ (bottom)

individual i were 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^* was then $\lambda(t) = \varphi(t)(1 - F(t))$.

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 were constructed with Epanechnikov kernels

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

We used a data-driven criterion for the selection of the bandwidth, by replacing

$V_0(h)$ in Definition (4.1) by

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

with

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

where $E_{\max}$ was the greatest observed recurrent event. We set the universal value of κ_0 at 10^{-2} after an extensive simulation study comparing 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 was

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

5.2. Illustration of the behavior of the adaptive estimator

For Figures 1−3, the intensity functions were estimated on a 20-point grid, regularly spaced on $[0, E_{\max}]$. The number of observations n, the mean number of recurrent $\bar{re}$ and the level of censoring pc are reported in the captions. On the left of the figures, the true intensity functions are drawn with solid lines, the estimators with dashed lines, and the set of all the estimators proposed to the selection algorithm with dotted lines. The right plots show the value of the selected windows for all points on the grid.

In Figures 1 and 2, we see the behavior of our estimators as the sample size grows. In scenario 1, where the intensity λ is smooth, and in scenario 2 where λ has a singularity, the estimator behaves as expected, improving with the sample size.

In Figure 3, we see the behavior of our estimator as 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 $\bar{re} = 0.89$, when $pc = 3\%$, to $\bar{re} = 0.25$, when $pc = 51\%$) as does the time

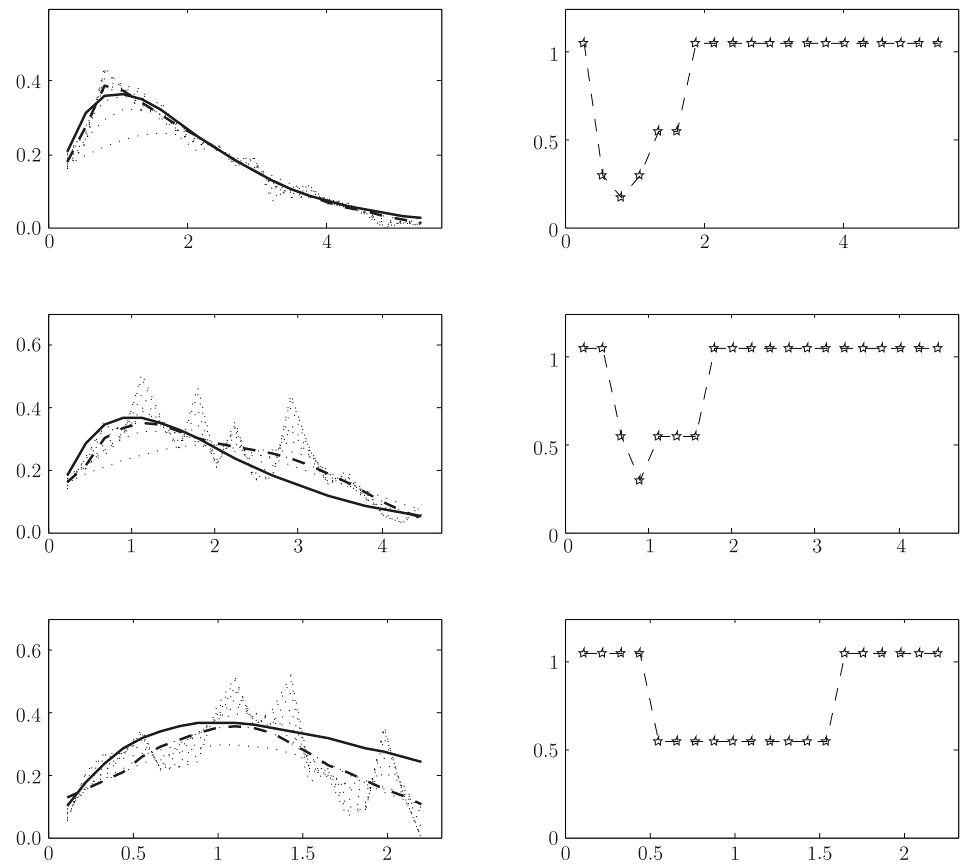

Figure 3. Scenario 1 with $\beta = 1$ and $n = 1,000$, $\bar{re} = 0.89$, $pc = 3\%$ (top), $n = 1,000$, $\bar{re} = 0.60$, $pc = 26\%$ (middle), $n = 1,000$, $\bar{re} = 0.25$, $pc = 51\%$ (bottom)

intervals, on which they are observed (from $[0, 5.7]$ when $pc = 3\%$, to $[0, 2.3]$, when $pc = 51\%$).

We 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 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 over time since none of the proposed curves would globally give a better estimate.

5.3. Monte Carlo study

Here, we compare our adaptive strategy for the selection of the bandwidth to a bootstrap-based strategy, in the spirit of Chiang, Wang, and Huang (2005). Note that no theoretical results are available for the latter in our setting; indeed Chiang, Wang, and Huang (2005) present only asymptotic results for a fixed bandwidth.

We conducted a Monte Carlo study (with $M = 100$ replications) and calculated the squared error for $m = 1, \ldots, M$

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

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

We compare two bandwidth selection procedures: the adaptive procedure described in the previous subsection, denoted by $SE(\hat{\lambda}^m_{Adapt})$, and its squared error, with $\hat{\lambda}^m_{\eta(t_k)} = \hat{\lambda}^m_{\hat{h}(t_k)}$, $\hat{h}(t_k)$ as in (4.3); a bootstrap-based procedure, denoted by $SE(\hat{\lambda}^m_{Boot})$, and its squared error, with $\hat{\lambda}^m_{\eta(t_k)} = \hat{\lambda}^m_{\hat{h}^{m,*}(t_k)}$,

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

with

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

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

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

where N^m_i and $\hat{G}^m$ are calculated on the mth Monte Carlo experiment and $K_{E,4} = (15/8) \times (1 - (7/2)u^2) K_{E,2}(u)$, see Chiang, Wang, and Huang (2005) and Schucany (1995) for the estimation of the bias, and Hansen (2005) for the definition of $K_{E,4}$.

In Tables 1 to 3, we display the mean and the median of $SE(\hat{\lambda}^m_{Adapt})$ and $SE(\hat{\lambda}^m_{Boot})$ obtained from $M = 100$ Monte Carlo experiments. The number of

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

$\times 10^3$	$n = 200$		$n = 500$		$n = 1,000$	
	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 2. Scenario 1 with $\sim 33\%$ of censoring and $\beta = 1$.

$\times 10^3$	$n = 200$		$n = 500$		$n = 1,000$	
	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 3. Scenario 2 with 0% of censoring and $\beta = 0.05$.

$\times 10^2$	$n = 200$		$n = 500$		$n = 1,000$	
	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

bootstrap samples was $B = 100$ 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 was borrowed from our theoretical proposal.

We can see from Tables 1−3 that the perfomances 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 partly due to the choice of the set of bandwidths $\mathcal{H}_n$ which sets the bootstrap under control. We also emphasize that the number of kernel estimators that are computed are $|\mathcal{H}_n|^2$ for our method versus $B|\mathcal{H}_n|$ for the bootstrap method, where $B = 100$ and $|\mathcal{H}_n| \simeq 10$ for $n = 1,000$; our method is approximately ten times faster.

6. Concluding Remarks

In this work, we provide a kernel estimator for the intensity function of a recurrent event process, and prove oracle type inequalities for the risk of an adaptive estimator with data-driven selected bandwidth. We have studied pointwise risk for pointwise-chosen bandwidth and integrated global risk with a globally selected bandwidth. Our bandwidth selection proposal differs from standard cross-validation methods. It is based on recent ideas developed by Goldenshluger and Lepski (2011); the results are new, and proofs are of interest. We also assessed the practical feasibility and performance of our proposal through a short simulation study.

7. Proofs

7.1. Proof of Lemma 1

The proof relies on some additional lemmas. First, write

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

and consider

$$\begin{aligned}
\Omega_G &= \left\{\omega : \forall t \in [0,\tau], G(t) - \hat{G}(t) \geq -\frac{c_G}{2}\right\}, \\
\Omega_G^\star &= \left\{\omega : \forall t \in [0,\tau], |G(t) - \hat{G}(t)| \leq c_0\sqrt{n^{-1}\log n}\right\}, \\
\Omega_{c_0} &= \Omega_G \cap \Omega_G^\star. \qquad (7.1)
\end{aligned}$$

We study the difference process $\hat{\lambda}_h - \tilde{\lambda}_h$ on Ω_{c_0} and its complement. The proof of the next lemma 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*

$$\mathbb{P}\left[\Omega_{c_0(p)}^c\right] \leq c_2 n^{-p}, \qquad (7.2)$$

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

Let $\Omega_p = \Omega_{c_0(p)}$ be such that (7.2) holds.

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

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

where

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

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

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

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

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

Proof of Lemmas 3 and 4. From $1-\hat{G}(t) \geq (n+1)^{-1}$ (see (3.2)) and $\|\hat{G}-G\|_\infty < 1$, we have for all $t \in [h, \tau-h]$,

$$\begin{aligned}
\mathbb{E}\left[\left(\hat{\lambda}_h(t)-\tilde{\lambda}_h(t)\right)^2 I(\Omega_p^c)\right] &\leq \frac{(n+1)^2}{n^2}\mathbb{E}\left[\left(\sum_{i=1}^n \int \frac{K_h\,(t-s)}{1-G(s-)}dN_i(s)\right)^2 I(\Omega_p^c)\right] \\
&\leq (n+1)^2 \mathbb{E}\left[\left(\int \frac{K_h\,(t-s)}{1-G(s-)}dN(s)\right)^2 I(\Omega_p^c)\right] \\
&\leq (n+1)^2 c_\tau \mathbb{E}\left[\int \frac{K_h^2\,(t-s)\, I(\Omega_p^c)}{(1-G(s-))^2}dN(s)\right], \qquad (7.3)
\end{aligned}$$

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

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

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 the Cauchy-Schwarz inequality, we get from (7.3) that

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

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

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

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

where $c_4 = 4c_0^2 c_G^{-2}$ *and* $c_0 = c_0(p)$. *For* $t \in [h, \tau - h]$, *we have*

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

where c *is a positive constant.*

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

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

with $\Lambda(\tau)$ *as in Proposition 1, and*

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

where c *is a positive constant.*

Proof of Lemmas 5 and 6. First, use $1 - \hat{G}(t) = 1 - G(t) + G(t) - \hat{G}(t) \geq c_G/2$ on Ω_G and $\|G(t) - \hat{G}(t)\|_\infty \leq c_0 \sqrt{n^{-1} \log n}$ on $\Omega_G^\star$, to write

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

Then

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

and

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

from Proposition 1. Combining these bounds gives Lemma 5.

The proof of Lemma 6 is much the same. From a change of variables and the Cauchy-Schwarz inequality we have

$$\int_h^{\tau-h} \left(\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \int \frac{|K_h(t-s)|}{1 - G(s-)} dN_i(s)\right]\right)^2 dt$$

$$
\begin{aligned}
&= \int_h^{\tau-h} \left(\int |K_h(t-s)| \lambda(s) ds \right)^2 dt \\
&\leq \left(\int_{-1}^1 K^2(u) du \right) \int_h^{\tau-h} \int_{-1}^1 \lambda^2(t-uh) du dt \\
&\leq 2\|K\|^2 \int_0^\tau \lambda^2(t) dt,
\end{aligned}
$$

where the last inequality is obtained with another change of variables. From 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.

The results of Lemmas 3 to 6 imply Lemma 1.

7.2. Proof of Theorem 2

For all $h \in \mathcal{H}_n$,

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

Since $V_0(h)$, given in (4.1), and $(\hat{\lambda}_h(t_0) - \lambda(t_0))^2$, bounded in Theorem 1a, have the right order (with an additional $\log(n)$ for V_0), we only study $A_0(h,t_0)$. With $\tilde{\lambda}_{h,h'} = K_{h'} * \tilde{\lambda}_h$, $\lambda_h(t_0) = \mathbb{E}[\tilde{\lambda}_h(t_0)]$ and $\lambda_{h,h'}(t_0) = \mathbb{E}[\tilde{\lambda}_{h,h'}(t_0)]$, $A_0(h,t_0)$ as

$$
\begin{aligned}
A_0(h,t_0) &= \sup_{h' \in \mathcal{H}_n} \left\{ \big(\hat{\lambda}_{h'}(t_0) - \hat{\lambda}_{h,h'}(t_0)\big)^2 - V_0(h') \right\}_+ \\
&\leq 5 \sup_{h' \in \mathcal{H}_n} \left\{ \big(\tilde{\lambda}_{h'}(t_0) - \lambda_{h'}(t_0)\big)^2 - \frac{V_0(h')}{10} \right\}_+ \\
&\quad + 5 \sup_{h' \in \mathcal{H}_n} \left\{ \big(\tilde{\lambda}_{h,h'}(t_0) - \lambda_{h,h'}(t_0)\big)^2 - \frac{V_0(h')}{10} \right\}_+ \\
&\quad + 5 \sup_{h' \in \mathcal{H}_n} \big(\hat{\lambda}_{h'}(t_0) - \tilde{\lambda}_{h'}(t_0)\big)^2 + 5 \sup_{h' \in \mathcal{H}_n} \big(\hat{\lambda}_{h,h'}(t_0) - \tilde{\lambda}_{h,h'}(t_0)\big)^2 \\
&\quad + 5 \sup_{h' \in \mathcal{H}_n} \big(\lambda_{h'}(t_0) - \lambda_{h,h'}(t_0)\big)^2
\end{aligned}
$$

$$:= 5(T_{0,1} + T_{0,2} + T_{0,3} + T_{0,4} + T_{0,5}).$$

Since

$$\begin{aligned}|\lambda_{h'}(t_0) - \lambda_{h,h'}(t_0)| &= |K_{h'} * \lambda(t_0) - K_{h'} * K_h * \lambda(t_0)| = |K_{h'} * (\lambda - K_h * \lambda)(t_0)| \\ &\leq \|K\|_1 \sup_{t\in[0,\tau]} |(\lambda - K_h * \lambda)(t)|, \\ T_{0,5} &\leq \|K\|_1^2 \, \|\lambda - K_h * \lambda\|_{\infty,\tau}^2 \leq (\|K\|_1)^2 c_1^2 h^{2\beta},\end{aligned}$$

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

We decompose $T_{0,3}$ into terms corresponding to $I(\Omega_p)$ and $I(\Omega_p^c)$, Ω_p as in (7.1). 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 $\mathrm{Card}(\mathcal{H}_n) \leq n$. This term is of order $1/n$ as soon as $p \geq 12$. On the other hand,

$$\begin{aligned}&\mathbb{E}\Big[\sup_{h'\in\mathcal{H}_n} (\hat{\lambda}_{h'} - \tilde{\lambda}_{h'})^2(t_0) I(\Omega_p)\Big] \\ &\leq \frac{4c_0^2}{c_G^2}\frac{\log(n)}{n}\mathbb{E}\left[\sup_{h'\in\mathcal{H}_n}\left(\int \frac{|K_{h'}(t_0-s)|}{1-G(s-)}\left(\frac{1}{n}\sum_{i=1}^n dN_i(s)\right)\right)^2\right] \\ &\leq \frac{8c_0^2}{c_G^2}\frac{\log(n)}{n}\mathbb{E}\left[\sup_{h'\in\mathcal{H}_n}\left(\int \frac{|K_{h'}(t_0-s)|}{1-G(s-)}\left(\frac{1}{n}\sum_{i=1}^n dN_i(s)-\lambda(s)(1-G(s-))ds\right)\right)^2\right] \\ &\quad + \frac{8c_0^2}{c_G^2}\frac{\log(n)}{n}\sup_{h'\in\mathcal{H}_n}\left(\int |K_{h'}(t_0-s)|\lambda(s)ds\right)^2 \\ &\leq \frac{8c_0^2}{c_G^2}\frac{\log(n)}{n}\sum_{k,h_k\in\mathcal{H}_n}\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^n\int\frac{|K_{h_k}(t_0-s)|}{1-G(s-)}dN_i(s)\right] + \frac{8c_0^2\|\lambda\|_{\infty,\tau}^2}{c_G^2}\frac{\log(n)}{n}\|K\|_1^2 \\ &\leq \frac{8c_0^2}{c_G^3}\frac{\log(n)}{n}\sum_{k,h_k\in\mathcal{H}_n}\frac{c_\tau\|\lambda\|_{\infty,\tau}\|K\|^2}{nh_k} + \frac{8c_0(2\|\lambda\|_{\infty,\tau})^2}{c_G^2}\frac{\log(n)}{n}\|K\|_1^2, \qquad (7.4)\end{aligned}$$

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 (4.5), and this bounds $T_{0,3}$.

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,

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

from Lemma 3, and

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

Using

$$\|f*g\|_q \le \|f\|_1\|g\|_q \quad \text{for} \quad q\ge 1, \tag{7.5}$$

it is easy to see that

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

and

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

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

To study $T_{0,1}$ and $T_{0,2}$, we recall the following.

Lemma 7 (Bernstein inequality)**.** *Let $\xi_1,\dots,\xi_n$ be independent and identically distributed random variables and $S_n(\xi) = \sum_{i=1}^n \xi_i$. Then, for $\eta > 0$,*

$$\mathbb{P}\left(|S_n(\xi) - \mathbb{E}[S_n(\xi)]| \ge n\eta\right) \le 2\max\left(\exp\left(-\frac{n\eta^2}{4w}\right), \exp\left(-\frac{n\eta}{4b}\right)\right), \tag{7.6}$$

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

We 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

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

Thus, (7.6) can be written in the following way: for some $x > 0$,

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

With $\kappa_0 \geq 80$,

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

On the other hand,

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

Taking $\kappa_1 \geq 4\kappa_2^{-2}$ in (4.4) gives

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

Therefore, we have

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

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

$$\begin{aligned}
&\mathbb{E}\left[\left\{|\tilde{\lambda}_h(t_0) - \lambda_h(t_0)|^2 - \frac{V_0(h)}{10}\right\}_+\right] \\
&\quad \leq \int_0^{+\infty} \mathbb{P}\left[|\tilde{\lambda}_h(t_0) - \lambda_h(t_0)| \geq \sqrt{\frac{V_0(h)}{10} + x}\right] dx
\end{aligned}$$

$$\leq 2n^{-2}\max\left(\int_0^{+\infty} e^{-\kappa_3 nhx}dx, \int_0^{+\infty} e^{-\kappa_4 nh\sqrt{x}}dx\right)$$
$$\leq 2n^{-2}\max\left(\frac{1}{\kappa_3 nh}, \frac{2}{\kappa_4^2(nh)^2}\right) \leq \kappa_5 n^{-2}$$

for some positive constant κ_5. Finally,

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

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

Now $T_{0,2}$ can be treated in a similar way. Write

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

and proceed as in 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 Theorem 2.

7.3. Proof of Theorem 3

As for 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 $V(h)$ and $\|\hat\lambda_h-\lambda\|^2$ (see Theorem 1(b)), have the right 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

$$\begin{aligned}A(h) &= \sup_{h'\in\mathcal{H}_n}\left\{\|\hat\lambda_{h'}-\hat\lambda_{h,h'}\|^2-V(h')\right\}_+\\ &\leq 5\sup_{h'\in\mathcal{H}_n}\left\{\|\tilde\lambda_{h'}-\lambda_{h'}\|^2-\frac{V(h')}{10}\right\}_+ +5\sup_{h'\in\mathcal{H}_n}\left\{\|\tilde\lambda_{h,h'}-\lambda_{h,h'}\|^2-\frac{V(h')}{10}\right\}_+\end{aligned}$$

$$+5 \sup_{h'\in\mathcal{H}_n} \|\hat{\lambda}_{h'} - \tilde{\lambda}_{h'}\|^2 + 5 \sup_{h'\in\mathcal{H}_n} \|\hat{\lambda}_{h,h'} - \tilde{\lambda}_{h,h'}\|^2 + 5 \sup_{h'\in\mathcal{H}_n} \|\lambda_{h'} - \lambda_{h,h'}\|^2$$
$$:= 5(T_1 + T_2 + T_3 + T_4 + T_5).$$

From

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

where we used (7.5) with $q = 2$, we obtain

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

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

Now, with

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

we have

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

where the last inequality is obtained from Lemma 1.

For T_3, from Lemma 4,

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

and this is of order $1/n$ as long as $p \ge 10$. Then, similar to (7.4),

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

and we conclude from (4.5) that $\mathbb{E}[T_3] \lesssim \log^{a+1}(n)/n$.

As in Theorem 2, T_1 and T_2 can be treated using a concentration inequality. First, we need to express each of them as a centered empirical process. For T_1, write

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

and recall that

$$\|\tilde{\lambda}_{h_k} - \lambda_{h_k}\|^2 = \sup_{f\in\mathbb{L}_2([h_k,\tau-h_k]),\|f\|=1} \langle \tilde{\lambda}_{h_k} - \lambda_{h_k}, f\rangle^2. \tag{7.7}$$

Now we introduce the centered empirical process

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

As $f \mapsto \nu_{n,h_k}(f)$ is continuous, the supremum in (7.7) 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,

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

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

Theorem 4 (Talagrand Inequality)**.** *Let $\xi_1,\ldots,\xi_n$ 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*

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

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

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

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

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

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

$$\begin{aligned}&\left|\int_{h_k}^{\tau-h_k} f(u)\int K_{h_k}(u-s)\frac{dN(s)}{1-G(s-)}du\right| \\ &\quad = \left|\int\left(\int_{h_k}^{\tau-h_k} f(u)K_{h_k}(u-s)\,du\right)\frac{dN(s)}{1-G(s-)}\right|\end{aligned}$$

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

To determine W, introduce the notation $K_{h_k}^-(s) = K_{h_k}(-s)$ and write

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

where we used Lemma 9 and (7.5) for $q = 2$. Therefore, W is a constant and we can apply the Talagrand Inequality:

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

for some positive constants ϑ_1, ϑ_2 and ϑ_3. Then, from (4.5), (4.11), and the fact that $\mathrm{Card}(\mathcal{H}_n) \leq n$, we have

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

Now, as with T_1,

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

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

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

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

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

8. Technical Lemmas

For Lemma 2, we consider a result which is a direct consequence of Theorem 1 in Bitouzé, Laurent and Massart (1999).

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 that can be formulated as follows (see Bitouzé, Laurent and Massart (1999)): there exists a positive constant η such that for any positive constant ε,

$$\mathbb{P}\left[\sqrt{n}\|(1-F)\,(\hat{G} - G)\|_{\infty,\tau} > \varepsilon\right] \leq 2.5\, e^{-2\varepsilon^2 + \eta\varepsilon} \tag{8.1}$$

and so

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

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

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

Thus, Lemma 8 implies that

$$\mathbb{P}\left[\Omega_G^c\right] \leq d_k n^{-k}, \text{ where } d_k > 0. \tag{8.2}$$

Next, we use (8.1) and write

$$\begin{aligned}
&\mathbb{P}\left[\|\hat{G}-G\|_{\infty,\tau} > c_0\sqrt{n^{-1}\log(n)}\right] \\
&\quad \le \mathbb{P}\left[\|(1-F)(\hat{G}-G)\|_{\infty,\tau} > c_0 c_F\sqrt{n^{-1}\log(n)}\right] \\
&\quad \le 2.5\exp(-2c_F^2c_0^2\log(n)+\eta c_F c_0\sqrt{\log(n)}) \le 2.5\exp((-2c_Fc_0+\eta)c_0c_F\log(n)).
\end{aligned}$$

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

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

This and (8.2) imply that $\mathbb{P}[\Omega^c] \le (d_k+2.5)n^{-k}$.

We conclude this section with a 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*

$$N(\tau)\int_{\tau 1}^{\tau_2} h^2(s)dN(s) \ge \left(\int_{\tau_1}^{\tau_2} h(s)dN(s)\right)^2,$$

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

Proof. We have

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

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

References

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer Series in Statistics. Springer-Verlag, New York.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Bartoszyński, R., Brown, B. W., McBride, C. M. and Thompson, J. R. (1981). Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process. *Ann. Statist.* **9**, 1050-1060.

Bitouzé, D., Laurent, B. and Massart, P. (1999). A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator. *Ann. Inst. H. Poincaré Probab. Statist.* **35**, 735-763.

Bouaziz, O., Geffray, S. and Lopez, O. (2010). Semi-parametric inference for the recurrent event process by means of a single-index model. arXiv:1005.4553.

Chiang, C.-T., Wang, M.-C. and Huang, C.-Y. (2005). Kernel estimation of rate function for recurrent event data. *Scand. J. Statist.* **32**, 77-91.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events.* Springer, New-York.

Dauxois, J.-Y. and Sencey, S. (2009). Non-parametric tests for recurrent events under competing risks. *Scand. J. Statist.* **36**, 649-670.

Ghosh, D. and Lin, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 554-562.

Ghosh, D. and Lin, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statist. Sinica* **12**, 663-688.

Ghosh, D. and Lin, D. Y. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* **59**, 877-885.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608-1632.

Hansen, B. E. (2005). Exact mean integrated squared error of higher order kernel estimators. *Econom. Theory* **21**, 1031-1057.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158-168.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B Stat. Methodol.* **62**, 711-730.

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

Nelson, W. (1995). Confidence limits for recurrence data – applied to cost or number of product repairs. *Technometrics* **37**, 147-157.

Pepe, M. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Amer. Statist. Assoc.* **88**, 811-820.

Schucany, W. R. (1995). Adaptive bandwidth choice for kernel regression. *J. Amer. Statist. Assoc.* **90**, 535-540.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation.* Springer, New York. (Revised and extended from the 2004 French original, translated by Vladimir Zaiats.)

MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints-Pères, 75006 Paris, France.

E-mail: olivier.bouaziz@parisdescartes.fr

MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints-Pères, 75006 Paris, France.

E-mail: fabienne.comte@univ-paris5.fr

Laboratoire de Statistique Theorique et Appliquee (LSTA), Equipe d'Accueil 3124, Bureau 1516-2-09, 4 place Jussieu, 75252 Paris cedex 05, France.

E-mail: agathe.guilloux@upmc.fr

(Received June 2011; accepted May 2012)