ESAIM: PS 20 (2016) 66–94
DOI: 10.1051/ps/2016004

ESAIM: Probability and Statistics
www.esaim-ps.org

EMPIRICAL LIKELIHOOD CONFIDENCE BANDS FOR MEAN FUNCTIONS OF RECURRENT EVENTS WITH COMPETING RISKS AND A TERMINAL EVENT

Jean-Yves Dauxois[1], Alexis Flesch[2] and Davit Varron[2]

Abstract. In this paper, we consider recurrent events with competing risks in the presence of a terminal event and a censorship. We focus our attention on the mean functions which give the expected number of events of a specific type that have occurred up to a time t. Using heuristics from empirical likelihood theory, we propose a method to build simultaneous (in t) confidence regions for these functions. To establish the consistency of this estimation method (as well as its bootstrap calibration), we prove a weak convergence (as stochastic processes) of the associated empirical likelihood ratio processes. Our approach almost entirely relies on empirical process methods. In the proofs, we also establish some results in empirical processes theory that may present some independent interest. Then we carry out a simulation study of our confidence bands, we compare those obtained by empirical likelihood to the ones obtained by bootstrap. Finally, our procedure is applied on a real data set of nosocomial infections in an intensive care unit of a French hospital.

Mathematics Subject Classification. 62N01, 62G15, 60G55.

Received September 17, 2014. Revised February 8, 2016. Accepted February 19, 2016.

1. Introduction

In this paper we consider a data set of nosocomial infections contracted by 7867 patients in an intensive care unit of a French hospital over a period of a decade. This data set has already been introduced and studied by Dauxois and Sencey [15]. For each patient, one knows if and when he contracted a nosocomial infection and what type of infection it was: pneumonia, septicemia, urinary tract infection and several other types of diseases. Each type of infection can affect the same patient several times. We also know if and when he died, and if not, when he left the hospital. This leads us to work on recurrent events with competing risks under random censorship and with a terminal event. The main aim of this paper is to build confidence bands over time for the mean number of infections of one or more types. This will be achieved using empirical likelihood and empirical process methods.The next sections are devoted to place our works in the frame of the existing literature, with an informal introduction of the contributions of our article.

Keywords and phrases. Censoring, competing risks, empirical likelihood, empirical processes, recurrent events, terminal event.

[1] Université de Toulouse – INSA, IMT UMR CNRS 5219, 135 Avenue de Rangueil, 31077 Toulouse cedex 4, France.

[2] Laboratoire de Mathématiques de Besançon, UMR CNRS 6623, Université de Franche-Comte, 16 route de Gray, 25030 Besançon cedex, France. davit.varron@gmail.com

Article published by EDP Sciences

1.1. Recurrent events and competing risks

The models of recurrent events are useful in many fields, like in social science for recurrent periods of unemployment, in reliability for recurrent occurrences of failures on the same device or also in biostatistics as in our case study of nosocomial infections.

Many statistical methods have been used over the last decades to study recurrent events based on Markov models, Martingale theory or Poisson processes. Andersen *et al.* [2] and recently Cook and Lawless [12] give a review of these methods. Some authors have considered the case with no assumption on the dependence structure between events. Among them, Wang and Wells [47], Lin *et al.* [31], Lin and Ying [29], Cai and Schaubel [5] and more recently Du [17] have studied the distribution of the gap time between events. On the other hand, Lawless and Nadeau [27] and Lawless [26] considered the mean function, which gives the mean number of events that have occurred up to a time t. Then Cook *et al.* [13] introduced robust tests of comparison between treatments. It has to be noted that semi-parametric inference in this models has been drawn by Lin *et al.* [30] and Ghosh [18] for accelerated failure time models, and by Lin *et al.* [32] for multiplicative rate or mean models.

In some situations, a terminal event may stop the recurrent event process, this terminal event being dependent from the recurrent event process. In the data set under consideration, the death of the patient stops the observation of nosocomial infections and is clearly dependent on the previous infections he suffered from. The end of the observation can also be due to an independent right censoring mechanism, specifically, as in our case, when the study comes to an end. Note that in other contexts, other causes of independent censoring can be observed like "lost to follow up".

This kind of situation has been first studied by Cook and Lawless [11]. Ghosh and Lin [19] demonstrate the weak convergence of their estimators and propose two sample tests of equality of two mean functions. In a semi-parametric setting, Ghosh and Lin [20] propose inferential procedures in the presence of a terminal event. The case of dependent censoring has also been considered in the references [21, 33].

Finally, there are some situations where the observed events are of multiple types. This is the case in our data set since several types of nosocomial infections can affect a patient: pneumonia, septicemia, urinary tract infection… In the same manner, the failure of a device can be caused by multiple mechanical parts. As an example, the engine, the gearbox or the tires can be a cause of failure for cars. This is also true for recurrent unemployment: one can loose his job for numerous different reasons. Thus, Chen and Cook [6] extend the set-up of recurrent events in presence of a terminal event to the multivariate case, taking into account different possible types of events. They consider two samples comparison tests and focus on the marginal effect of each type of event, using estimators of the specific mean functions associated to each type of event. Chen *et al.* [7] propose semi-parametric inference in an interval-censored setting. Finally, Dauxois and Sencey [15] introduce a model of competing risks for recurrent events in presence of a terminal event and have derived non-parametric tests in this setup. The consistency of their testing procedure (as well as its ability to detect adjacent alternatives) relies on an extension of convergence results (as bivariate stochastic processes) formerly established by Cook and Lawless [11]. The *first contribution* of the present paper is to extend the aforementioned convergence results to the bootstrap versions of those estimators (see Sect. 2.2 below).

1.2. Empirical likelihood

In another respect, empirical likelihood was first introduced by Thomas and Grunkemeier [43] in a setup of survival analysis. Owen [34–36] generalizes their concept to obtain *e.g.* confidence intervals or regions around a point estimator $\hat{\theta}_n$ of a finite dimensional parameter θ. Owen [37] provides a comprehensive overview of this methodology up to that time. One of the main underlying ideas of the empirical likelihood is that one can capture some geometric aspects of the data set (the simplest one being the asymmetry of a univariate sample), by proceeding as follows: slightly (and continuously) unbalance the empirical weights $1/n$ that are the leading weights in the definition of $\hat{\theta}_n$, and consider the set of all values of $\hat{\theta}_n$ under those unbalanced weights. We informally call leading weights the ones that play the main role for proving the asymptotic normality of $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$. One of the advantages of this empirical likelihood method is that one does not need to estimate the variance in order to build the confidence region. Moreover, the shape of those regions strongly depends on the geometry of the data whereas classical central limit theorem gives ellipse-shaped confidence regions. Finally, the confidence band always lies inside of the convex hull defined by the data. Since that time, many papers have considered the advantage of empirical likelihood in almost all the areas of statistic. The case of the regression models has been extensively considered. One can find in Chen and Van Keilegom [9] an extensive review of empirical likelihood method for regression-type inference problems, including parametric, semiparametric, and nonparametric models. The large list of references cited in this paper shows the growing interest of these techniques in the literature. And this has increased more rapidly again these last years. One can cite, among others, the following recent papers: [8,10,24,25,41]. In survival analysis, one can cite Li *et al.* [28] which gives an other overview with focus on two important regression models for survival data: the Cox proportional hazards model and the Accelerated Failure Time model. Other authors have considered nonparametric models like, among others, Adimari [1] and Ren [40] who use empirical likelihood ratio techniques to study the mean under random censorship or Wang and Jing [46] who use empirical likelihood for a class of functionals of survival distribution. Finally, let us mention the important paper from Hjort *et al.* [23] where a generalization of Owen's result is given and allows to consider the case where the traditional assumptions in this area are violated.

The *second contribution* of the present article is that we take advantage of the empirical likelihood methodology in order to build confidence bands for the mean functions of a multiple-type recurrent events process with terminal event and under right censoring. This approach is fully non-parametric and uses Hjort *et al.*'s results [23], as well as empirical processes theory. For a self-contained presentation of that theory, we refer to van der Vaart and Wellner [44], and we will often make a free use of its notations.

It has to be noted that, as far as we know, there doesn't exist any work which introduces such confidence bands in the literature. The only paper with some work in this direction is the paper from Ghosh and Lin [21]. But it considers the case with an unique cause of death (and not multiple causes of death as in our case) and derives only pointwise confidence intervals (see Sect. 2 of their paper). From Theorem 2 of Dauxois and Sencey [15], obtained in the case of multiple causes of death, one could exploit the idea of Ghosh and Lin [21] to obtain confidence bands for the mean functions. But some mathematical developments are still needed to get confidence bands rather than pointwise confidence intervals.

However, it is of interest to compare our empirical likelihood confidence bands with other. From the result of Dauxois and Sencey [15], there is at least two ways which could lead to the construction of an alternative confidence band for the mean function. One can follow the idea of Ghosh and Lin [21] or one can use a bootstrap method justified by our Theorem 2.2, where we prove the weak convergence of the bootstrapped mean function estimators. We have thought more natural and coherent, with what contains this paper, to consider the second way. The comparison by simulation is carried out in Section 4 after some explanations on how one can use this theorem to build confidence bands.

1.3. Organization of the article

The overall organization of the paper is as follows. We state all our main results in Section 2. Then, in Section 3, we apply our results to the data set of nosocomial infections and we obtain confidence bands for the specific mean functions. Next, in Section 4, we conduct a simulation study to check the accuracy of our method on finite samples.The remaining sections are then dedicated to proofs.

Note that we have developed a Python library to build confidence bands like those introduced in this article. The library is distributed under the Creative Commons CC BY-NC-SA licence and can be downloaded at `http://alexisfles.ch/en/research/rec.html`. One has only to download the file "elrec.tar.gz" (forget the library "emplik" which is already included in this compressed file) and follow the instruction given on the above web-page. Python is an open-source programming language, which can be downloaded from `http://www.python.org`.

2. Main result

2.1. The statistical framework

For the sake of simplicity, we shall suppose that only two types of events are observed, since the generalization is straightforward. For $j \in \{1, 2\}$, let $N_j^\star(t)$ denote the total number of events of type j that have occurred up to time t. We make the assumption that the counting processes $N_j^\star(\cdot)$ are almost surely bounded by a constant A. We also suppose the presence of a terminal event D, *a priori* dependent on the $N_j^\star(\cdot)$'s, after which the counting processes cannot jump. Finally, we assume that the observation of the processes suffers from random right-censoring by a random variable C assumed to be independent of the $N_j^\star(\cdot)$'s and D. Write $X := D \wedge C$ for the time after which the process doesn't jump anymore and $\delta := I(D \leq C)$, where $I(\cdot)$ denotes the (logical) indicator function. This last random variable informs us on whether it was the terminating event that stopped the process or the censorship. Writing $N_j(t) := N_j^\star(t \wedge C)$ for $t \geq 0$ and $j \in \{1, 2\}$, we assume that the observed data are i.i.d. replicates of $(N_{i,1}(\cdot), N_{i,2}(\cdot), X_i, \delta_i)$ of $(N_1(\cdot), N_2(\cdot), X, \delta)$, where the considered processes are indexed by $t \in [0, \tau]$ with τ fulfilling $\mathbb{P}(C > \tau)\mathbb{P}(X > \tau) > 0$. The statistical framework is that of estimating the mean functions

$$
\begin{aligned}
t \mapsto \mu_j(t) &:= \mathbb{E}\left(N_j^\star(t)\right) \\
&= \int_{(0,t]} \frac{\mathbb{P}\left(D \geq u\right)}{\mathbb{P}\left(X \geq u\right)} \mathbb{E}\left(\mathrm{d}N_j(u)\right),
\end{aligned}
\tag{2.1}
$$

where (2.1) is a consequence of the independence between C and $(D, N_1^\star, N_2^\star)$. Using this latter representation, Cook and Lawless [11] proposed the following estimator

$$
\begin{aligned}
\widehat{\mu}_j(t) &:= \frac{1}{n}\sum_{i=1}^{n} \int_{(0,t]} \frac{\widehat{S}^-(u)}{\bar{Y}(u)} \mathrm{d}N_{i,j}(u), \text{ where} \\
\overline{Y}(u) &:= \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \geq u\right), \ u \geq 0,
\end{aligned}
\tag{2.2}
$$

and where $\widehat{S}(\cdot)$ is the Kaplan–Meier estimator of $S(\cdot)$, the survival function of D (see, for example, [2] for a definition of $\widehat{S}(\cdot)$). Note that, in the preceding expression, we used the conventional notation

$$
f^-(t) := \overline{\lim_{\epsilon>0,\ \epsilon\to 0}} f(t-\epsilon), \quad \text{for } t > 0, \text{ and } f^-(0) := f(0). \tag{2.3}
$$

We shall also use the notation $f(t^-)$ for $f^-(t)$, when it comes to unburden notations. Let $D([0, \tau])$ be the space of functions on $[0, \tau]$ that are right continuous on $[0, \tau]$ and which admit a limit from the left at every $t \in (0, \tau]$ (in which case the limit superior in (2.3) is a true limit for any $f(\cdot) \in D([0, \tau])$). Throughout this article, the space $D([0, \tau])$ will be implicitly endowed with the supremum norm:

$$
\| f \|_{[0,\tau]} := \sup_{t\in[0,\tau]} | f(t) | .
$$

Dauxois and Sencey [15] proved the following result.

Theorem 2.1 ([15]). *We have, as $n \to \infty$,*

$$
\sqrt{n}\begin{pmatrix} \widehat{\mu}_1(\cdot) - \mu_1(\cdot) \\ \widehat{\mu}_2(\cdot) - \mu_2(\cdot) \end{pmatrix} \xrightarrow{D} \begin{pmatrix} G_1(\cdot) \\ G_2(\cdot) \end{pmatrix}, \ \textit{in } D^2([0, \tau]),
$$

where $G(\cdot) = (G_1(\cdot), G_2(\cdot))$ is a mean zero bivariate Gaussian process on $[0, \tau]$.

To prove that result, the authors used arguments from empirical processes theory (see, *e.g.*, van der Vaart and Wellner [44]) with convergence of martingales. An interesting aspect of their proof (which will be crucial for our work) is that they make use of martingales at only one single step (in the middle of page 666, when they invoke results of Ghosh and Lin [19]). A byproduct of the present work is an alternative proof of Theorem 2.1 without using any martingale argument (see Sect. 5).

2.2. A bootstrap limit theorem for $(\hat{\mu}_1(\cdot), \hat{\mu}_2(\cdot))$

Our first main result is a generalization of Theorem 2.1 to bootstrap versions of the $\widehat{\mu}_j(\cdot)$. Due to some measurability concerns, we have to completely formalize the bootstrap sample as follows:

(1) Consider a probability space $\left(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}\right)$ and a triangular array $(\mathfrak{r}_{i,n})_{n\geq 1,\, i\leq n}$ of mutually independent random variables on $\left(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}\right)$ for which each $\mathfrak{r}_{i,n}$ has the uniform distribution on $\{1,\dots,n\}$.
(2) Define the observation space $\mathcal{Z} := D^2([0,\tau]) \times \mathbb{R}^+ \times \{0,1\}$ endowed with the product Borel σ-algebra, and write $\mathbf{P}_0$ for the law of $Z := (N_1, N_2, X, \delta)$ on $\mathcal{Z}$.
(3) Now consider the product space $\Omega := \mathcal{Z}^{\mathbb{N}} \times \widetilde{\Omega}$ endowed with $\mathbb{P} := \mathbf{P}_0^{\otimes\mathbb{N}} \otimes \widetilde{\mathbb{P}}$ (on the product σ algebra). Given $\omega = ((z_i)_{i\geq 1}, \widetilde{\omega}) \in \Omega$ define

$$Z_i(\omega) = (N_{i,1}(\omega), N_{i,2}(\omega), X_i(\omega), \delta_i(\omega)) := z_i,$$

and

$$Z_{i,n}^B(\omega) = \left(N_{i,1,n}^B(\omega), N_{i,2,n}^B(\omega), X_{i,n}^B(\omega), \delta_{i,n}^B(\omega)\right) := z_{\mathfrak{r}_{i,n}(\widetilde{\omega})}.$$

In other words, the Z_i are represented in a canonical way (being coordinate projections), and each $Z_{i,n}^B$ is constructed by randomly picking (through $\mathfrak{r}_{i,n}$) a coordinate of $(Z_1,\dots,Z_n)$. Then let us denote by $\left(\widehat{\mu}_j^B(\cdot)\right)_{j\in\{1,2\}}$ the estimators $(\widehat{\mu}_j(\cdot))_{j\in\{1,2\}}$ built from the sample $Z_{i,n}^B,\ i=1,\dots,n$. Note that the letter B stands for *bootstrap*. From now on, all the presented results shall be stated in this probabilistic framework.

It is well-known that a very large panel of techniques from empirical processes theory are very well adapted to bootstrap statistics. However, to the best of our knowledge, such an interplay between the bootstrap and martingale methods has not yet been made clear. To bypass that obstacle we will avoid any martingale argument in our generalization. The mathematical rigor (due to a well-known lack of measurability of empirical processes-see [44], pp. 3–4) imposes us to write our result as an *outer almost sure* convergence in law. For more details on this notion we refer to ([44], Thm. 3.6.2, p. 347). These technical details can be ignored from readers that are more interested in the applied part of the our work. In this case, the heuristic of our next theorem is that "for almost all observed sample sequence $(Z_i(\omega))_{i\geq 1} = (z_i)_{i\geq 1}$, we have a convergence in distribution under the sole randomness of the boostrap scheme, namely, that of the random indices $\tilde{\omega} \mapsto \mathfrak{r}_{i,n}(\tilde{\omega})$".

Theorem 2.2. *As $n\to\infty$, we have outer almost surely:*

$$\sqrt{n}\begin{pmatrix}\widehat{\mu}_1^B(\cdot) - \widehat{\mu}_1(\cdot)\\ \widehat{\mu}_2^B(\cdot) - \widehat{\mu}_2(\cdot)\end{pmatrix} \xrightarrow{D} G(\cdot),$$

where $G(\cdot)$ is the limiting bivairate gaussian process of Theorem 2.1.

The proof of Theorem 2.2 is provided in Section 5.

2.3. Asymptotic simultaneous confidence regions by empirical likelihood

Our proposed method to build confidence intervals (or regions) clearly takes advantage of the heuristic exposed in Section 1.2 and can be formalized as follows. Let us write for all t in $[0,\tau]$, all $\theta\in\mathbb{R}$ and $j\in\{1,2\}$:

$$\mathrm{EL}_n^{(j)}(\theta,t) := \max\left\{\prod_{i=1}^n np_i,\ \widehat{\mu}_{j,\boldsymbol{p}}(t) = \theta,\ p_i \geq 0 \text{ and } \sum_{i=1}^n p_i = 1\right\},$$

where

$$\widehat{\mu}_{j,\boldsymbol{p}}(t) := \sum_{i=1}^{n} p_i \int_{(0,t]} \frac{\widehat{S}^{-}(u)}{\bar{Y}(u)} \mathrm{d}N_{i,j}(u), \text{ for } \boldsymbol{p} = (p_1, \ldots, p_n).$$

For fixed $c \in (0,1]$ and $t \in [0,\tau]$, the region $R^{(j)}(c,t) := \left\{\theta \in \mathbb{R},\ \mathrm{EL}_n^{(j)}(\theta,t) \geq c\right\}$ is an interval containing the point estimator $\hat{\mu}_j(t)$, but non necessarily symmetric around $\hat{\mu}_j(t)$.

The preceding method can be extended to build bivariate confidence regions: for fixed $t \in [0,\tau]$ and $\boldsymbol{\theta} := (\theta_1, \theta_2) \in \mathbb{R}^2$, define

$$\mathrm{EL}_n(\boldsymbol{\theta}, t) := \max\left\{\prod_{i=1}^{n} np_i,\ \forall j \in \{1,2\}\ \widehat{\mu}_{j,\boldsymbol{p}}(t) = \theta_j,\ p_i \geq 0 \text{ and } \sum_{i=1}^{n} p_i = 1\right\}.$$

In that case, for fixed $c \in (0,1]$ and $t \in [0,\tau]$, the region $R(c,t) := \left\{\boldsymbol{\theta} \in \mathbb{R}^2,\ \mathrm{EL}_n(\boldsymbol{\theta},t) \geq c\right\}$ is a bounded region contained in $R^{(1)}(c,t) \times R^{(2)}(c,t)$ which is not necessarily an ellipsoid.

Our second result shows that, given $0 < \tau_1 \leq \tau$, it is possible, for fixed $\alpha \in (0,1)$, to calibrate c_α so that all the $R(c_\alpha, t)$, for $t \in [\tau_1, \tau]$, define *simultaneous* confidence regions. Our only assumption is a moment condition upon the following centered random variables

$$\mathbf{m}_i(t) := (m_{i,1}(t), m_{i,2}(t))^t,\ \text{with } m_{i,j}(t) := \int_{(0,t]} \frac{S(u^-)}{\mathbb{E}\left(Y(u)\right)} \mathrm{d}N_{i,j}(u) - \mu_j(t).$$

Theorem 2.3. *Let $V(t)$ be the covariance matrix of $\mathbf{m}_1(t)$ and Θ the unit sphere in $\mathbb{R}^2$. Suppose that there exists $\tau_1 > 0$ such that*

$$0 < \inf_{t \in [\tau_1,\tau]} \inf_{\theta \in \Theta} \mathbb{E}\left(\theta^t \mathbf{m}_1(t)\right) \text{ and } \sup_{t \in [\tau_1,\tau]} \sup_{\theta \in \Theta} \theta^t V(t) \theta < \infty. \tag{2.4}$$

Then, as $n \to \infty$:

$$-2 \log EL_n(\boldsymbol{\mu}(\cdot), \cdot) \xrightarrow{D} G(\cdot)^t V^{-1}(\cdot) G(\cdot),\ \text{ in } D([\tau_1,\tau]),$$

where $\boldsymbol{\mu}(\mathbf{t}) := (\mu_1(t), \mu_2(t))$.

Now make the weaker assumption that, for fixed $j \in \{1,2\}$, the functions $V_j(\cdot) := \mathbb{V}ar(m_{1,j}(\cdot))$ and $\mathbb{E}\left(|m_{1,j}(\cdot)|\right)$ are bounded away from zero and infinity on $[\tau_1,\tau]$, for some $\tau_1 > 0$. Then, as $n \to \infty$:

$$-2 \log EL_n^{(j)}(\mu_j(\cdot), \cdot) \xrightarrow{D} G_j^2(\cdot) V_j^{-1}(\cdot),\ \text{ in } D([\tau_1,\tau]).$$

Remark 2.4. The asymptotic validity of empirical likelihood methods relies on a convex hull condition as well as an internal studentization phenomenon. Condition (2.4) is essential for the latter to happen. The lower bound can be roughly interpreted as "all the $\mathbf{m}_1(t)$, for $t \in [\tau_1,\tau]$ are far enough from degeneracy".

2.4. Bootstrapping the empirical likelihood

The main statistical impact of Theorem 2.3 is that

$$\inf_{t \in [\tau_1,\tau]} \mathrm{EL}_n(\boldsymbol{\mu}(t), t) \xrightarrow{D} C := \exp\left(-\frac{\|\ G(\cdot)^t V^{-1}(\cdot) G(\cdot)\ \|_{[\tau_1,\tau]}}{2}\right).$$

Hence, if we explicitly knew the law of C (through that of the underlying Gaussian process) then, with a choice of c_α such that $\mathbb{P}\left(C \leq c_\alpha\right) = \alpha$, all the $R(c_\alpha, t)$, for $t \in [\tau_1,\tau]$, would give us a simultaneous confidence regions with asymptotic uniform level less than α. To estimate the limiting distribution of

$$U := \sup_{t\in[\tau_1,\tau]} -2\log \mathrm{EL}_n\left(\boldsymbol{\mu}(t),t\right),$$

we propose the following bootstrap procedure.

Recalling the notations of Subsection 2.2, from an observed sample $(z_1,\dots,z_n)$ let us define $\hat{S}^B(\cdot)$ and $\overline{Y}^B(\cdot)$ as the analogues of $S(\cdot)$ and $\overline{Y}(\cdot)$ with the formal replacement of Z_i by $Z^B_{i,n}$. Then define the following random variables on $(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{\mathbb{P}})$ (the dependency upon $(z_1,\dots,z_n)$ being sometimes omitted to unburden the notations):

$$\begin{aligned}
\widehat{\mu}^B_{j,\boldsymbol{p}}(t) &:= \sum_{i=1}^n p_i \int_{(0,t]} \frac{\hat{S}^B-(u)}{\overline{Y}^B(u)} \mathrm{d}N^B_{i,j,n}(u)\\
U^B_{(z_1,\dots,z_n)} &:= \sup_{t\in[\tau_1,\tau]} -2\log \mathrm{EL}^B_n\left(\widehat{\boldsymbol{\mu}}(t),t\right), \text{where}\\
\mathrm{EL}^B_n(\boldsymbol{\theta},t) &:= \max\left\{\prod_{i=1}^n np_i,\ \forall j\in\{1,2\},\ \widehat{\mu}^B_{j,\boldsymbol{p}}(t)=\theta_j,\ p_i\ge 0 \text{ and } \sum_{i=1}^n p_i = 1\right\}, \boldsymbol{\theta}\in\mathbb{R}^2,\ t\in[0,\tau],
\end{aligned}$$

and where $\widehat{\boldsymbol{\mu}}(\cdot) := (\widehat{\mu}_1(\cdot),\widehat{\mu}_2(\cdot))$ is considered as non random (*i.e.* built from $(z_1,\dots,z_n)$). Also define, for $j\in\{1,2\}$:

$$\begin{aligned}
U^{B,j}_{(z_1,\dots,z_n)} &:= \sup_{t\in[\tau_1,\tau]} -2\log \mathrm{EL}^{B,j}_n\left(\widehat{\mu}_j(t),t\right), \text{where}\\
\mathrm{EL}^{B,j}_n(\theta,t) &:= \max\left\{\prod_{i=1}^n np_i,\ \ \widehat{\mu}^B_{j,\boldsymbol{p}}(t)=\theta,\ p_i\ge 0 \text{ and } \sum_{i=1}^n p_i = 1\right\}, \theta\in\mathbb{R},\ t\in[0,\tau].
\end{aligned}$$

Our next result shows the bootstrap consistency of U^B.

Theorem 2.5. *Under the assumptions of Theorem 2.3, for almost any sequence $z_i := Z_i(\omega),\ i\ge 1$, we have, as $n\to\infty$:*

$$U^B_{(z_1,\dots,z_n)} \xrightarrow{D} ||\ G(\cdot)^t V^{-1}(\cdot) G(\cdot)\ ||_{[\tau_1,\tau]}$$

and

$$U^{B,j}_{(z_1,\dots,z_n)} \xrightarrow{D} ||\ V_j^{-1}(\cdot) G_j^2(\cdot)\ ||_{[\tau_1,\tau_2]},\ \textit{for } j\in\{1,2\},$$

where the aforementioned random variables have baseline probability space $(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{\mathbb{P}})$.

Note that we do not state here a convergence of stochastic processes, but the simpler convergence of their suprema. This is again due to technical measurability conditions: denoting by $BL_1(\mathbb{R})$ the set of 1-Lipschitz functions on $\mathbb{R}$ that are bounded by 1, the maps

$$(z_1,\dots,z_n) \to \sup_{h\in BL_1(\mathbb{R})} \left| \mathbb{E}\left(h\left(U^B_{(z_1,\dots,z_n)}\right)\right) - \mathbb{E}\left(h\left(||\ G(\cdot)^t V^{-1}(\cdot) G(\cdot)\ ||_{[\tau_1,\tau]}\right)\right)\right|,\ n\ge 1$$

are respectively measurable from $\mathcal{Z}^n$ to $\mathbb{R}^+$ (the latter endowed with their Borel σ-algebras). Hence, almost sure convergence is here equivalent to (and not just weaker than) "outer almost sure convergence". This is of particular importance, since our proof of Theorem 2.5 will rely on arguments that are almost sure properties.

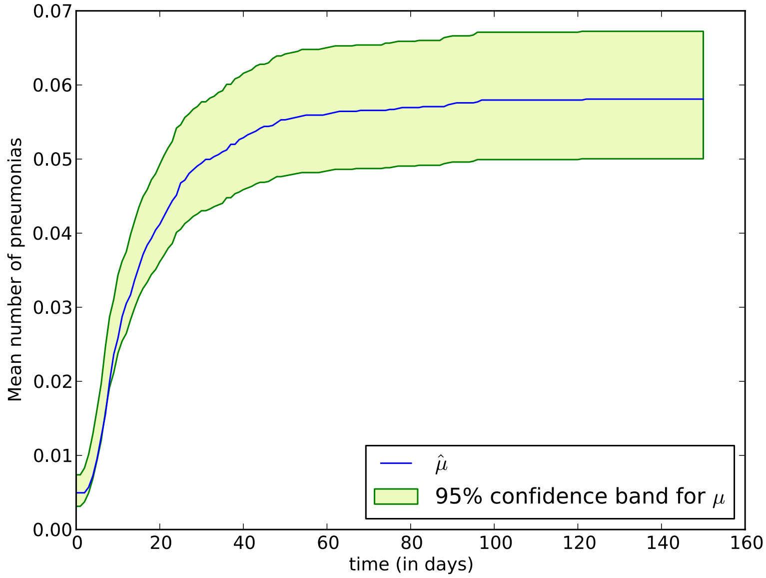

FIGURE 1. Nosocomial infections dataset. Estimated mean number of pneumonias and 95% confidence band (based on empirical likelihood ratio) plotted against time.

3. Application to nosocomial infections

3.1. Algorithm in dimension 1

In this section, we will apply our results to the data set of nosocomial infections introduced in Section 1. We will restrict our attention to pneumonias, septicemias and urinary tract infections, as those are the most frequent infections in the data set: for example, 373 patients experienced a total amount of 463 pneumonias during their stay. Using Theorems 2.3 and 2.5, one can build simultaneous confidence intervals for the mean function of each type of infection, using the bootstrap to determine the limiting law. For sake of completeness, let us precisely describe the algorithm used to obtain a 95% confidence band for the mean function of one type of infection. Using the same notations as before, we have $n = 7867$ and $\max_{1\leq i\leq n} X_i = 380$. First of all, we approximate the 95%-quantile of the limiting law by the bootstrap approximation that was proved consistent in Theorem 2.5. To this effect, we draw $(Z^B_{1,n}, \dots, Z^B_{n,n})$ uniformly in $\{Z_1, \dots, Z_n\}$. Then, given $j \in \{1, 2\}$, we compute $-2\log \mathrm{EL}^{B,j}_n(\hat{\mu}(k), k)$ for each day $k \in [\![0, 380]\!]$, and we store the supremum (over k) of those values. Finally, we repeat this 1000 times (performing a Monte Carlo simulation), sort the values and take the 950th as our 95%-estimated quantile $\hat{u}_{95}$. Then, given $j \in \{1, 2\}$, for each day k in $[\![0, 380]\!]$, we determine by dichotomy the interval C_k defined by:

$$\theta \in C_k \Leftrightarrow -2\log \mathrm{EL}^{(j)}_n(\theta, k) \leq \hat{u}_{95}. \tag{3.1}$$

Notice that the confidence band stays the same after the last jump: for example, no patient contracted a pneumonia after the 125th day as can be seen on Figure 1. Notice also that the confidence band is not centered around $\hat{\mu}$: unlike central limit theorem, empirical likelihood builds confidence regions whose shape strongly depends on the geometry of the data.

3.2. Algorithm in dimension 2

In this section, we show an application of our bidimensional results of Theorems 2.3 and 2.5 to our data set, by building a confidence tube for (μ_1, μ_2). For illustration purpose, we have arbitrarily considered μ_1 as the mean number of pneumonias and μ_2 the mean number of septicemias. In this context, the number of patients

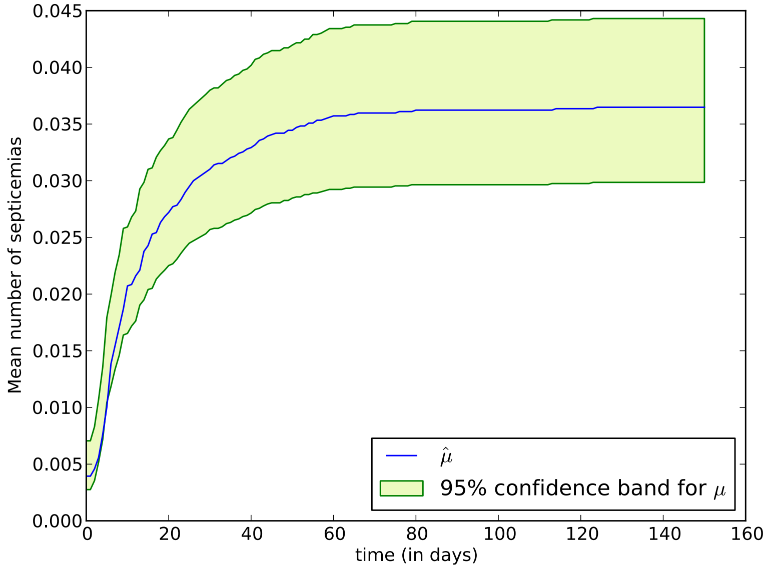

FIGURE 2. Nosocomial infections dataset. Estimated mean number of septicemias and 95% confidence band (based on empirical likelihood ratio) plotted against time.

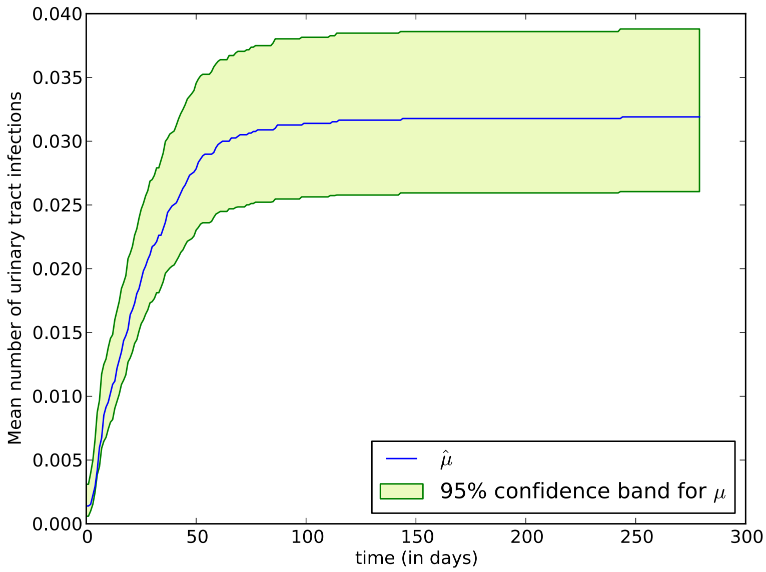

FIGURE 3. Nosocomial infections dataset. Estimated mean number of urinary tract infections and 95% confidence band (based on empirical likelihood ratio) plotted against time.

who experienced either pneumonia or septicemia is 547, with 71 patients experiencing both types of infections at least one time. The total number of pneumonias is 463 (373 patients experienced it) and the total amount of septicemias is 289. We use the same algorithm procedure as in the preceding section, with the formal replacement of $\mathrm{EL}_n^{(j)}(\cdot,\cdot)$ by $\mathrm{EL}_n(\cdot,\cdot)$ (as well as for their bootstrap versions). Hence, for each k, the set C_k defined in (3.1) is not an interval anymore, but a bivariate region, and plotting C_k against k is a 3D graph. In order to make

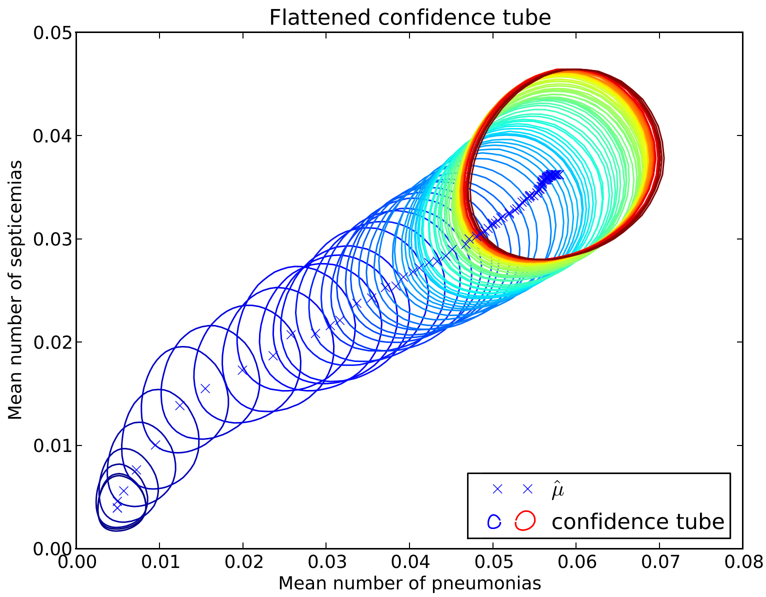

FIGURE 4. Nosocomial infections dataset. Joint estimate of the mean number of pneumonias and septicemias and 95% confidence band (based on empirical likelihood ratio) plotted in function of time (in days).

that graph more readable, we projected it on a plane, eliminating the dimension corresponding to time k, but keeping the information upon k by introducing a progressive change of colors. Each convex set corresponds to a day, the first day being at the bottom left corner and the last day at the upper right corner. As in dimension 1, one can notice that this confidence tube is not elliptic as would be expected if we had used the functional central limit theorem. Once again, this is due to empirical likelihood procedure (Fig. 4).

4. Simulation study

In order to assess the performance of empirical likelihood method to build confidence bands for the mean function $\mu(\cdot)$, we have driven a simulation study. We will restrict our attention to the case where the recurrent events are of a single type. This is easier and however sufficient to understand how this empirical likelihood ratio method works on finite samples. Let $N(\cdot)$ denote the counting process of the number of events. Following Dauxois and Sencey [15], we use a bivariate exponential distribution to simulate the counting process N and the terminal event D. Recall that (X_1, X_2) is bivariate exponentially distributed with parameters $(\lambda_1, \lambda_2, \lambda_{12})$ if and only if there exists $U_1 \sim \mathcal{E}(1/\lambda_1)$, $U_2 \sim \mathcal{E}(1/\lambda_2)$ and $U_{12} \sim \mathcal{E}(1/\lambda_{12})$ mutually independent such that $X_1 = U_1 \wedge U_{12}$ and $X_2 = U_2 \wedge U_{12}$. We shall write $(X_1, X_2) \sim \mathrm{BVE}\,(\lambda_1, \lambda_2, \lambda_{12})$.

Let $T_1 < T_2 < \ldots < T_n < \ldots$ denote the successive occurrences times of the recurrent event process. The random vector (T_1, D) is drawn according to the $\mathrm{BVE}\,(\lambda_1, \lambda_D, \lambda_{12})$. The others jumps of N are then simulated such that $(T_i, T_{i+1} - T_i) \sim \mathrm{BVE}\,(\lambda_1, \lambda_2, \lambda_{12})$. The parameters are chosen in order to have $\mathrm{corr}(T_i, T_{i+1} - T_i) = \mathrm{corr}(D, T_1) = \rho$. It then follows that:

$$\lambda_1 = \lambda_2 = \frac{1-\rho}{1+\rho} \quad \text{and} \quad \lambda_{12} = \frac{2\rho}{1+\rho}\cdot$$

Two values of ρ are selected: $\rho = 0$ and $\rho = 0.25$.

Table 1. Simulation results. Monte Carlo estimates of the confidence band levels obtained by the *algorithm in dimension* 1 (resp. *by direct bootstrap algorithm*, in italic and between parentheses) for selected values of the sample size n, the mean number of events, the percentage of censoring, the correlation ρ and the nominal level ($\alpha_1 = 1\%$, $\alpha_5 = 5\%$ and $\alpha_{10} = 10\%$).

		Mean number of events: 1			
		$\rho = 0$		$\rho = 0.25$	
		0%	30%	0%	30%
	α_1	1.40 (7.51)	1.16 (10.41)	2.12 (4.46)	1.91 (6.77)
$n = 30$	α_5	6.13 (12.26)	5.78 (15.98)	6.92 (8.95)	6.18 (11.82)
	α_{10}	11.46 (17.61)	11.59 (21.98)	12.18 (14.55)	11.50 (17.36)
	α_1	1.23 (3.11)	0.93 (4.44)	1.12 (2.00)	1.00 (2.71)
$n = 100$	α_5	5.31 (7.41)	5.05 (9.51)	5.31 (6.03)	5.34 (7.29)
	α_{10}	10.66 (12.52)	10.47 (15.32)	10.06 (11.17)	10.70 (12.72)
		Mean number of events: 2			
		$\rho = 0$		$\rho = 0.25$	
		0%	30%	0%	30%
	α_1	0.96 (5.96)	1.09 (11.59)	1.48 (6.04)	1.00 (8.94)
$n = 30$	α_5	4.99 (10.42)	5.18 (16.86)	5.95 (10.56)	5.38 (14.65)
	α_{10}	10.28 (15.85)	10.92 (22.77)	11.12 (15.82)	10.66 (20.6)
	α_1	0.94 (2.55)	0.90 (5.02)	1.00 (2.59)	1.01 (4.29)
$n = 100$	α_5	4.84 (6.75)	4.99 (10.43)	4.70 (7.12)	4.76 (9.54)
	α_{10}	9.98 (11.94)	10.37 (16.35)	9.95 (12.49)	10.55 (14.78)

The censoring random variable C is chosen to be uniformly distributed on an interval $[0, c]$, where c is a constant. The values of λ_D and c are chosen empirically to get approximately 1 or 2 observed events per subject. Finally, two different sample sizes are selected: $n = 30$ and $n = 100$.

We used the *algorithm in dimension* 1 decribed in Section 3.1 to build asymptotic confidence bands, considering three different levels: 1%, 5% and 10%.

These procedures have been replicated 10 000 times and we have calculated the percentage of obtained confidence intervals containing the true value of the parameter, giving us an empirical estimation of the confidence band level. The results are listed in Table 1. One can see that the empirical level generally becomes closer to the nominal level when one increases the sample size or the mean number of events. It is interesting to note that this is also the case when the percentage of censoring increases. This kind of phenomenon, which at a first glance appears to be paradoxical, has also been observed by Dauxois and Sencey [15]. It is nothing but the fact that the estimation is easier when the observation of the recurrent process is stopped by an independent right censoring rather than by a dependent terminal event.

Finally, one can observe that the empirical levels appear to be slightly greater than the nominal values. This should be, at least partially, due to the bootstrap procedure used to estimate the limiting distribution of the empirical likelihood ratio. Indeed, it may happen that the bootstrap sample $(Z^B_{1,n}, \dots Z^B_{n,n})$ does not contain the process N_{i_0} which jumps the first among all the initial simple counting processes $(N_i)_{1 \le i \le n}$. Thus the domain where is searched the supremum of the empirical likelihood ratio is reduced. This induces a negative bias on this supremum and leads to a positive bias on the confidence band level.

We have compared our empirical likelihood confidence bands built through the *algorithm in dimension* 1 to another type of bootstrapped confidence bands that one can obtain thanks to our Theorem 2.2. Let us now explain this other algorithm - which we will call the *direct bootstrap algorithm*- recalling notations of Section 2.2, from the simulation of a sample $(Z_1, \dots, Z_n)$, we derive the estimated mean function $\hat{\mu}(\cdot)$. Then, we draw 1000 bootstrap samples $(Z^B_{1,n}, \dots, Z^B_{n,n})$ from $(Z_1, \dots, Z_n)$. For each sample, we calculate the estimator $\hat{\mu}^B(\cdot)$ and the supremum of the process $|\hat{\mu}^B(\cdot) - \hat{\mu}(\cdot)|$. With u_α defined as the α-quantile of the family of those 1000 supremums, a $(1-\alpha)$-confidence band for $\mu(\cdot)$ is given by the two processes: $\hat{\mu}(\cdot) \pm u_{1-\alpha}$.

This procedure has also been replicated 10 000 times and we have obtained an empirical estimation of the confidence band level as before. These empirical confidence band levels are listed between parentheses in the same Table 1. One can observe that, as it is rather frequently observed in the literature, the levels are better with Empirical Likelihood rather than directly by bootstrap. The two usual heuristic explanations of this phenomenon can again be served here: the internal studentization of the procedure allows to avoid the variance estimation phase and the Empirical Likelihood approach captures the first order asymmetry of the sample. On the other hand, as expected, the empirical levels are improved when one increases n, but the convergence appears to be slower than with Empirical Likelihood. Finally, one can see that, when one changes the percentage of censoring, the empirical level of the bootstrapped confidence band behaves oppositely than with empirical likelihood. Here, the empirical level is worse under censoring than without. The reasons of this contradictory phenomenon could be the followings. As already mentioned by Dauxois and Sencey, increasing the percentage of censoring improves slightly the consistency of the estimator. On the opposite, resampling is negatively affected by the presence of censoring since the real percentage of censoring in the initial sample is often far away from the theoretical percentage. But this negative phenomenon seems to be attenuated by the Empirical Likelihood procedure, probably still for the same reasons: this method possesses the property of auto-normalization and the capability to capture the asymmetric features of the sample.

5. Proof of Theorem 2.2

In parallel with our proof of Theorem 2.2, we will also give another proof of Theorem 2.1 in order to emphasize the nonnecessity of martingales in the present statistical framework. Recall that $\mu_j,\ j=1,2$ have been formally defined in (2.1). To make the link with Section 3, diseases corresponding to $j=1$ and $j=2$ denote pneumonias and septicemias, respectively.

5.1. Plug in representations

Let us first note that (2.1) can be rewritten through the following plug-in expression, as processes indexed by $[0,\tau]$:

$$\begin{aligned}\mu_j(\cdot) &:= \int_{(0,\cdot]} \frac{\mathbb{P}(D\geq u)}{\mathbb{P}(X\geq u)}\mathbb{E}\left(\mathrm{d}N_j(u)\right)\\ &= \int_{(0,\cdot]} \frac{KM(F_0,F_1)^-}{1-(F_0+F_1)^-}\mathrm{d}\widetilde{\mu}_j \qquad (5.1)\\ &=: \phi(\widetilde{\mu}_j,F_0,F_1)(\cdot) \qquad (5.2)\end{aligned}$$

where

$$\begin{aligned}F_\ell(t) &:= \mathbb{P}\left(X\leq t,\delta=\ell\right),\ \text{for } \ell=0,1 \text{ and } t\in[0,\tau],\\ \widetilde{\mu}_j(t) &:= \mathbb{E}(N_j(t)),\ \text{for } j\in\{1,2\} \text{ and } t\in[0,\tau],\end{aligned}$$

and where $KM(F_0,F_1)(\cdot)$ is the image of subdistribution functions (F_0,F_1) by the Kaplan–Meier function KM. This function is defined in details Proposition A.5 of Section A. Note that $KM(F_0,F_1)(\cdot)=\mathbb{P}\left(D\geq\cdot\right)$.

Similarly, write

$$\begin{aligned}F_{n,0}(t) &:= \frac{1}{n}\sum_{i=1}^n(1-\delta_i)\mathbb{1}_{[0,t]}(X_i),\ \text{for } t\in[0,\tau],\\ F_{n,1}(t) &:= \frac{1}{n}\sum_{i=1}^n\delta_i\mathbb{1}_{[0,t]}(X_i),\quad \text{for } t\in[0,\tau],\\ \overline{N}_j(t) &:= \frac{1}{n}\sum_{i=1}^n N_{i,j}(t),\ \text{for } j\in\{1,2\} \text{ and } t\in[0,\tau],\end{aligned}$$

and note that $KM(F_{n,0}, F_{n,1})(\cdot) = \hat{S}(\cdot)$. Then, by definition of $\hat{\mu}_j(\cdot)$ in (2.2), we have almost surely:

$$\hat{\mu}_j(\cdot) = \boldsymbol{\phi}\left(\overline{N}_j, F_{n,0}, F_{n,1}\right)(\cdot).$$

Similarly, the same representation holds for the bootstrap versions. Namely, for any $n \geq 1$, we have:

$$\hat{\mu}_j^B(\cdot) = \boldsymbol{\phi}\left(\overline{N}_j^B, F_{n,0}^B, F_{n,1}^B\right)(\cdot)$$

with

$$\begin{aligned}
F_{n,0}^B(t) &:= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_{i,n}^B)\mathbb{1}_{[0,t]}\left(X_{i,n}^B\right), \text{ for } t \in [0,\tau],\\
F_{n,1}^B(t) &:= \frac{1}{n}\sum_{i=1}^{n}\delta_{i,n}^B\mathbb{1}_{[0,t]}\left(X_{i,n}^B\right), \text{ for } t \in [0,\tau],\\
\overline{N}_j^B(t) &:= \frac{1}{n}\sum_{i=1}^{n}N_{i,j,n}^B(t), \text{ for } j \in \{1,2\} \text{ and } t \in [0,\tau].
\end{aligned}$$

Such plug-in representations clearly show the importance of convergence result for empirical processes and their bootstrap analogues. This is the subject of the following section.

5.2. Some convergence results for empirical processes and their bootstrap version

The following result will be the very base of our methodology.

Proposition 5.1. *The following assertions hold.*

(I) *We have, in* $D^4([0,\tau])$*:*

$$\sqrt{n}\begin{pmatrix}\overline{N}_1(\cdot)-\widetilde{\mu}_1(\cdot)\\ \overline{N}_2(\cdot)-\widetilde{\mu}_2(\cdot)\\ F_{n,0}(\cdot)-F_0(\cdot)\\ F_{n,1}(\cdot)-F_1(\cdot)\end{pmatrix} \xrightarrow{D} \mathbf{G}(\cdot),$$

where $\mathbf{G}(\cdot)$ *is a* $\mathbb{R}^4$ *valued Gaussian process.*

(II) *In addition, all the four aforementioned processes between parentheses have (measurable)* $||\cdot||_{[0,\tau]}$ *norms which are almost surely* $o(1)$*.*

(III) *Moreover we have*

$$\sqrt{n}\begin{pmatrix}\overline{N}_1^B(\cdot)-\overline{N}_1(\cdot)\\ \overline{N}_2^B(\cdot)-\overline{N}_2(\cdot)\\ F_{n,0}^B(\cdot)-F_{n,0}(\cdot)\\ F_{n,1}^B(\cdot)-F_{n,1}(\cdot)\end{pmatrix} \xrightarrow{D} \mathbf{G}(\cdot),$$

outer almost surely.

(IV) *In addition, all the four aforementioned processes between parentheses have (measurable)* $||\cdot||_{[0,\tau]}$ *norms which are almost surely* $o(1)$*.*

The proof of this Theorem clearly uses arguments from empirical processes theory and Vapnik–Chervonenkis (VC) combinatorics. The following preliminary lemma is a generalization of the arguments of Bilias *et al.* [3] and can be of independent interest. Its statement involves notions for which the reader can find definitions in ([44], Chap. 2.6).

Lemma 5.2. *Let $\mathcal{F}$ be a class of real valued functions on a set $\mathcal{X}$. Assume that there exists a totally ordered set $(\mathcal{T},\prec)$ such that one can represent $\mathcal{F}$ by $\mathcal{F}=\{f_t,\ t\in\mathcal{T}\}$ in a way such that*

$$\forall x\in\mathcal{X},\ \forall(t,t')\in\mathcal{T}^2,\ t\prec t'\Rightarrow f_t(x)\leq f_{t'}(x).$$

Then $\mathcal{F}$ is a VC subgraph class of functions on $\mathcal{X}$ with VC dimension equal to 1.

Proof of Lemma 5.2*.* Consider the class of subgraphs of $\mathcal{F}$, namely:

$$\mathcal{C}:=\left\{\{(u,x)\in\mathbb{R}\times\mathcal{X},\ u\leq f_t(x)\}\,,\ t\in\mathcal{T}\right\}.$$

Now take an arbitrary subset $C_0:=\{(u_1,x_1),(u_2,x_2)\}\subset\mathbb{R}\times\mathcal{X}$. We need to show that C is not shattered by $\mathcal{C}$ (see [44], p. 135), *i.e.* that the class $\{C\cap C_0,\ C\in\mathcal{C}\}$ is strictly included in that of all subsets of C_0. Assume that this is untrue. Then their would exist $t\in\mathcal{T}$ such that $u_1\leq f_t(x_1)$ and $u_2>f_t(x_2)$, and their would also exist $t'\in\mathcal{T}$ such that $u_1>f_{t'}(x_1)$ and $u_2\leq f_{t'}(x_2)$. Now since $(\mathcal{T},\prec)$ is totally ordered, we have either $t\prec t'$ or $t'\prec t$. If $t\prec t'$, we have $f_t(x_1)\leq f_{t'}(x_1)$. Hence one cannot have simultaneously $u_1\leq f_t(x_1)$ and $u_1>f_{t'}(x_1)$. Now if $t'\prec t$, one cannot either simultaneously have $u_2>f_t(x_2)$ and $u_2\leq f_{t'}(x_2)$. This leads to a contradiction. □

We will divide the proof of Proposition 5.1 is in two parts: the first one is dedicated to the convergences in law, while the second one is dedicated to the results of Glivenko–Cantelli type.

Proof of (I) *and* (III) *in Proposition* 5.1*.* Denote by $\mathbb{1}_C$ the indicator function of a set C. Note that all the considered processes are empirical processes (or bootstrap empirical processes) indexed by the following classes of functions

$$\begin{aligned}\mathcal{F}_N&:=\left\{f_t:=\{\psi\to\psi(t)\}\,,\ t\in[0,\tau]\right\},\\ \mathcal{F}_\ell&:=\left\{\mathbb{1}_{[0,t]\times\{\ell\}},t\in[0,\tau]\right\},\ \text{for } \ell=0,1,\end{aligned}$$

where the elements of $\mathcal{F}_N$ are evaluation maps on $\mathcal{X}:=D^+([0,\tau],A)\subset D([0,\tau])$, defined as the set of all increasing functions on $[0,\tau]$ that are bounded by A. Hence, by ([44], Thm. 3.6.2, p. 347), it is sufficient to prove that each of these classes is uniformly bounded and universally Donsker. It is trivially the case for both $\mathcal{F}_0$ and $\mathcal{F}_1$ because they are VC classes of (indicators of) sets. For a definition of a VC class of sets and a VC subgraph class of functions (see, *e.g.* [44], Chap. 2.6). Now $\mathcal{F}_N$ is uniformly bounded by A and satisfies the following monotonicity property:

$$\forall\psi\in D^+([0,\tau],B)\,,\ t\to f_t(\psi)\text{ is increasing on }[0,\tau].$$

Hence, it is a VC subgraph class by Lemma 5.2. This concludes the proof of (I) and (III) of Proposition 5.1. □

In order to prove (IV) of Proposition 5.1, we first have to introduce a result of almost sure convergence for bootstrap empirical measures. Such a result is clearly new in regard of the existing literature on bootstrap empirical measures and may present an interest in itself. For a definition of a pointwise measurable class, we refer to ([44], p. 110). We will also use the notation

$$J(\mathcal{F}):=\int_0^\infty\sup_{Q\text{ probab. measure}}\sqrt{\log N(\epsilon,\mathcal{F},Q)}\mathrm{d}\epsilon,$$

where, for a probability measure Q, the integer $N(\epsilon,\mathcal{F},Q)$ denotes the minimal number of balls with radius $\epsilon>0$ for the $L^2(Q)$ norm that are needed to cover the class $\mathcal{F}$.

Proposition 5.3. *Let $\mathcal{F}_{(z_1,\ldots,z_n)}$, for $n \geq 1$ and $(z_1,\ldots,z_n) \in \mathcal{Z}^n$, be a collection of classes of functions such that $\mathcal{F}_{(z_1,\ldots,z_n)}$ is pointwise measurable, for each $n \geq 1$, $(z_1,\ldots,z_n) \in \mathcal{Z}^n$, and such that*

$$\sup_{z\in\mathcal{Z},\, f\in\mathcal{F}_{(z_1,\ldots,z_n)}} \mid f(z) \mid \leq C_0(z_1,\ldots,z_n), \tag{5.3}$$

from some measurable function $C_0(\cdot)$ fulfilling, as $n \to \infty$:

$$\frac{C_0(Z_1,\ldots,Z_n)}{n^{1-\alpha}} \to 0,\ \mathbb{P}-a.s. \text{ for some } \alpha > 0. \tag{5.4}$$

Assume that, for each $n \geq 1$ and $(z_1,\ldots,z_n) \in \mathcal{Z}^n$, we have

$$J(\mathcal{F}_{(z_1,\ldots,z_n)}) \leq C_1(z_1,\ldots,z_n),$$

for some measurable function $C_1(\cdot)$ fulfilling

$$\lim_{n\to\infty} n^{-1/2} C_1(Z_1,\ldots,Z_n) = 0,\ \mathbb{P}-a.s. \tag{5.5}$$

Then we have

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}_{(Z_1,\ldots,Z_n)}} \left| \frac{1}{n}\sum_{i=1}^n \left[f(Z^B_{i,n}) - \frac{1}{n}\sum_{i=1}^n f(Z_i) \right] \right| \to 0,\ \mathbb{P}-a.s. \tag{5.6}$$

To the best of our knowledge, the most general results on bootstrap empirical measures (see [44], Chap. 3.6 for an overview of them) does not imply an almost sure convergence like in (5.6). They imply only convergences in probability. Hence the first novelty of Proposition 5.3 is an almost sure convergence to 0, which will be needed at further stages of this article. The second innovation is that the class of functions may depend of the (non bootstrap) sample. These two points are, of course, obtained at the price of conditions that are more stringent, but still covering a large range of statistical applications.

Proof of Proposition 5.3. The begining of the proof makes use of Mc Diarmid's bounded difference inequality (see, *e.g.*, [16], p. 7). Recall that the $\mathfrak{r}_{i,n}$ and the probability $\widetilde{\mathbb{P}}$ have been defined in Subsection 2.2. Fix $n \geq 1$ and $(z_1,\ldots,z_n) \in \mathcal{Z}^n$. Now consider the following map on $\mathcal{Z}^n$:

$$g_{(z_1,\ldots,z_n)} : (\widetilde{z}_1,\ldots,\widetilde{z}_n) \to \sup_{f\in\mathcal{F}_{(z_1,\ldots,z_n)}} \left| \frac{1}{n}\sum_{i=1}^n \left[f(\widetilde{z}_i) - \frac{1}{n}\sum_{i=1}^n f(z_i) \right] \right|.$$

By (5.3), this map satisfies the *bounded difference assumption* (see, *e.g.*, [16], p. 7), with bounding constant $2C_0(z_1,\ldots,z_n)/n$, and we can hence apply Mc Diarmid's bounded difference inequality to obtain, for each $t > 0$:

$$\widetilde{\mathbb{P}}\left(\left| g_{(z_1,\ldots,z_n)}\left(z_{\mathfrak{r}_{1,n}},\ldots,z_{\mathfrak{r}_{n,n}}\right) - \mathbb{E}_{\widetilde{\mathbb{P}}}\left(g_{(z_1,\ldots,z_n)}\left(z_{\mathfrak{r}_{1,n}},\ldots,z_{\mathfrak{r}_{n,n}}\right)\right) \right| \geq t \right) \leq 2\exp\left(-\frac{nt^2}{C_0(z_1,\ldots,z_n)} \right). \tag{5.7}$$

Now writing $A_n := \left\{ C_0(Z_1,\ldots,Z_n) \leq n^{1-\alpha/2} \right\}$ and integrating (5.7) on A_n entails (recalling the independence between $(\mathfrak{r}_{i,n})_{i\leq n}$ and $(Z_i)_{i\leq n}$)

$$\begin{aligned} &\mathbb{P}\left(\left\{ \left| g_{(Z_1,\ldots,Z_n)}\left(Z^B_{1,n},\ldots,Z^B_{n,n}\right) - \mathbb{E}_{\widetilde{\mathbb{P}}}\left(g_{(Z_1,\ldots,Z_n)}\left(Z^B_{1,n},\ldots,Z^B_{n,n}\right)\right) \right| \geq t \right\} \cap A_n \right) \\ \leq & 2\mathbb{E}\left(\exp\left(-\frac{nt^2}{C_0(Z_1,\ldots,Z_n)} \right) \mathbb{1}_{A_n} \right) \leq 2\exp\left(-n^{\alpha/2} t^2 \right), \end{aligned}$$

for all large enough n. Hence, by the Borel–Cantelli lemma, in conjunction with

$$\mathbb{P}\left(\bigcup_{n_0\geq 1}\bigcap_{n\geq n_0} A_n\right) = 1,$$

which is guaranteed by (5.4), we only need to show the almost sure convergence to 0 of

$$\mathbb{E}_{\widetilde{\mathbb{P}}}\left(g_{(Z_1,\ldots,Z_n)}\left(Z^B_{1,n},\ldots,Z^B_{n,n}\right)\right) = \mathbb{E}_{\widetilde{\mathbb{P}}}\left(\sup_{f\in\mathcal{F}_{(Z_1,\ldots,Z_n)}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z^B_{i,n}) - \frac{1}{n}\sum_{i=1}^{n} f(Z_i)\right|\right).$$

Fix (once again) $(z_1,\ldots,z_n)\in\mathcal{Z}^n$. Under the probability $\widetilde{\mathbb{P}}$, the random variable

$$M(z_1,\ldots,z_n) := \sup_{f\in\mathcal{F}_{(z_1,\ldots,z_n)}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[f(z_{\mathfrak{r}_{i,n}}) - \frac{1}{n}\sum_{i=1}^{n} f(z_i)\right]\right|$$

is the supremum of an empirical process indexed by $\mathcal{F}_{(z_1,\ldots,z_n)}$, with underlying sample law $n^{-1}\sum_{i=1}^{n}\delta_{z_i}$ (here δ_{z_i} denotes the Dirac point mass at z_i). Hence, using the usual symmetrization and subgaussian chaining arguments (see, *e.g.*, [44], pp. 127–128, but without making the change of variable in the uniform entropy integral at the middle of p. 128), we have

$$\mathbb{E}_{\widetilde{\mathbb{P}}}\left(M(z_1,\ldots,z_n)\right) \leq \frac{\mathfrak{C}}{\sqrt{n}} J\left(\mathcal{F}_{(z_1,\ldots,z_n)}\right) \leq \frac{\mathfrak{C} C_1(z_1,\ldots,z_n)}{\sqrt{n}},$$

where $\mathfrak{C}$ is universal. By (5.5) the latter defines a sequence which is $o(1)$ for $\mathbf{P}_0^{\otimes\mathbb{N}}$ almost any sequence $(z_i)_{i\geq 1}$. This concludes the proof of Proposition 5.3. □

Proof of (II) *and* (IV) *of Proposition 5.1.* Since $\mathcal{F}_N$, $\mathcal{F}_0$, $\mathcal{F}_1$ are universally Donsker, they are also universally Glivenko–Cantelli, which implies (II). Now (IV) can be reformulated as

$$\max\left\{\|\overline{N}^B_1 - \overline{N}_1\|_{[0,\tau]}; \|\overline{N}^B_2 - \overline{N}_2\|_{[0,\tau]}; \| F^B_{n,0} - F_{n,0}\|_{[0,\tau]}; \| F^B_{n,1} - F_{n,1}\|_{[0,\tau]}\right\} \to 0\ \mathbb{P} - a.s.$$

Such a convergence is the consequence of the almost sure convergence result for bootstrap empirical measures established in previous Proposition 5.3. □

5.3. Proofs of Theorems 2.1 and 2.2 by the Functional Delta method

Given Proposition 5.1, Theorems 2.1 and 2.2 will follow from the Functional Delta method with the map $\boldsymbol{\phi}(\cdot,\cdot,\cdot)$ defined in (5.2). Let us first recall some definitions borrowed from ([44], Chap. 3.9).

Definition 5.4. Let $(\mathbb{D},\|\cdot\|_{\mathbb{D}})$ and $(\mathbb{E},\|\cdot\|_{\mathbb{E}})$ be two normed spaces and let $\mathbb{D}_0\subset\mathbb{D}$. Let ϕ be a map defined on a set $D_\phi\subset\mathbb{D}$, taking values in $\mathbb{E}$, and let $\theta\in D_\phi$. We shall say that ϕ is locally uniformly Hadamard differentiable (LUHD) at θ tangentially to $\mathbb{D}_0$ whenever, for each $h\in\mathbb{D}_0$ and for each sequences $(\theta_n)_{n\geq 1}\in D_\phi^{\mathbb{N}}$, $(t_n)_{n\geq 1}\in]0,\infty[^{\mathbb{N}}$, $(h_n)_{n\geq 1}\in\mathbb{D}^{\mathbb{N}}$ fulfilling

$$(\theta_n + t_n h_n)_{n\geq 1}\in D_\phi^{\mathbb{N}},\quad \|\theta_n - \theta\|_{\mathbb{D}}\to 0,\quad \|h_n - h\|_{\mathbb{D}}\to 0,\quad t_n\to 0,$$

we have, for some $\mathrm{d}\phi_\theta(h)\in\mathbb{E}$ depending only on θ and h:

$$\left\|t_n^{-1}\left[\phi(\theta_n + t_n h_n) - \phi(\theta_n)\right] - \phi'_\theta(h)\right\|_{\mathbb{D}} \to 0, \tag{5.8}$$

and when, in addition, $\mathrm{d}\phi_\theta(\cdot)$ defines a continuous linear map on $(\mathbb{D}_0,\|\cdot\|_{\mathbb{D}})$.

We shall say that ϕ is Hadamard differentiable (HD) at θ tangentially to $\mathbb{D}_0$ whenever (5.8) is only required to hold for constant sequences $\theta_n\equiv\theta$.

The LUHD notion, which is clearly stronger than the usual HD notion, is useful since it adapts very well to bootstrap empirical measures (see, e.g [44], p. 379). For ψ in $D([0,\tau])$, we shall denote by $TV(\psi)$ the (possibly infinite) total variation of ψ on $[0,\tau]$. For fixed $M>0$, we shall write

$$\begin{aligned} BV_M([0,\tau]) &:= \{\psi \in D([0,\tau]),\ TV(\psi)\le M\},\\ B_M([0,\tau]) &:= \left\{f\in D([0,\tau]),\ \|\, f\,\|_{[0,\tau]}\le M\right\}. \end{aligned}$$

Using the general results on the functional Delta method for empirical processes and their bootstrap analogues (see [44], Thms. 3.9.5 and 3.9.13), both Theorems 2.1 and 2.2 will be proved if we prove the following proposition.

Proposition 5.5. *For fixed $M>0$ and $\epsilon>0$, the map ϕ defined on*

$$D_{\phi,\epsilon,M} := \left\{(\psi_1,\psi_2,\psi_3,\psi_4)\in BV_M([0,\tau])^4,\ \inf_{[0,\tau]} \mid 1-F_0-F_1 \mid \ge \epsilon\right\}\subset D^4([0,\tau]) \text{ by}$$

$$\phi(\psi_1,\psi_2,\psi_3,\psi_4) := (\boldsymbol{\phi}(\psi_1,\psi_3,\psi_4),\boldsymbol{\phi}(\psi_2,\psi_3,\psi_4))$$

is LUHD at each $(N_1,N_2,F_0,F_1)\in D_\phi$, tangentially to $D^4([0,\tau])$.

Indeed, since we have assumed $\mathbb{P}(C>\tau)\,\mathbb{P}(X>\tau)>0$ and $N_j(\tau)\le A$ almost surely, we have $(\widetilde{\mu}_1,\widetilde{\mu}_2,F_0,F_1)\in D_{\phi,\epsilon,M}$, for some $M>0$ and $\epsilon>0$. Moreover, Proposition 5.1 in conjunction with the fact that each $N_{i,j}$ is a.s. bounded by A entail that the following measurable random variables satisfy, almost surely:

$$\overline{\lim_{n\to\infty}}\max\left\{TV(\overline{N}_j),TV(\overline{N}_j^B)\right\}\le 2A \tag{5.9}$$

and

$$\underline{\lim_{n\to\infty}}\min\left\{\inf_{[0,\tau]} 1-F_{n,1}-F_{n,0},\ \inf_{[0,\tau]} 1-F_{n,1}^B-F_{n,0}^B\right\}>\epsilon/2. \tag{5.10}$$

Hence, the functional Delta method is fully applicable if Proposition 5.5 is proved. Its proof is made by decomposing ϕ into a succession of simpler maps and and a careful application of a chain rule. All the technicalities of this proof are postponed to Section A.

6. Proof Theorems 2.3 and 2.5

6.1. The main tool

Straightforward calculus shows that:

$$\widehat{\mu}_{j,\boldsymbol{p}}(t)=\mu_j(t)\Longleftrightarrow \sum_{i=1}^n p_i\left(\int_{(0,t]}\frac{\widehat{S}(u^-)}{\bar{Y}(u)}\mathrm{d}N_{i,j}(u)-\mu_j(t)\right)=0.$$

Hence, writing

$$\begin{aligned}\widehat{m}_{i,j}(\cdot) &:= \int_{(0,\cdot]}\frac{\widehat{S}(u^-)}{\bar{Y}(u)}\mathrm{d}N_{i,j}(u)-\mu_j(\cdot)\\ &= \boldsymbol{\phi}(N_{i,j},F_{n,0},F_{n,1})-\boldsymbol{\phi}(\tilde{\mu}_j,F_0,F_1),\end{aligned}$$

we have (with $\boldsymbol{\mu}(t):=(\mu_1(t),\mu_2(t))$)

$$\mathrm{EL}_n^{(j)}(\mu_j(t),t)=\max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i\widehat{m}_{i,j}(t)=0,\ p_i\ge 0 \text{ and } \sum_{i=1}^n p_i=1\right\}$$

and

$$\mathrm{EL}_n(\boldsymbol{\mu}(t),t)=\max\left\{\prod_{i=1}^n np_i, \forall j\in\{1,2\}\ \ \sum_{i=1}^n p_i\widehat{m}_{i,j}(t)=0,\ p_i\geq 0 \text{ and } \sum_{i=1}^n p_i=1\right\}.$$

From now on we will focus on proving the bivariate parts of Theorems 2.3 and 2.5. The proofs of the univariate parts follow exactly the same lines (with sometimes even simpler arguments). Hence we will omit those proofs for sake of brevity.

The main ingredient of our proof is a general result in empirical likelihood theory, due to Hjort, McKeague and van Keilegom. Unfortunately, their result is not present in the final version of their work (see [23]). Thus, it seems convenient to write it down in the present article (see also [45] for a generalization to the multi sample case). Recall that Θ denotes the Euclidian sphere in $\mathbb{R}^2$.

Theorem 6.1 (Hjort *et al.*, unpublished)**.** *Let $0\leq\tau_1<\tau_2$ and $(W_{i,n})_{n\geq 1,\ i\leq n}$ be a sequence of random elements taking values in $D^2([\tau_1,\tau_2])$. Assume that, for a 2-2 matrix-valued function $V(\cdot)$ fulfilling*

$$0<\inf_{t\in[\tau_1,\tau_2],\ \theta\in\Theta}\theta^tV(t)\theta\leq\sup_{t\in[\tau_1,\tau_2],\ \theta\in\Theta}\theta^tV(t)\theta<\infty,$$

we have, as $n\to\infty$:

$$(\mathrm{A0}^\star):\inf_{\substack{t\in[\tau_1,\tau],\\ \theta\in\Theta}}\ \max_{i\leq n} I\left(\theta^tW_{i,n}(t)>0\right)\xrightarrow{D}1;$$

$$(\mathrm{A1}^\star):\sum_{i=1}^n W_{n,i}\xrightarrow{D}G(\cdot),\ \text{ in } D^2([\tau_1,\tau_2]);$$

$$(\mathrm{A2}^\star):\Big|\Big|\sum_{i=1}^n W_{i,n}(\cdot)^{\otimes 2}-V(\cdot)\Big|\Big|_{[\tau_1,\tau_2]}\xrightarrow{D}0;$$

$$(\mathrm{A3}^\star):\max_{1\leq i\leq n}\Big|\Big|W_{i,n}\Big|\Big|_{[\tau_1,\tau_2]}\xrightarrow{D}0.$$

Then, writing

$$EL_n(t):=\max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_iW_{i,n}(t)=0,\ p_i\geq 1,\ \sum_{i=1}^n p_i=1\right\},$$

we have, as $n\to\infty$

$$-2\log EL_n(\cdot)\xrightarrow{D}G^t(\cdot)V^{-1}(\cdot)G(\cdot),\ \ in\ D([\tau_1,\tau_2]).$$

Note that we took care to reformulate assumptions of Hjort *et al.* [23] by four convergences in law, pointing thus out that the underlying probability space may change with n (which is, in some sense, the case for the bootstrap).

From Theorem 6.1, one can see that Theorem 2.3 will be proved if we verify $(\mathrm{A0}^\star)$–$(\mathrm{A3}^\star)$ for the triangular array of bivariate processes $W_{i,n}(\cdot):=n^{-1/2}(\widehat{m}_{i,1}(\cdot),\widehat{m}_{i,2})$. Similarly, Theorem 2.5 will be proved if we show that, for almost any sequence $z_i:=Z_i(\omega),\ i\geq 1$, the sequence of processes (with baseline probability space $(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{\mathbb{P}})$)

$$W_{i,n}^B(\cdot):=n^{-1/2}\left(\widehat{m}_{i,1,n}^B(\cdot),\widehat{m}_{i,2,n}^B(\cdot)\right),$$

where

$$\begin{aligned}\widehat{m}_{i,j,n}^B(\cdot)&:=\int_{(0,\cdot]}\frac{KM(F_{n,0}^B,F_{n,1}^B)(u^-)}{1-(F_{n,0}^B+F_{n,1}^B)(u^-)}\mathrm{d}N_{i,j,n}^B(u)-\widehat{\mu}_j(\cdot)\\&=\boldsymbol{\phi}(N_{i,j,n}^B,F_{n,0}^B,F_{n,1}^B)(\cdot)-\boldsymbol{\phi}(\overline{N}_j,F_{n,0},F_{n,1})(\cdot)\end{aligned}$$

fulfill conditions $(\mathrm{A0}^\star)$–$(\mathrm{A3}^\star)$, as $n\to\infty$.

First, note that (A1*) is a consequence of Theorems 2.1 and 2.2. The proof of the remaining conditions (A0*), (A2*) and (A3*) will take place in three separate propositions, that are given in the next sections.

6.2. Verification of (A2*)

We start with two technical lemmas. Recall that KM stands for the Kaplain–Meier map (see Prop. A.5).

Lemma 6.2. *The map ϕ defined by*

$$\phi(F_0, F_1) := \frac{KM(F_0, F_1)}{1 - (F_0 + F_1)}$$

on the set

$$D_\phi := \left\{(F_0, F_1) \in D\left([\tau_1, \tau]\right);\ \inf_{[\tau_1,\tau]}(1 - F_0 - F_1) > 0\right\}$$

is continuous with respect to $||\cdot||_{[\tau_1,\tau]}$. In addition, $(F_{n,0}, F_{n,1})$ and $(F^B_{n,0}, F^B_{n,1})$ both belong to D_ϕ, $\mathbb{P}$-almost surely, for n large enough.

Proof. The continuity is a straightforward consequence of the continuity of the product integral with respect to $||\cdot||_{[\tau_1,\tau]}$ (see *e.g.* [44], p. 391), while the second assertion of Lemma 6.2 is a consequence of (5.10) in conjunction with $\mathbb{P}\left(D > \tau\right) > 0$. □

Lemma 6.3. *Let $\mathbf{h}$ be a positive and bounded (by a constant C) function of $D([0,\tau])$. Then, for $(\ell,\ell') \in \{0,1\}^2$, the class of functions on $D^{+,2}\left([0,\tau], A\right)$*

$$\mathcal{F}^{\ell,\ell'}_{\mathbf{h}} := \left\{f^{\ell,\ell'}_{t,\mathbf{h}},\ t \in [\tau_1,\tau]\right\},$$

where

$$f^{\ell,\ell'}_{t,\mathbf{h}} : (\psi_1,\psi_2) \to \left(\int_{(0,t]} \mathbf{h}^- \mathrm{d}\psi_1\right)^\ell \times \left(\int_{(0,t]} \mathbf{h}^- \mathrm{d}\psi_2\right)^{\ell'},$$

is uniformly bounded by $(AC)^{\ell+\ell'}$ and is V-C subgraph. Hence, it is universally Donsker.

Proof. The boundedness is immediate, while the VC property is a consequence of Lemma 5.2, taking $\mathcal{T} = [\tau_1,\tau]$ endowed with its natural order. □

We can now prove the following proposition, which is strong enough to check (**A2***), taking $(\ell,\ell') = (1,1)$.

Proposition 6.4. *We have, for each $j,j' \in \{1,2\}$ and $\ell,\ell' \in \{0,1\}^2$, $(\ell,\ell') \neq (0,0)$*

$$\left|\left|\frac{1}{n}\sum_{i=1}^n \widehat{m}^\ell_{i,j}(\cdot)\widehat{m}^{\ell'}_{i,j'}(\cdot) - \mathrm{Cov}\left(m^\ell_{1,j}(\cdot), m^{\ell'}_{1,j'}(\cdot)\right)\right|\right|_{[\tau_1,\tau]} \to 0,\ a.s., \tag{6.1}$$

$$\left|\left|\frac{1}{n}\sum_{i=1}^n \widehat{m}^B\ell_{i,j,n}(\cdot)\widehat{m}^B\ell'_{i,j',n}(\cdot) - \mathrm{Cov}\left(m^\ell_{1,j}(\cdot), m^{\ell'}_{1,j'}(\cdot)\right)\right|\right|_{[\tau_1,\tau]} \to 0,\ a.s. \tag{6.2}$$

Proof. Let us first prove (6.1). Fix $(\ell,\ell')\in\{0,1\}^2\setminus(0,0)$. By the triangle inequality we have

$$\begin{aligned}
&\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\widehat{m}_{i,j}^{\ell}(\cdot)\widehat{m}_{i,j'}^{\ell'}(\cdot)-\mathrm{Cov}\left(m_{1,j}^{\ell}(\cdot),m_{1,j'}^{\ell'}(\cdot)\right)\Big|\Big|_{[\tau_1,\tau]}\\
&\le\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{m}_{i,j}^{\ell}(\cdot)\widehat{m}_{i,j'}^{\ell'}(\cdot)-m_{i,j}^{\ell}(\cdot)m_{i,j'}^{\ell'}(\cdot)\right)\Big|\Big|_{[\tau_1,\tau]}\\
&\quad+\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}m_{i,j}^{\ell}(\cdot)m_{i,j'}^{\ell'}(\cdot)-\mathrm{Cov}\left(m_{1,j}^{\ell}(\cdot),m_{1,j'}^{\ell'}(\cdot)\right)\Big|\Big|_{[\tau_1,\tau]}\\
&=:A_n+B_n.
\end{aligned}\tag{6.3}$$

Let us consider the first term A_n. Elementary algebraic computations (recalling that $(\ell,\ell')\in\{0,1\}^2$) shows that:

$$\mid a^{\ell}b^{\ell'}-c^{\ell}d^{\ell'}\mid\le\mid a^{\ell}\mid\times\mid b-\mathrm{d}\mid+\mid d^{\ell'}\mid\times\mid a-c\mid.$$

Then, we have with probability one

$$A_n\le\max_{i\le n}\Big|\Big|\widehat{m}_{i,j}(\cdot)\Big|\Big|_{[\tau_1,\tau]}^{\ell},\times\max_{i\le n}\Big|\Big|\widehat{m}_{i,j'}(\cdot)-m_{i,j'}(\cdot)\Big|\Big|_{[\tau_1,\tau]}+\max_{i\le n}\Big|\Big|m_{i,j'}(\cdot)\Big|\Big|_{[\tau_1,\tau]}^{\ell'}\times\max_{i\le n}\Big|\Big|\widehat{m}_{i,j}(\cdot)-m_{i,j}(\cdot)\Big|\Big|_{[\tau_1,\tau]}.$$

But, since for each i and j, $N_{i,j}(\cdot)$ is a.s. bounded by A on $[0,\tau]$, we can write:

$$\max_{i\le n}\mid\mid\widehat{m}_{i,j}(\cdot)\mid\mid_{[\tau_1,\tau]}\le\frac{A}{\overline{Y}(\tau)}+\frac{A}{\mathbb{P}\left(X>\tau\right)}\text{ and }\max_{i\le n}\mid\mid m_{i,j}(\cdot)\mid\mid_{[\tau_1,\tau]}\le\frac{2A}{\mathbb{P}\left(X>\tau\right)}.\tag{6.4}$$

Hence, in order to get $A_n\to0$ a.s., it remains to prove the a.s. convergence to 0 of

$$\max_{i\le n}\Big|\Big|\widehat{m}_{i,j'}(\cdot)-m_{i,j'}(\cdot)\Big|\Big|_{[\tau_1,\tau]}\le\Big|\Big|\frac{KM(F_{n,0},F_{n,1})}{1-(F_{n,0}+F_{n,1})}-\frac{KM(F_0,F_1)}{1-(F_0+F_1)}\Big|\Big|_{[\tau_1,\tau]}\times A.$$

This is true by Lemma 6.2 and point (II) of Proposition 5.1.

Now, for the second term B_n, let us first consider the most complicated case where $(\ell,\ell')=(1,1)$. We will follow the notation of Lemma 6.3. Write $\mathbf{h}(\cdot):=S^-(\cdot)/\mathbb{P}\left(X\ge\cdot\right)$. Expanding an empirical covariance and recalling that $\mathbb{E}\left(m_{i,j}(\cdot)\right)\equiv0$, we can split B_n as follows:

$$\begin{aligned}
B_n&=\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\left(f_{\cdot,\mathbf{h}}^{1,0}(N_{i,j},N_{i,j'})-\mathbb{E}\left(f_{\cdot,\mathbf{h}}^{1,0}(N_{i,j},N_{i,j'})\right)\right)\left(f_{\cdot,\mathbf{h}}^{0,1}(N_{i,j},N_{i,j'})-\mathbb{E}\left(f_{\cdot,\mathbf{h}}^{0,1}(N_{i,j},N_{i,j'})\right)\right)\\
&\qquad-\mathbb{E}\left(m_{i,j}(\cdot)m_{i,j'}(\cdot)\right)\Big|\Big|_{[\tau_1,\tau]}\\
&\le\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}f_{\cdot,\mathbf{h}}^{1,1}(N_{i,j},N_{i,j'})-\mathbb{E}\left(f_{\cdot,\mathbf{h}}^{1,1}(N_{i,j},N_{i,j'})\right)\Big|\Big|_{[\tau_1,\tau]} &&(6.5)\\
&\quad+\Big|\Big|\mathbb{E}\left(f_{\cdot,\mathbf{h}}^{0,1}(N_{1,j},N_{1,j'})\right)\Big|\Big|_{[\tau_1,\tau]}\times\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}f_{\cdot,\mathbf{h}}^{1,0}(N_{i,j},N_{i,j'})-\mathbb{E}\left(f_{\cdot,\mathbf{h}}^{1,0}(N_{i,j},N_{i,j'})\right)\Big|\Big|_{[\tau_1,\tau]} &&(6.6)\\
&\quad+\Big|\Big|\mathbb{E}\left(f_{\cdot,\mathbf{h}}^{1,0}(N_{1,j},N_{1,j'})\right)\Big|\Big|_{[\tau_1,\tau]}\times\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}f_{\cdot,\mathbf{h}}^{0,1}(N_{i,j},N_{i,j'})-\mathbb{E}\left(f_{\cdot,\mathbf{h}}^{0,1}(N_{i,j},N_{i,j'})\right)\Big|\Big|_{[\tau_1,\tau]}. &&(6.7)
\end{aligned}$$

This majorant tends to zero almost surely, because of its expression with respect of the suprema of empirical processes indexed by $\mathcal{F}_{\mathbf{h}}^{0,1}$, $\mathcal{F}_{\mathbf{h}}^{1,0}$ and $\mathcal{F}_{\mathbf{h}}^{1,1}$ (recall Lem. 6.3 and note that the expectation terms are bounded by $A/\mathbb{P}(X > \tau)$). Now considering the case where $(\ell, \ell') = (0,1)$ or $(1,0)$, we have

$$B_n = \Big|\Big| \frac{1}{n} \sum_{i=1}^{n} f_{\cdot,\mathbf{h}}^{\ell,\ell'}(N_{i,j}, N_{i,j'}) - \mathbb{E}\left(f_{\cdot,\mathbf{h}}^{\ell,\ell'}(N_{i,j}, N_{i,j'}) \right) \Big|\Big|_{[\tau_1,\tau]},$$

and hence the same arguments apply. This proves (6.1).

Let us now consider (6.2). We use the same kind of inequalities as above, formally replacing Z_i by $Z_{i,n}^B$ and $(\widetilde{\mu}_1, \widetilde{\mu}_2, F_0, F_1)$ by $(\overline{N}_1, \overline{N}_2, F_{n,0}, F_{n,1})$. Hence, writing

$$m_{i,j,n}^B(\cdot) := \phi(N_{i,j,n}^B, F_{n,0}, F_{n,1}) - \phi(\overline{N}_j, F_{n,0}, F_{n,1})$$

for the bootstrap analogue of $m_{i,j}(\cdot)$ and denoting A_n^B and B_n^B the bootstrap analogues of A_n and B_n in (6.3), it will be sufficient to prove that the following points hold almost surely:

$$\overline{\lim_{n\to\infty}} \max_{i\le n} ||\ \widehat{m}_{i,j,n}^B(\cdot)\ ||_{[\tau_1,\tau]} < \infty, \tag{6.8}$$

$$\overline{\lim_{n\to\infty}} \max_{i\le n} ||\ m_{i,j,n}^B(\cdot)\ ||_{[\tau_1,\tau]} < \infty, \tag{6.9}$$

$$\Big|\Big| \frac{KM(F_{n,0}^B, F_{n,1}^B)}{1-(F_{n,0}^B + F_{n,1}^B)} - \frac{KM(F_{n,0}, F_{n,1})}{1-(F_{n,0}+F_{n,1})} \Big|\Big|_{[\tau_1,\tau]} \to 0, \tag{6.10}$$

$$\forall (\ell,\ell') \in \{0,1\}^2 \setminus \{(0,0)\},\ \Big|\Big| \frac{1}{n} \sum_{i=1}^{n} \left(m_{i,j,n}^B(\cdot)\right)^{\ell} \left(m_{i,j,n}^B(\cdot)\right)^{\ell'} - \mathbb{E}\left((m_{1,j,n}(\cdot))^{\ell} (m_{1,j,n}(\cdot))^{\ell'} \right) \Big|\Big|_{[\tau_1,\tau]} \to 0. \tag{6.11}$$

Indeed, the a.s. convergence to 0 of A_n^B will be a consequence of (6.8), (6.9) and (6.10) while the a.s convergence to 0 of B_n^B will be a consequence of (6.11).

First, note that we have, almost surely, for each $n \ge 1$:

$$\max_{i\le n} ||\ \widehat{m}_{i,j,n}^B(\cdot)\ ||_{[\tau_1,\tau]} \le \frac{A}{1-(F_{n,0}^B + F_{n,1}^B)(\tau^-)} + \frac{A}{1-(F_{n,0}+F_{n,1})(\tau^-)}$$

$$\max_{i\le n} \Big|\Big| m_{i,j,n}^B(\cdot) \Big|\Big|_{[\tau_1,\tau]} \le \frac{2A}{1-(F_{n,0}+F_{n,1})(\tau^-)}.$$

Hence, (6.8) and (6.9) are obtained thanks to Lemma 6.2. On the other hand, (6.10) is an immediate consequence of Lemma 6.2 together with point (IV) of Proposition 5.1.

Finally, it remains to prove (6.11). We will use the same arguments as those used to bound B_n. Writing

$$\hat{\mathbf{h}} = \hat{\mathbf{h}}(Z_1, \ldots, Z_n) := \frac{\hat{S}^-}{\overline{Y}},$$

we can assert that the bootstrap analogues of (6.5), (6.6) and (6.7) converge almost surely to zero if

$$\Big|\Big| \frac{1}{n} \sum_{i=1}^{n} f_{\cdot,\hat{\mathbf{h}}}^{1,0}(N_{i,j}, N_{i,j'}) \Big|\Big|_{[\tau_1,\tau]} \times \Big|\Big| \frac{1}{n} \sum_{i=1}^{n} \left[f_{\cdot,\hat{\mathbf{h}}}^{0,1}(N_{i,j,n}^B, N_{i,j',n}^B) - \frac{1}{n} \sum_{i=1}^{n} f_{\cdot,\hat{\mathbf{h}}}^{0,1}(N_{i,j}, N_{i,j'}) \right] \Big|\Big|_{[\tau_1,\tau]} \to 0,$$

$$\Big|\Big| \frac{1}{n} \sum_{i=1}^{n} f_{\cdot,\hat{\mathbf{h}}}^{0,1}(N_{i,j}, N_{i,j'}) \Big|\Big|_{[\tau_1,\tau]} \times \Big|\Big| \frac{1}{n} \sum_{i=1}^{n} \left[f_{\cdot,\hat{\mathbf{h}}}^{1,0}(N_{i,j,n}^B, N_{i,j',n}^B) - \frac{1}{n} \sum_{i=1}^{n} f_{\cdot,\hat{\mathbf{h}}}^{1,0}(N_{i,j}, N_{i,j'}) \right] \Big|\Big|_{[\tau_1,\tau]} \to 0, \text{ and}$$

$$\Big|\Big| \frac{1}{n} \sum_{i=1}^{n} \left[f_{\cdot,\hat{\mathbf{h}}}^{1,1}(N_{i,j,n}^B, N_{i,j',n}^B) - \frac{1}{n} \sum_{i=1}^{n} f_{\cdot,\hat{\mathbf{h}}}^{1,1}(N_{i,j}, N_{i,j'}) \right] \Big|\Big|_{[\tau_1,\tau]} \to 0.$$

But, from (6.1) we have:

$$\Big|\Big|\frac{1}{n}\sum_{i=1}^{n} f^{0,1}_{\cdot,\hat{\mathbf{h}}}(N_{i,j},N_{i,j'})\Big|\Big|_{[\tau_1,\tau]} = \Big|\Big|\frac{1}{n}\sum_{i=1}^{n} \hat{m}_{i,j'}(\cdot)+\mu_{j'}(\cdot)\Big|\Big|_{[\tau_1,\tau]} = O(1),$$

$$\Big|\Big|\frac{1}{n}\sum_{i=1}^{n} f^{1,0}_{\cdot,\hat{\mathbf{h}}}(N_{i,j},N_{i,j'})\Big|\Big|_{[\tau_1,\tau]} = \Big|\Big|\frac{1}{n}\sum_{i=1}^{n} \hat{m}_{i,j}(\cdot)+\mu_{j}(\cdot)\Big|\Big|_{[\tau_1,\tau]} = O(1),$$

Thus, it is sufficient to prove that, for any $(\ell,\ell')\in\{0,1\}^2$, we have, as $n\to\infty$:

$$\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\left[f^{\ell,\ell'}_{\cdot,\hat{\mathbf{h}}}(N^B_{i,j,n},N^B_{i,j',n})-\frac{1}{n}\sum_{i=1}^{n} f^{\ell,\ell'}_{\cdot,\hat{\mathbf{h}}}(N_{i,j},N_{i,j'})\right]\Big|\Big|_{[\tau_1,\tau]}\to 0 \text{ a.s.} \tag{6.12}$$

In this order, let us recall the setup of Proposition 5.3. From Lemma 6.3, we know that, for each $n\geq 1$ and $(z_1,\dots,z_n)\in\mathcal{Z}^n$, the class $\mathcal{F}_{(z_1,\dots,z_n)} := \mathcal{F}^{\ell,\ell'}_{\hat{\mathbf{h}}(z_1,\dots,z_n)}$ is uniformly bounded by

$$C_0(z_1,\dots,z_n) := \left(\frac{A}{\frac{1}{n}\sum\limits_{i=1}^{n} I\left(x_i>\tau\right)}\right)^{\ell+\ell'}.$$

Moreover, since $\mathcal{F}_{(z_1,\dots,z_n)}$ is VC subgraph with dimension 1 and admits the constant $C_0(z_1,\dots,z_n)$ as an envelope, we have

$$J(\mathcal{F}_{(z_1,\dots,z_n)})\leq \mathcal{K}C_0(z_1,\dots,z_n) =: C_1(z_1,\dots,z_n),$$

where $\mathcal{K}$ only depends on the VC dimension of $\mathcal{F}_{(z_1,\dots,z_n)}$ and hence is universal (see [44], p. 141). Now since, we have almost surely

$$C_0^{-1}(Z_1,\dots,Z_n)\to\left(\frac{\mathbb{P}\left(X>\tau\right)}{A}\right)^{\ell+\ell'}>0,$$

conditions (5.4) and (5.5) are satisfied. Hence, an application of Proposition 5.3 proves (6.12) and concludes the proof of Proposition 6.4. □

6.3. Proof of (A3)*

From (6.4), we have

$$\max_{i\leq n}||\ \hat{m}_{i,j}\ ||_{[\tau_1,\tau]} = o(\sqrt{n})$$
$$\max_{i\leq n}||\ \hat{m}^B_{i,j}\ ||_{[\tau_1,\tau]} = o(\sqrt{n})$$

which prove that assumption (A3)* is satisfied for the $W_{i,n}(\cdot)$ and $W^B_{i,n}(\cdot)$.

6.4. Proof of (A0)*

In order to check (A0)* (and hence conclude the proofs of Thms. 2.3 and 2.5), we will use the following lemma.

Lemma 6.5. *Let $(a_1,\dots,a_n)\in\mathbb{R}^n$ and let*

$$\eta := \frac{1}{4n}\sum_{i=1}^{n}|\ a_i-\overline{a}\ | \quad \textit{and } \sigma^2 := \frac{1}{n}\sum_{i=1}^{n}\left(a_i-\overline{a}\right)^2,$$

where $\overline{a} = \sum_{i=1}^{n} a_i/n$. *Then, for all* $\delta > 0$, *we have*

$$\frac{1}{n}\sum_{i=1}^{n} I\left(a_i \geq \delta\right) \geq \frac{\eta^2}{\sigma^2} - I\left(\overline{a} + \eta \leq \delta\right).$$

Proof. Since

$$\frac{1}{n}\sum_{i=1}^{n} I\left(a_i \geq \delta\right) \geq \frac{1}{n}\sum_{i=1}^{n} I\left(a_i \geq \overline{a} + \eta\right) - I\left(\overline{a} + \eta \leq \delta\right),$$

we just need to bound the first term from below. Now, the Cauchy–Schwartz inequality entails

$$\begin{aligned}\sqrt{\frac{1}{n}\sum_{i=1}^{n} I\left(a_i - \overline{a} \geq \eta\right) \times \sigma^2} &\geq \frac{1}{n}\sum_{i=1}^{n}(a_i - \overline{a}) I\left(a_i - \overline{a} \geq \eta\right) \\ &\geq \frac{1}{n}\sum_{i=1}^{n}(a_i - \overline{a}) I\left(a_i - \overline{a} \geq 0\right) - \frac{1}{n}\sum_{i=1}^{n}(a_i - \overline{a}) I\left(0 \leq a_i - \overline{a} \leq \eta\right) \\ &\geq 2\eta - \eta,\end{aligned}$$

where, for the last inequality, we used the identity

$$\sum_{i=1}^{n} \mid a_i - \overline{a} \mid = 2\sum_{i=1}^{n}(a_i - \overline{a}) I\left(a_i - \overline{a} \geq 0\right).$$

□

Now, set $\hat{\mathbf{m}}_i = (\hat{m}_{i,1}, \hat{m}_{i,2})^t$ and $\hat{\mathbf{m}}_{i,n}^B = (\hat{m}_{i,1,n}^B, \hat{m}_{i,2,n}^B)^t$. The condition $(\mathbf{A0})^\star$ will be fulfilled in the two cases under consideration once the following proposition has been established.

Proposition 6.6. *Under assumption* (2.4) *we have, almost surely,*

$$\varliminf_{n\to\infty} \inf_{t\in[\tau_1,\tau],\ \theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} I\left(\theta^t \hat{\mathbf{m}}_i(t) > 0\right) > 0 \tag{6.13}$$

$$\varliminf_{n\to\infty} \inf_{t\in[\tau_1,\tau],\ \theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} I\left(\theta^t \hat{\mathbf{m}}_{i,n}^B(t) > 0\right) > 0, \tag{6.14}$$

which entail

$$\inf_{t\in[\tau_1,\tau],\ \theta\in\Theta} \max_{i\leq n} I\left(\theta^t \hat{\mathbf{m}}_i(t) > 0\right) \to 1$$

$$\inf_{t\in[\tau_1,\tau],\ \theta\in\Theta} \max_{i\leq n} I\left(\theta^t \hat{\mathbf{m}}_{i,n}^B(t) > 0\right) \to 1,$$

as $n \to \infty$

Proof. Fix $n \geq 1, \omega \in \Omega,\ \theta \in \Theta$ and $t \in [\tau_1, \tau]$. An application (for any specified ω) of Lemma 6.5 with $a_i := \theta^t \hat{\mathbf{m}}_i(t,\omega)$ (resp. $a_i = \theta^t \hat{\mathbf{m}}_i^B(t,\omega)$) entails the following almost sure inequalities, for any $\delta > 0$:

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} I\left(\theta^t \hat{\mathbf{m}}_i(t) > 0\right) + I\left(\frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_i(t) + \frac{1}{4n}\sum_{i=1}^{n}\left|\left[\theta^t \hat{\mathbf{m}}_i(t) - \frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_i(t)\right]\right| \leq \delta\right) \\ &\geq \frac{\left(\frac{1}{4n}\sum_{i=1}^{n}\left|\left[\theta^t \hat{\mathbf{m}}_i(t) - \frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_i(t)\right]\right|\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(\theta^t \hat{\mathbf{m}}_i(t) - \frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_i(t)\right)^2}\end{aligned} \tag{6.15}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} I\left(\theta^t \hat{\mathbf{m}}_{i,n}^B(t) > 0\right) + I\left(\frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i,n}^B(t) + \frac{1}{4n}\sum_{i=1}^{n}\left|\left[\theta^t \hat{\mathbf{m}}_{i}^B(t) - \frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i,n}^B(t)\right]\right| \le \delta\right)$$
$$\ge \frac{\left(\frac{1}{4n}\sum\limits_{i=1}^{n}\left|\left[\theta^t \hat{\mathbf{m}}_{i,n}^B(t) - \frac{1}{n}\sum\limits_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i,n}^B(t)\right]\right|\right)^2}{\frac{1}{n}\sum\limits_{i=1}^{n}\left(\theta^t \hat{\mathbf{m}}_{i,n}^B(t) - \frac{1}{n}\sum\limits_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i,n}^B(t)\right)^2}.$$

Now, a repeated use of Proposition 6.4 shows that, almost surely (recalling Assumption (2.4))

$$\varliminf_{n\to\infty}\inf_{t\in[\tau_1,\tau]}\frac{\left(\frac{1}{4n}\sum\limits_{i=1}^{n}\left|\left[\theta^t \hat{\mathbf{m}}_{i}(t) - \frac{1}{n}\sum\limits_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i}(t)\right]\right|\right)^2}{\frac{1}{n}\sum\limits_{i=1}^{n}\left(\theta^t \hat{\mathbf{m}}_{i}(t) - \frac{1}{n}\sum\limits_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i}(t)\right)^2} > \inf_{t\in[\tau_1,\tau]}\frac{\left(\frac{1}{4}\mathbb{E}\left(\mid \theta^t \mathbf{m}_1(t)\mid\right)\right)^2}{\mathbb{E}\left(\left(\theta^t \mathbf{m}_1(t)\right)^2\right)} > 0, \text{ by (2.4)}, \tag{6.16}$$

$$\varliminf_{n\to\infty}\inf_{t\in[\tau_1,\tau]}\frac{\left(\frac{1}{4n}\sum\limits_{i=1}^{n}\left|\left[\theta^t \hat{\mathbf{m}}_{i}^B(t) - \frac{1}{n}\sum\limits_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i}^B(t)\right]\right|\right)^2}{\frac{1}{n}\sum\limits_{i=1}^{n}\left(\theta^t \hat{\mathbf{m}}_{i}^B(t) - \frac{1}{n}\sum\limits_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i}^B(t)\right)^2} > \inf_{t\in[\tau_1,\tau]}\frac{\left(\frac{1}{4}\mathbb{E}\left(\mid \theta^t \mathbf{m}_1(t)\mid\right)\right)^2}{\mathbb{E}\left(\left(\theta^t \mathbf{m}_1(t)\right)^2\right)} > 0, \tag{6.17}$$

$$\varliminf_{n\to\infty}\inf_{t\in[\tau_1,\tau]}\frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i}(t) + \frac{1}{4n}\sum_{i=1}^{n}\left|\left[\theta^t \hat{\mathbf{m}}_{i}(t) - \frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i}(t)\right]\right| \ge \inf_{t\in[\tau_1,\tau]}\mathbb{E}\left(\mid \theta^t \mathbf{m}_1(t)\mid\right) > 0, \tag{6.18}$$

$$\varliminf_{n\to\infty}\inf_{t\in[\tau_1,\tau]}\frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i,n}^B(t) + \frac{1}{4n}\sum_{i=1}^{n}\left|\left[\theta^t \hat{\mathbf{m}}_{i,n}^B(t) - \frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i,n}^B(t)\right]\right| \ge \inf_{t\in[\tau_1,\tau]}\mathbb{E}\left(\mid \theta^t \mathbf{m}_1(t)\mid\right) > 0. \tag{6.19}$$

Now take $\delta := \delta_0/2$, where δ_0 is the infimum on the right hand side of (6.18). This entails

$$I\left(\frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i}(t) + \frac{1}{4n}\sum_{i=1}^{n}\left|\left[\theta^t \hat{\mathbf{m}}_{i}(t) - \frac{1}{n}\sum_{i=1}^{n}\theta^t \hat{\mathbf{m}}_{i}(t)\right]\right| \le \delta\right) \to 0 \tag{6.20}$$

almost surely. Now inject both (6.16) and (6.20) in (6.15), and take the almost sure inferior limit to prove (6.13). Finally (6.14) is proved following the same way with bootstrap versions (6.17) and (6.19). □

Appendix A: proof of Proposition 5.5

The idea to prove Proposition 5.5 is to use the chain (or composition) rule. Such a rule is often essential to prove a HD (see, *e.g.* Van der Vaart and Wellner [44], p. 373). But, this chain rule is not in general sufficient to prove a LUHD property. An additional continuity condition is necessary. This is the aim of the next lemma to clarify this point. Even if its demonstration is obvious (and not detailed here), we state it as a lemma since we could not find any appropriate statement in the existing literature.

Lemma A.1 (Chain rule)**.** *Let ϕ be a map from $D_\phi \subset \mathbb{D}$ to $(\mathbb{E}, ||\cdot||_{\mathbb{E}})$ and let $\theta \in D_\phi$. Assume that ϕ is LUHD at θ tangentially to $\mathbb{D}_0$.*

Let $\tilde{\phi}$ be a map from $D_{\tilde{\phi}} \subset \mathbb{E}$ to a normed space $(\mathbb{F}, ||\cdot||_F)$, and suppose that $\phi(D_\phi) \subset D_{\tilde{\phi}}$. Assume that $\tilde{\phi}$ is LUHD at $\phi(\theta)$ tangentially to $\mathrm{d}\phi_\theta(\mathbb{D}_0)$. Also assume that ϕ is $||\cdot||_{\mathbb{D}}$-continuous at θ. Then $\tilde{\phi}\circ\phi$ is LUHD at θ tangentially to $\mathbb{D}_0$, with derivative

$$\mathrm{d}(\tilde{\phi}\circ\phi)_\theta \equiv \mathrm{d}\tilde{\phi}_{\phi(\theta)} \circ \mathrm{d}\phi_\theta.$$

Our proof of Proposition 5.5 relies on this chain rule. As a consequence, all the following technicalities (from Lem. A.2 to the end of this section) are careful verifications of the applicability of the chain rule. Our first lemma is straightforward and its proof is therefore omitted.

Lemma A.2. *For $M, \epsilon > 0$, define*

$$D^{(1)}_{\epsilon,M} := \left\{ (A,B) \in BV_M([0,\tau])^2,\ \inf_{[0,\tau]} A \geq \epsilon \right\}.$$

The map $\phi^{(1)}(\cdot,\cdot)$ defined on $D_{\phi^{(1)}} := D^{(1)}_{\epsilon,M}$ by

$$\phi^{(1)}:\ (A,B) \to (1/A, B) \in D([0,\tau])^2$$

is continuous on $D^{(1)}_{\epsilon,M}$. It is also LUHD at each $(A,B) \in D^{(1)}_{\epsilon,M}$, tangentially to $D([0,\tau])^2$, with derivative given by

$$\mathrm{d}\phi^{(1)}_{A,B}:\ (h^A, h^B) \to \left(\frac{-h^A}{A^2}, h^B \right) \in D([0,\tau])^2.$$

Moreover we have

$$\phi^{(1)}\left(D^{(1)}_{\epsilon,M}\right) \subset \left\{ (A,B) \in BV_{\frac{M}{\epsilon^2}}([0,\tau]) \times BV_M([0,\tau]),\ \|\ A\ \|_{[0,\tau]} \leq \epsilon \right\} =: D^{(2)}_{\epsilon,M}.$$

Throughout the remainder of this section, we extend the notation $\int_{(0,t]} A\mathrm{d}B$ to the case where B is not of bounded variation but A is, thanks to the following (integration by parts) formula:

$$\int_{(0,t]} A\mathrm{d}B = AB(t) - AB(0) - \int_{(0,t]} B^- dA. \tag{A.1}$$

Lemma A.3. *The map $\phi^{(2)}$ defined on $D_{\phi^{(2)}} := D^{(2)}_{\epsilon,M}$ by*

$$\phi^{(2)}:\ (A,B) \to \int_{(0,\cdot]} A^- \mathrm{d}B \in D([0,\tau])$$

is continuous on $D_{\phi^{(2)}}$. It is also LUHD at each $(A,B) \in D_{\phi^{(2)}}$, tangentially to $D([0,\tau])^2$, with derivative

$$\mathrm{d}\phi^{(2)}_{A,B}:\ (h^A, h^B) \to \int_{(0,\cdot]} h^A \mathrm{d}B + \int_{(0,\cdot]} A\mathrm{d}h^B \in D([0,\tau]).$$

Moreover we have $\phi^{(2)}\left(D^{(2)}_{\epsilon,M}\right) \subset BV_{M/\epsilon}([0,\tau])$.

Proof. Let us write shortly $\int$ instead of $\int_{(0,\cdot]}$. In order to prove the continuity of $\phi^{(2)}$, let us consider a sequence (A_n, B_n) of elements of $D_{\phi^{(2)}}$ such that $\|\ A_n - A\ \|_{[0,\tau]}$ and $\|\ B_n - B\ \|_{[0,\tau]} \to 0$ for some $(A,B) \in D_{\phi^{(2)}}$. By the triangle inequality, its continuity is obtained if we prove separately that

$$\left\| \int A^- \mathrm{d}\,(B_n - B) \right\|_{[0,\tau]} \to 0, \tag{A.2}$$

$$\left\| \int (A_n^- - A^-) \mathrm{d}B_n \right\|_{[0,\tau]} \to 0, \tag{A.3}$$

as $n \to \infty$. Note that (A.3) is a direct consequence of $B_n \in BV_M$ for all $n \geq 1$. Now, by integration by parts, we have:

$$\int A^- \mathrm{d}\,(B_n - B) = A^-(B_n - B) - A^-(B_n - B)(0) - \int (B_n^- - B^-)\mathrm{d}(A^-). \tag{A.4}$$

Thus, the convergence (A.2) followed from the hypothesis that A is of bounded variation.

Let us now prove the LUHD property. First, write

$$\begin{aligned}
&t_n^{-1}\left(\phi^{(2)}\left(A_n+t_nh_n^A, B_n+t_nh_n^B\right)-\phi^{(2)}\left(A_n,B_n\right)\right)-\mathrm{d}\phi^{(2)}_{A,B}(h^A,h^B)\\
&=\int h_n^{A-}\mathrm{d}(B_n-B)+\int\left(A_n+t_nh_n^A-A\right)^-\mathrm{d}h_n^B+\int A^-\mathrm{d}(h_n^B-h^B)+\int(h_n^A-h^A)^-\mathrm{d}B\\
&=: R_{1,n}+R_{2,n}+R_{3,n}+R_{4,n}.
\end{aligned}$$

By the same arguments as above, both $\|\ R_{3,n}\ \|_{[0,\tau]}$ and $\|\ R_{4,n}\ \|_{[0,\tau]}$ straightforwardly tend to 0. By integration by parts, the term $\|\ R_{2,n}\ \|_{[0,\tau]}$ tends to 0 if we show that

$$\left\|\int h_n^{B-}\mathrm{d}\left((A_n+t_nh_n^A-A)^-\right)\right\|_{[0,\tau]}\to 0. \tag{A.5}$$

This is done (as well as proving that $\|\ R_{1,n}\ \|_{[0,\tau]}\to 0$) by following arguments of Van der Vaart and Wellner ([44], p. 382), and also borrowing their notations. For fixed $\eta>0$, let us approximate $h_B{}^-$ by $\tilde{h}^B:=\sum_{i=1}^{m}h^{B-}(t_i)\mathbb{1}_{(t_{i-1},t_i]}$ so that $\|\ h^{B-}-\tilde{h}^B\ \|_{[0,\tau]}\le\eta$. Then,

$$\begin{aligned}
&\left\|\int h_n^{B-}\mathrm{d}\left(A_n+t_nh_n^A-A\right)\right\|_{[0,\tau]}\le\|\ h_n^B-\tilde{h}^B\ \|_{[0,\tau]}\times\frac{2M}{\epsilon^2}\\
&+\sum_{i=1}^{m}\left|h^B(t_i)\right|\left|\left(A_n+t_nh_n^A-A\right)(t_i)-\left(A_n+t_nh_n^A-A\right)(t_{i-1})\right|+\left|h^B(\tau)\right|\left|\left(A_n+t_nh_n^A-A\right)\{\tau\}\right|,
\end{aligned}$$

where the factor $2M/\epsilon^2$ is due to the fact that both A and $A_n+t_nh_n^A$ belong to $BV_{M/\epsilon^2}([0,\tau])$. This proves (A.5) since both $\|\ h_n^B-h^B\ \|_{[0,\tau]}$ and $\|\ A_n+t_nh_n^A-A\ \|_{[0,\tau]}$ tend to 0 by assumption. The control of $R_{1,n}$ is done very similarly and is hence omitted. □

Our next lemma states the appropriate regularity properties of the product integral. Once again, we will borrow the notations of Van der Vaart and Wellner ([44], p. 390).

Lemma A.4. *The map $\phi^{(3)}$ defined on $D_{\phi^{(3)}}:=BV_M([0,\tau])\cap B_{M'}([0,\tau])$ by*

$$\phi^{(3)}:\ A\to\prod_{0<s\le\cdot}(1+dA)\in D([0,\tau])$$

is continuous on $D_{\phi^{(3)}}$. It is also LUHD at each $A\in D_{\phi^{(3)}}$, tangentially to $D([0,\tau])$, with derivative

$$\mathrm{d}\phi^{(3)}_A(h):\ A\to\phi^{(3)}(A)\times\int\frac{\phi^{(3)}(A)}{\left(\phi^{(3)}(A)\right)^-}dh\in D([0,\tau]).$$

Moreover, we have $\phi^{(3)}\left(D_{\phi^{(3)}}\right)\subset BV_{(M+M')\exp(M+M')}\cap B_{\exp(M+M')}$.

Proof. The continuity of the product integral with respect to $\|\cdot\|_{[\tau_1,\tau]}$ is well known, as well as the following equality, as functions on $[0,\tau]$, known as the Duhamel equation (see [44], pp. 390–391):

$$\phi^{(3)}(A)-\phi^{(3)}(B)=\phi^{(3)}(B)\times\int_{(0,\cdot]}\frac{\phi^{(3)}(A)^-}{\phi^{(3)}(B)}\mathrm{d}(B-A). \tag{A.6}$$

Hence we have

$$t_n^{-1}\left(\phi^{(3)}(A_n+t_nh_n)-\phi^{(3)}(A_n)\right)=\phi^{(3)}(A_n)\int_{(0,\cdot]}\frac{\phi^{(3)}(A_n+t_nh_n)^-}{\phi^{(3)}(A_n)}\mathrm{d}h_n, \tag{A.7}$$

Now, the arguments of Van der Vaart and Wellner [44], page 391 (where they approximate the h_n and their limit h by a common function $\tilde{h}$ of bounded variation) still apply in this context. Thus, it suffices to show that, for any function $\tilde{h}$ of bounded variation, we have:

$$\phi^{(3)}(A_n)\int_{(0,\cdot]}\frac{\phi^{(3)}(A_n+t_nh_n)^-}{\phi^{(3)}(A_n)}\mathrm{d}\tilde{h}\to\phi^{(3)}(A)\int_{(0,\cdot]}\frac{\phi^{(3)}(A)^-}{\phi^{(3)}(A)}\mathrm{d}\tilde{h},\ \text{as } n\to\infty.$$

This is true by continuity of $\phi^{(3)}$ with respect to $||\cdot||_{[0,\tau]}$ together with the fact that $\tilde{h}$ is of bounded variation. We omit details. □

A first consequence of the preceding lemmas is that the maps defining the Nelson–Aalen and the Kaplan–Meier estimators are both LUHD.

Proposition A.5. *For $M,\epsilon>0$, define*

$$D^{(4)}_{M,\epsilon}:=\left\{(F_0,F_1)\in BV_M([0,\tau])^2;\ \inf_{[0,\tau]}\mid 1-F_0-F_1\mid\geq\epsilon\right\}.$$

The Nelson-Aalen map $\phi^{(4)}$, defined on $D_{\phi^{(4)}}:=D^{(4)}_{M,\epsilon}$ by

$$\phi^{(4)}:=(F_0,F_1)\to\int\frac{1}{1-(F_0+F_1)^-}\mathrm{d}F_1\in D([0,\tau]),$$

is continuous on $D^{(4)}_{M,\epsilon}$. It is also LUHD at each $(F_0,F_1)\in D^{(4)}_{M,\epsilon}$, tangentially to $D([0,\tau])^2$, with derivative given by

$$(h^{F_0},h^{F_1})\to\int\frac{(h^{F_0}+h^{F_1})^-}{(1-(F_0+F_1)^-)^2}\mathrm{d}F_1+\int\frac{\mathrm{d}h^{F_1}}{1-(F_0+F_1)^-}\in D([0,\tau]).$$

Moreover, we have $\phi^{(4)}\left(D^{(4)}_{M,\epsilon}\right)\subset BV_{M/\epsilon}([0,\tau])\cap B_{M/\epsilon}([0,\tau])$.

As a consequence, the Kaplan–Meier function $KM:=\phi^{(3)}\circ\phi^{(4)}$ is LUHD at each $(F_0,F_1)\in D^{(4)}_{M,\epsilon}$, tangentially to $D([0,\tau])^2$, with derivative given by

$$(h^{F_0},h^{F_1})\to KM(F_0,F_1)\times\int\frac{KM(F_0,F_1)}{KM(F_0,F_1)^-}\left[\frac{(h^{F_0}+h^{F_1})^-}{(1-(F_0+F_1)^-)^2}\mathrm{d}F_1+\frac{\mathrm{d}h^{F_1}}{1-(F_0+F_1)^-}\right]\in D([0,\tau]).$$

In addition, we have $KM\left(D^{(4)}_{M,\epsilon}\right)\subset BV_{2M\epsilon^{-1}\exp(2M\epsilon^{-1})}\cap B_{\exp(2M\epsilon^{-1})}$.

Proof. Apply the chain rule for the composition $\phi^{(2)}\circ\phi^{(1)}\circ\phi$, where ϕ is the continuous linear map $(F_0,F_1)\to(1-(F_0+F_1)^-),F_1)$, for which $\phi(D^{(4)}_{M,\epsilon})\subset D^{(1)}(2M,\epsilon)$. We omit details. □

Now we just have to recall the previous results in order to prove the Proposition 5.5.

Proof of Proposition 5.5. One can decompose the map $\boldsymbol{\phi}(\cdot,\cdot,\cdot)$ like

$$\boldsymbol{\phi}(\tilde{\mu},F_0,F_1)=\phi^{(2)}\left(\phi^{(1)}\left(KM(F_0,F_1),1-F_1-F_0\right),\tilde{\mu}\right).$$

From the above lemmas and the chain rule (Lem. A.1) one can get the result of Proposition 5.5. □

References

[1] G. Adimari, Empirical likelihood type confidence intervals under random censorship. *Ann. Inst. Statist. Math.* **49** (1997) 447–466.

[2] P.K. Andersen, Ø. Borgan, R.D. Gill and N. Keiding, Statistical models based on counting processes. *Springer Ser. Stat.* Springer-Verlag, New York (1993).

[3] Y. Bilias, M. Gu and Zh. Ying, Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.* **25** (1997) 662–682.

[4] J. Cai and D.E. Schaubel, Analysis of recurrent event data. In *Advances in survival analysis*, Vol. 23 of *Handbook of Statist.* Elsevier, Amsterdam (2004) 603–623.

[5] J. Cai and D.E. Schaubel, Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Anal.* **10** (2004) 121–138.

[6] B.E. Chen and R.J. Cook, Tests for multivariate recurrent events in the presence of a terminal event. *Biostatistics* **5** (2004) 237–251.

[7] B.E. Chen, R.J. Cook, J.F. Lawless and M. Zhan, Statistical methods for multivariate interval-censored recurrent events. *Stat. Med.* **24** (2005) 671–691.

[8] J. Chen, L. Peng and Y. Zhao, Empirical likelihood based confidence intervals for copulas. *J. Multivariate Anal.* **100** (2009) 137–151

[9] S.X. Chen and I. Van Keilegom, A review on empirical likelihood methods for regression. *TEST* **18** (2009) 415–447.

[10] S.X. Chen, L. Peng and Y. Qin, Effects of data dimension on empirical likelihood. *Biometrika* **96** (2009) 711–722.

[11] R.J. Cook and J.F. Lawless, Marginal analysis of recurrent events and a terminating event. *Stat. Med.* **16** (1997) 911–924.

[12] R.J. Cook and J.F. Lawless, The Statistical Analysis of Recurrent Events. *Statistics for Biology and Health*. Springer-Verlag (2007).

[13] R.J. Cook, J.F. Lawless and C. Nadeau, Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* **52** (1996) 557–571.

[14] R.J. Cook, J.F. Lawless, L. Lakhal-Chaieb and K. Lee, Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: application to skeletal complications in cancer metastatic to bone. *J. Amer. Statist. Assoc.* **104** (2009) 60–75.

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

[16] L. Devroye and G. Lugosi, Combinatorial methods in density estimation. Springer, New York (2001).

[17] P. Du, Nonparametric modeling of the gap time in recurrent event data. *Lifetime Data Anal.* **15** (2009) 256–277.

[18] D. Ghosh, Accelerated rates regression models for recurrent failure time data. *Lifetime Data Anal.* **10** (2004) 247–261.

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

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

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

[22] D.J. Hand, The statistical analysis of recurrent events by Richard R. Cook, Jerald F. Lawless. *Int. Stat. Rev.* **76** (2008) 142–143.

[23] N.L. Hjort, I.W. McKeague and I. Van Keilegom, Extending the scope of empirical likelihood. *Ann. Statist.* **37** (2009) 1079–1111.

[24] B. Jing, J. Yuan and W. Zhou, Jackknife empirical likelihood. *J. Amer. Statist. Assoc.* **104** (2009) 1224–1232.

[25] A. Keziou and S. Leoni-Aubin, On empirical likelihood for semiparametric two-sample density ratio models. *J. Statist. Plann. Inference* **138** (2008) 915–928.

[26] J.F. Lawless, The analysis of recurrent events for multiple subjects. *J. Roy. Statist. Soc. Ser. C* **44** (1995) 487–498.

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

[28] G. Li, R. Li and M. Zhou, Empirical likelihood in survival analysis. In Contemporary multivariate analysis and design of experiments. Vol. 2 of *Ser. Biostat.* World Sci. Publ., Hackensack, NJ (2005) 337–349.

[29] D.Y. Lin and Z. Ying, Nonparametric tests for the gap time distributions of serial events based on censored data. *Biometrics* **57** (2001) 369–375.

[30] D.Y. Lin, L.J. Wei and Zhiliang Ying. Accelerated failure time models for counting processes. *Biometrika* **85** (1998) 605–618.

[31] D.Y. Lin, W. Sun and Z. Ying, Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika* **86** (1999) 59–70.

[32] 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** (2000) 711–730.

[33] M. Miloslavsky, S. Keleş, M.J. van der Laan and S. Butler. Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *J. R. Stat. Soc. Ser. B* **66** (2004) 239–257.

[34] A.B. Owen, Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** (1988) 237–249.

[35] A.B. Owen, Empirical likelihood for linear models. *Ann. Statist.* **19** (1991) 1725–1747.

[36] A.B. Owen, Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** (1990) 90–120.

[37] A.B. Owen, Empirical Likelihood. *Monographs on Statistics and Applied Probability.* Taylor & Francis (2001).
[38] A. Pakes and D. Pollard, Simulation and the asymptotics of optimization estimators. *Econometrica* **57** (1989) 1027–1057.
[39] D. Pollard, Empirical processes: theory and applications. Vol. 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, CA (1990).
[40] J. Ren, Weighted empirical likelihood ratio confidence intervals for the mean with censored data. *Ann. Inst. Statist. Math.* **53** (2001) 498–516.
[41] J. Ren, Smoothed weighted empirical likelihood ratio confidence intervals for quantiles. *Bernoulli* **14** (2008) 725–748.
[42] J. Ren, Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *Ann. Statist.* **36** (2008) 147–166.
[43] D.R. Thomas and G.L. Grunkemeier, Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.* **70** (1975) 865–871.
[44] A.W. Van der Vaart and J.A. Wellner, Weak convergence and empirical processes. Springer (1996).
[45] D. Varron, Empirical likelihood confidence bands for functional parameters in plug-in estimation. Preprint (2014).
[46] Q. Wang and B. Jing, Empirical likelihood for a class of functionals of survival distribution with censored data. *Ann. Inst. Statist. Math.* **53** (2001) 517–527.
[47] W. Wang and M.T. Wells, Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* **85** (1998) 561–572.