ANNALES DE L'I. H. P., SECTION B

D. BITOUZÉ
B. LAURENT
P. MASSART

A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator

Annales de l'I. H. P., section B, tome 35, nº 6 (1999), p. 735-763
<http://www.numdam.org/item?id=AIHPB_1999__35_6_735_0>

Ann. Inst. Henri Poincaré,

Vol. 35, n° 6, 1999, p. 735-763. *Probabilités et Statistiques*

A Dvoretzky–Kiefer–Wolfowitz type inequality for the Kaplan–Meier estimator

by

D. BITOUZÉ[a,1], **B. LAURENT**[b,2], **P. MASSART**[b,3]

[a] L.M.P.A. Joseph Liouville, Université du Littoral-Côte d'Opale, Centre Universitaire de la Mi-Voix, 50 rue F. Buisson, B.P. 699, F-62228 Calais Cédex, France

[b] Université Paris Sud. Bât. 425, F-91405 Orsay Cédex, France

Article received on 10 February 1998, revised 12 May 1999

ABSTRACT. – We prove a new exponential inequality for the Kaplan–Meier estimator of a distribution function in a right censored data model. This inequality is of the same type as the Dvoretzky–Kiefer–Wolfowitz inequality for the empirical distribution function in the non-censored case. Our approach is based on Duhamel equation which allows to use empirical process theory. © Elsevier, Paris

Key words: Kaplan–Meier estimator, censored data, exponential inequality, law of iterated logarithm, empirical process

RÉSUMÉ. – Nous prouvons une nouvelle inégalité exponentielle pour l'estimateur de Kaplan–Meier d'une fonction de répartition dans un modèle de données censurées à droite. Cette inégalité est du même type que l'inégalité de Dvoretzky–Kiefer–Wolfowitz pour la fonction de répartition empirique dans le cas non censuré. Notre approche est basée sur l'équation de Duhamel qui nous permet d'utiliser la théorie des processus empiriques. © Elsevier, Paris

[1] E-mail: bitouze@univ-littoral.fr.

[2] E-mail: Beatrice.Laurent@math.u-psud.fr.

[3] E-mail: Pascal.Massart@math.u-psud.fr.

1. INTRODUCTION

Let $Z_1, Z_2, \ldots, Z_n, \ldots$ be a sequence of independent random variables with common distribution function F on the real line. The properties of the empirical distribution function $\widehat{F_n}$ based on the sample $Z_1, \ldots, Z_n$ as an estimator of F have been investigated for a long time. In particular, Donsker's theorem, see Donsker [7], which ensures the weak convergence of $\sqrt{n}(\widehat{F_n} - F)$ towards $B^0 \circ F$ where B^0 is a Brownian bridge, is an essential tool for studying the asymptotic behaviour of several statistics (the Kolmogorov–Smirnov or Von Mises statistics, for example). As a matter of fact, this limit theorem may be completed by a sharp nonasymptotic bound which is due to Dvoretzky, Kiefer, and Wolfowitz (DKW) [10]:

$$\mathbb{P}\left(\sqrt{n}\left\|\widehat{F_n} - F\right\|_\infty > \lambda\right) \leqslant C\,\mathrm{e}^{-2\lambda^2}.$$

Moreover, Massart [19] proved that C may be taken as 2 which makes the use of this bound relevant for small samples.

The estimator $\widehat{F_n}$ may be viewed as a NonParametric Maximum Likelihood Estimator (NPMLE) of F. In the random censorship model, the NPMLE of F can be also explicitly computed, it is the celebrated Kaplan–Meier estimator introduced by Kaplan and Meier [18]. Our aim in this paper is to prove an exponential bound for the Kaplan–Meier estimator which is of the same type as the DKW inequality.

Let us first describe the right censored data model. Let $X_1, \ldots, X_n, \ldots$ and $Y_1, \ldots, Y_n, \ldots$ be independent sequences of independent and identically distributed nonnegative random variables with distribution functions F and G respectively. One observes $Z_1, \ldots, Z_n$ such that for $1 \leqslant i \leqslant n$, $Z_i = (W_i, \delta_i)$ where

$$W_i = X_i \wedge Y_i, \qquad \delta_i = \mathbf{1}_{\{X_i \leqslant Y_i\}}.$$

In survival analysis, the X_i's represent lifetimes, and the Y_i's censoring times. The Kaplan–Meier estimator $\widehat{F_n}$ of the distribution function F is defined by

$$1 - \widehat{F_n}(x) = \prod_{\{i \in \{1,\ldots,n\},\, W_i \leqslant x\}} \left(\frac{n - r_i}{n - r_i + 1}\right)^{\delta_i},$$

where r_i is the rank of $(W_i, 1 - \delta_i)$ in the set $\{(W_j, 1 - \delta_j),\ j \in \{1, \ldots, n\}\}$. The ordering is lexicographical, this means that in the case of ties among the set $\{W_i, 1 \leqslant i \leqslant n\}$, death observations ($\delta_i = 1$) come before censored observations ($\delta_i = 0$). Note that in the complete data model $\widehat{F_n}$ reduces to the empirical distribution function.

Many efforts have been made to extend known results on the empirical distribution function to the Kaplan–Meier estimator and there is a huge literature on the topic. Donsker's theorem (see Donsker [7]) has been extended in various directions. The weak convergence of the process $\sqrt{n}(\widehat{F_n} - F)$ to a Gaussian process was established by Breslow and Crowley [5] on a fixed interval $]-\infty, w]$ with $w < \tau$, where τ is defined by

$$\tau = \inf\{x \in \mathbb{R}, (1-F)(1-G)(x) = 0\}$$

(with the convention that $\inf\emptyset = +\infty$). Gill [12] proved the weak convergence of a conveniently weighted version of the Kaplan–Meier process on the whole interval $]-\infty, \tau]$.

Concerning exponential bounds, Dinwoodie [6] has established a large deviation principle for censored data which allows to study the asymptotic behaviour of

$$\frac{1}{n}\ln \mathbb{P}\big(\|\widehat{F_n} - F\|_\infty > \varepsilon\big),$$

when n is large and ε is fixed. Unfortunately, this limit theorem does not provide any information about the moderate deviations of $\|\widehat{F_n} - F\|_\infty$ and therefore, a DKW type inequality cannot be deduced from such a result.

Up to our knowledge, the first nonasymptotic exponential bound for the Kaplan–Meier estimator is due to Földes and Rejtő [11]. They proved that if $w < \tau$:

$$\mathbb{P}\Big(\sqrt{n}\sup_{t\leqslant w}\big|(\widehat{F_n} - F)(t)\big| \geqslant \lambda\Big) \leqslant C\,\mathrm{e}^{-\eta((1-F)(1-G)(w))^4\lambda^2},$$

where C and η are absolute constants.

The main result of this paper improves on the preceding bound of Földes and Rejtő and may be stated as follows:

THEOREM 1. – *Let $\widehat{F_n}$ be the Kaplan–Meier estimator of the distribution function F. There exists an absolute constant C such that, for any positive λ,*

$$\mathbb{P}\big(\sqrt{n}\|(1-G)(\widehat{F_n} - F)\|_\infty > \lambda\big) \leqslant 2.5\,\mathrm{e}^{-2\lambda^2 + C\lambda}. \tag{1}$$

It is worth noticing that, for censored data, the Kaplan–Meier estimator may fail to be uniformly consistent on the whole line, this happens whenever $G^{-}(\tau) = 1$ and $F^{-}(\tau) < 1$ (see Gill [13]). The weighting factor $1 - G$ that we use allows to control the uniform deviation on the whole line almost as well as in the complete data model. One can wonder whether our bound is sharp or not. The asymptotic behaviour of the left hand side of inequality (1) is well known in the non censored case. Provided that F is continuous, it converges as n goes to infinity towards

$$Q(\lambda) = \mathbb{P}\Big(\sup_{t\in[0,1]} |B^0(t)| \geqslant \lambda\Big),$$

where B^0 is a Brownian bridge. Since $Q(\lambda) \sim 2\,\mathrm{e}^{-2\lambda^2}$, as λ goes to infinity, it means that Massart's upper bound $2\,\mathrm{e}^{-2\lambda^2}$ is optimal and that, in the censored case, we miss this bound by a factor $1.25\,\mathrm{e}^{C\lambda}$. Therefore, our bound has the right exponential decay with respect to λ^2. Of course, it would be desirable to compute C, but unfortunately, our techniques are not sharp enough to do that efficiently.

We shall in fact prove the following slightly better inequality than (1):

$$\mathbb{P}\Big(\sup_{1\leqslant k\leqslant n} \frac{k}{\sqrt{n}} \big\|(1-G)\big(\widehat{F_k} - F\big)\big\|_\infty > \lambda\Big) \leqslant 2.5\,\mathrm{e}^{-2\lambda^2 + C\lambda}$$

from which one can straightforwardly derive a bounded Law of Iterated Logarithm (LIL):

$$\limsup_{n\to+\infty} \frac{\sqrt{n}\|(1-G)(\widehat{F_n} - F)\|_\infty}{\sqrt{2\ln\ln n}} \leqslant \frac{1}{2} \quad \text{a.s.} \tag{2}$$

Again this almost sure upper bound is known to be optimal in the non censored case. This result seems to be new. Gu and Lai [16] actually proved a functional LIL which is apparently stronger, but since this result does not hold on the whole line it does not imply (2).

To prove Theorem 1, we follow the idea that Van der Laan has developed in his Thesis [25]. He noticed that, for some indirectly observed models, the NPMLE $\widehat{F_n}$ of a distribution function F satisfies an identity of the type:

$$\sqrt{n}\big(\widehat{F_n}(t) - F(t)\big) = \int I_{\widehat{F_n},t}(z)\,\mathrm{d}\nu_n(z),$$

where ν_n is a centered and normalized empirical measure and $I_{K,t}$ denotes some influence function for the (sub-probability) distribution function K. He showed that $\sqrt{n}$ asymptotics for $\widehat{F_n}$ can be derived from this implicit equation provided that the class of influence functions $I_{K,t}$ has the Donsker property when K and t vary.

Such an identity indeed holds for the Kaplan–Meier estimator and follows from Duhamel equation. Entropy with bracketing computations allows as Duhamel equation to derive Theorem 1 from a nonasymptotic exponential bound for empirical processes which is of independent interest.

The paper is organized as follows: in Section 2, we state nonasymptotic and asymptotic bounds for the Kaplan–Meier estimator, Section 3 is devoted to a general exponential inequality for the empirical process. The main proofs are postponed to Section 4.

2. EXPONENTIAL INEQUALITY AND ASYMPTOTIC BOUNDS FOR THE KAPLAN–MEIER ESTIMATOR

2.1. An exponential inequality

In this section, we consider the right censored data model as defined in Section 1. The Kaplan–Meier estimator of the distribution of interest F is denoted by $\widehat{F_n}$ and G denotes the distribution function of the censoring times.

We normalize the Kaplan–Meier process $\widehat{F_n} - F$ by $1 - G^-$, where G^- is the left-continuous version of G. This allows us to formulate an exponential inequality for the deviation of the supremum of the process $(1 - G^-)(\widehat{F_n} - F)$ on the whole line. We shall in fact prove a slightly stronger result than Theorem 1 which has the advantage to readily imply a bounded LIL.

THEOREM 2. – *The following inequality holds for the Kaplan–Meier estimator $\widehat{F_n}$. For any $\lambda > 0$,*

$$\mathbb{P}\left(\sup_{1\leqslant k\leqslant n} \frac{k}{\sqrt{n}}\left\|(1 - G^-)\left(\widehat{F_k} - F\right)\right\|_\infty \geqslant \lambda\right) \leqslant 2.5\, \mathrm{e}^{-2\lambda^2 + C\lambda}, \tag{3}$$

where C is an absolute constant.

Remark 1. – Since $1 - G \leqslant 1 - G^-$, the same result holds if we replace G^- by G.

One gets as a corollary of Theorem 2:

COROLLARY 1. – *The sequence $(\widehat{F_n})_{n\geqslant 1}$ of Kaplan–Meier estimators satisfies the bounded law of the iterated logarithm*:

$$\limsup_{n\to+\infty} \frac{\sqrt{n}\|(1-G^-)(\widehat{F_n}-F)\|_\infty}{\sqrt{2\ln\ln n}} \leqslant \frac{1}{2} \quad a.s. \tag{4}$$

Remark 2. – In the complete data model, inequality (4) is sharp since it is known that

$$\limsup_{n\to+\infty} \frac{\sqrt{n}\|\widehat{F_n}-F\|_\infty}{\sqrt{2\ln\ln n}} = \frac{1}{2} \quad \text{a.s.}$$

This result is due to Smirnov (see Shorack and Wellner [21]).

Remark 3. – Strong consistency results for the Kaplan–Meier estimator can be immediately derived from the bounded LIL. As a matter of fact, provided that $G^-(\tau) < 1$, we easily get that

$$\sup_{t\leqslant\tau}\left|\left(\widehat{F_n}-F\right)(t)\right| \xrightarrow[n\to+\infty]{\text{a.s.}} 0.$$

If $G^-(\tau) = 1$ and $F^-(\tau) = 1$, the strong consistency can be derived easily from the strong consistency in the case $G^-(\tau) < 1$ via some monotonicity argument as pointed out by Gill [13]. We recover here that the uniform consistency of the Kaplan–Meier estimator holds on $[0, \tau]$ provided that condition (5) is fulfilled:

$$G^-(\tau) < 1 \quad \text{or} \quad F^-(\tau) = 1. \tag{5}$$

This result is due to Stute and Wang [23] who proved it by using martingale technics. Note that Stute and Wang [23] have furthermore investigated the case where condition (5) does not hold. They proved that in this case $\widehat{F_n}$ is not consistent since it converges to a limit which is different from F. In this situation, one needs to modify the Kaplan–Meier estimator to get a consistent estimator (see Gill [13]).

2.2. An asymptotic upper bound

We now state an asymptotic upper bound which allows to understand how sharp is our DKW type inequality for the Kaplan–Meier estimator.

THEOREM 3. – *Let $\widehat{F_n}$ be the Kaplan–Meier estimator of F, and let B^0 be a Brownian bridge. For any positive λ,*

$$\limsup_{n\to+\infty} \mathbb{P}\left(\sqrt{n}\left\|(1-G)\left(\widehat{F_n}-F\right)\right\|_\infty > \lambda\right) \leqslant \mathbb{P}\Big(\sup_{t\in[0,1]} \left|B^0(t)\right| > \lambda\Big) \leqslant 2\,\mathrm{e}^{-2\lambda^2}. \tag{6}$$

Remark 4. – Inequality (6) ensures that, asymptotically, the uniform deviation of the conveniently weighted Kaplan–Meier process is stochastically smaller than the standard Kolmogorov–Smirnov limiting distribution. We do not know whether the inequality

$$\mathbb{P}\left(\sqrt{n}\left\|(1-G)\left(\widehat{F_n}-F\right)\right\|_\infty > \lambda\right) \leqslant 2\,\mathrm{e}^{-2\lambda^2} \tag{7}$$

holds or not for the Kaplan–Meier estimator. Indeed, this is a natural question since on the one hand we know from (6) that it holds asymptotically and on the other hand that it is valid for all n in the non censored case (see Massart [19]). Inequality (1) does not of course imply (7) but can be considered as a first step towards such a result.

Proof. – The proof of Theorem 3 follows from well known arguments. Let $\mathcal{T} = \{t,\ G^-(t) < 1\}$, let C be the continuous, nondecreasing and nonnegative function

$$C(t) = \int_0^t \frac{\mathrm{d}F(s)}{(1-F^-(s))^2(1-G^-)(s)}$$

and let $K(t) = C(t)/(1+C(t))$ if $C(t) < +\infty$ and $K(t) = 1$ if $C(t) = +\infty$. Combining the weak convergence result towards a Gaussian process given by Gill [14], p. 173, with the representation of that process from a Brownian bridge given by Gill [12], we have that the process $\sqrt{n}(1-G)(\widehat{F_n}-F)$ converges towards $((1-F)(1-G)/(1-K))B^0(K)$ in $\mathcal{D}(\mathcal{T})$. Therefore,

$$\lim_{n\to+\infty} \sqrt{n}\left\|(1-G)\left(\widehat{F_n}-F\right)\right\|_\infty = \sup_{t\in\mathcal{T}}\left[\frac{(1-F)(t)(1-G)(t)}{(1-K)(t)}\left|B^0\big(K(t)\big)\right|\right]$$

and inequality (6) follows from the fact that $(1-G)(1-F)/(1-K) \leqslant 1$. □

2.3. Connection with empirical processes: van der Laan's identity

A connection between the Kaplan–Meier process $\widehat{F_n} - F$ and the empirical process $\mathrm{P}_n - \mathrm{P}$ based on the variables $Z_i = (W_i, \delta_i)$ is given by van der Laan's identity (van der Laan [26]). In our particular case, this identity is equivalent to the well-known Duhamel equation for the univariate product integral (van der Laan [26], p. 14, and Gill and Johansen [15], Theorem 6). The version presented below is to be found in Gill [14], p. 172. Let $I_{K,x}$ be the influence function defined, for any [0, 1] valued and nondecreasing function K, and for any x such that $G^-(x) < 1$, by

$$\begin{aligned} I_{K,x}(w,\delta) = {} & \delta \frac{(1-K)(x)(1-G^-)(x)}{(1-K)(w)(1-G^-)(w)} \mathbf{1}_{]-\infty,x]}(w) \\ & - \int_{-\infty}^{w} \frac{(1-K)(x)(1-G^-)(x)}{(1-G^-)(u)} \mathbf{1}_{]-\infty,x]}(u)\, \mathrm{d}\left(\frac{1}{1-K}\right)(u), \end{aligned}$$

where $\mathbf{1}_A$ denotes the indicator function of any set A. Moreover, $I_{K,x} \equiv 0$ if $G^-(x) = 1$.

THEOREM 4 (van der Laan). – *For any* $x \in \mathbb{R}$

$$(1-G^-)(x)\big(F - \widehat{F_n}\big)(x) = (\mathrm{P}_n - \mathrm{P})(I_{\widehat{F_n},x}). \tag{8}$$

Remark 5. – In the definition of the influence function, it may happen that some division by 0 occurs. We just refer to Gill [14], p. 128, where it is shown that the convention 0/0 equals 1 is adequate.

Denoting by $\mathcal{K}$ the class of [0, 1] valued and non decreasing functions on $\mathbb{R}$, and by $\mathcal{I}$ the set of functions $\mathcal{I} = \{I_{K,x},\ K \in \mathcal{K},\ x \in \mathbb{R}\}$, we deduce from identity (8) that

$$\left\|(1-G^-)\big(\widehat{F_n} - F\big)\right\|_\infty \leqslant \sup_{f \in \mathcal{I}} \left|(\mathrm{P}_n - \mathrm{P})(f)\right|.$$

So that we are in a position to derive Theorem 2 from general exponential bounds for empirical processes involving metric properties of $\mathcal{I}$. The purpose of the next section is precisely to establish such bounds.

3. EXPONENTIAL BOUNDS FOR EMPIRICAL PROCESSES

Let $Z_1, \ldots, Z_n$ be n independent identically distributed random variables with distribution P on some measurable space $(\mathcal{Z}, \mathcal{C})$. Let P_n denote the empirical probability measure based upon $(Z_1, \ldots, Z_n)$. For $p \in [1, \infty[$, we denote by $\mathbb{L}^p(\mathrm{P})$ the set of measurable real valued functions f on $\mathcal{Z}$ such that $|f|^p$ is integrable with respect to P. We set $\nu_n = \sqrt{n}(\mathrm{P}_n - \mathrm{P})$. This means that

$$\nu_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(f(Z_i) - \int f \, \mathrm{dP} \right),$$

whenever $f \in \mathbb{L}^1(\mathrm{P})$. Moreover, for $f \in \mathbb{L}^2(\mathrm{P})$, we denote by $\mathrm{Var}_{\mathrm{P}}(f)$ the quantity $\int f^2 \, \mathrm{dP} - (\int f \, \mathrm{dP})^2$. We shall deal with $\mathbb{L}^2$-entropy with bracketing as defined below.

DEFINITION 1. – *Let μ be some nonnegative measure. Given $\varepsilon > 0$, a bracket with $\mathbb{L}^2(\mu)$-diameter ε is defined from a pair of functions f, g in $\mathbb{L}^2(\mu)$ with $f \leqslant g$ and $\|f - g\|_2 = \varepsilon$ as*

$$[f, g] = \{h \in \mathbb{L}^2(\mu),\ f \leqslant h \leqslant g\}.$$

Let $\mathcal{F} \subset \mathbb{L}^2(\mu)$. Denoting by $N(\varepsilon)$ the minimum cardinality of a covering of $\mathcal{F}$ by a finite set of brackets with diameter not larger than ε, the $\mathbb{L}^2(\mu)$-entropy with bracketing of $\mathcal{F}$ is defined as the function $\varepsilon \mapsto \ln(N(\varepsilon) \vee e)$.

The notion of entropy with bracketing has been introduced by Dudley [9] and the importance of $\mathbb{L}^2$-entropy with bracketing has been pointed out by Ossiander [20]. It refines on $\mathbb{L}^\infty$-entropy and is especially well suited for studying classes of functions with uniformly bounded variations for which, of course, $\mathbb{L}^\infty$-entropy is irrelevant. The control of the entropy with bracketing of the class of influence functions $\mathcal{F}$ introduced in Section 2, will be precisely derived from the entropy computation of the class of functions with uniformly bounded variations, due to van de Geer [24].

LEMMA 1 (van de Geer). – *Let μ be some nonnegative bounded measure on some interval $U \subset \mathbb{R}$. Given some positive number M, the $\mathbb{L}^2(\mu)$-entropy with bracketing H of the class of functions defined on U with variation bounded by M satisfies the inequality,*

$$H(\varepsilon) \leqslant \gamma M \sqrt{\mu(U)}/\varepsilon$$

for any $\varepsilon > 0$, where γ is some absolute constant.

We now provide some general exponential bounds for empirical processes which are refinements of related bounds in Birgé and Massart [2]. As compared to their Proposition 3, the main novelty here is that we provide a Hoeffding type inequality with an explicit optimal sub-Gaussian rate of decay. The proof uses the adaptive truncation procedure introduced by Bass [1] for partial sum processes and Ossiander [20] for empirical processes. Throughout the sequel, $\mathbb{P}^*$ will denote the outer probability measure associated with $\mathbb{P}$ (we recall that, generally speaking, $\sup_{f\in\mathcal{F}}|\nu_n(f)|$ is not necessarily measurable).

THEOREM 5. – *Let $\mathcal{F}$ be a class of measurable functions defined on $\mathcal{Z}$. We assume that for some constants m and M*

$$\forall f \in \mathcal{F} \quad \forall z \in \mathcal{Z}, \quad m \leqslant f(z) \leqslant M. \tag{9}$$

Let $a = M - m$. Then, $\mathrm{Var}_{\mathrm{P}}(f) \leqslant a^2/4$ for any f in $\mathcal{F}$. Let H denote the $\mathbb{L}^2(\mathrm{P})$-entropy with bracketing of $\mathcal{F}$. Assume that $H^{1/2}$ is integrable at 0, and let

$$\varphi(t) = \int_0^t H^{1/2}(x)\,\mathrm{d}x$$

for any positive t. Let $\varepsilon_0 \in]0,1[$, and σ be some positive number such that $\sup_{f\in\mathcal{F}}\mathrm{Var}_{\mathrm{P}}(f) \leqslant \sigma^2 \leqslant a^2/4$. For any positive t and λ, we set $\psi(t) = t + 1 - \sqrt{1+2t}$ and

$$h_1(\lambda) = \frac{2\lambda^2}{a^2}, \qquad h_2(\lambda) = \frac{9n\sigma^2}{a^2}\psi\left(\frac{a\lambda}{3\sqrt{n}\sigma^2}\right).$$

Then, for any positive λ,

$$\begin{aligned}\mathbb{P}^*\Big(\sup_{1\leqslant k\leqslant n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}|\nu_k(f)| > \lambda + C\Big(\varepsilon_0\frac{\lambda}{\sigma} + \varphi(\varepsilon_0)\Big)\Big)&\\ \leqslant 2.5\,\mathrm{e}^{H(\varepsilon_0)}\,\mathrm{e}^{-h(\lambda)},&\end{aligned} \tag{10}$$

where C is an absolute constant, and h can be taken as h_1 or h_2.

We first deduce from Theorem 5 a sharp Hoeffding type inequality taking benefit of a special shape for the entropy function. Because of the entropy computation by van de Geer [24], this inequality will imply Theorem 2.

COROLLARY 2. – *Let $\mathcal{F}$ be a class of real functions defined on $\mathcal{Z}$ and satisfying condition* (9). *Suppose the $\mathbb{L}^2(\mathrm{P})$-entropy with bracketing* *of $\mathcal{F}$, denoted with H, satisfies $H(\varepsilon) \leqslant \gamma/\varepsilon$ for some γ in $\mathbb{R}^+$. Then for any positive real number λ:*

$$\mathbb{P}^*\Big(\sup_{1\leqslant k\leqslant n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}|\nu_k(f)| > \lambda\Big) \leqslant 2.5\ \mathrm{e}^{-2\lambda^2/a^2+C\lambda},$$

where C is an absolute constant and $a = M - m$.

Since the function ψ defined in Theorem 5 satisfies $\psi(t) \geqslant \frac{t^2}{2t+2}$, one can also straightforwardly get from Theorem 5 a Bernstein type inequality. Although we shall not present here applications of this bound, it should be noticed that it can be used to derive local properties of the Kaplan–Meier estimator in the spirit of Bitouzé's Thesis [4].

COROLLARY 3. – *Let $\mathcal{F}$ be a class of real functions defined on $\mathcal{Z}$ such that condition* (9) *holds. Let H be the $\mathbb{L}^2(\mathrm{P})$-entropy with bracketing of $\mathcal{F}$. We assume that $H^{1/2}$ is integrable at* 0 *and we define*

$$\varphi(t) = \int_0^t H(x)^{1/2}\,\mathrm{d}x.$$

Let $\theta \in]0, 1]$. The following inequality holds for any positive σ such that $\sup_{f\in\mathcal{F}} \mathrm{Var_P}(f) \leqslant \sigma^2 \leqslant a^2/4$ and for any positive λ:

$$\mathbb{P}^*\bigg(\sup_{1\leqslant k\leqslant n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}|\nu_k(f)| > \lambda(1+C\theta) + C\frac{\varphi(\theta\sigma)}{\theta}\Big(1+\frac{\varphi(\theta\sigma)}{\sqrt{n}\theta\sigma^2}\Big)\bigg)$$
$$\leqslant 2.5\ \mathrm{e}^{-h(\lambda)},$$

where C is an absolute constant and

$$h(\lambda) = \frac{\lambda^2}{(2\sigma^2 + \frac{2a\lambda}{3\sqrt{n}})}.$$

4. PROOFS

4.1. Proof of Theorem 2

Identity (8) ensures that

$$\mathbb{P}\Big(\sup_{1\leqslant k\leqslant n}\frac{k}{\sqrt{n}}\big\|(1-G^-)(\widehat{F_k}-F)\big\|_\infty \geqslant \lambda\Big)$$

$$\leqslant \mathbb{P}^*\biggl(\sup_{1\leqslant k\leqslant n}\sup_{f\in\mathcal{I}}\sqrt{\frac{k}{n}}\bigl|\nu_k(f)\bigr|\geqslant\lambda\biggr),$$

where

$$\mathcal{I}=\bigl\{I_{K,x},\ K\in\mathcal{K},\ x\in\mathbb{R}\bigr\}$$

and $\mathcal{K}$ is the class of [0, 1] valued and non decreasing functions on $\mathbb{R}$. Therefore, inequality (3) may be derived from Corollary 2 applied with the class $\mathcal{I}$. In order to apply Corollary 2, we have to verify that, for some constants m and M, any f in $\mathcal{I}$ satisfies $m\leqslant f\leqslant M$ and we have to evaluate the $\mathbb{L}^2(\mathrm{P})$-entropy with bracketing of $\mathcal{I}$.

Since ν_n is centered, we may add a constant to the function f without affecting $\nu_n(f)$. More precisely, the following equality is equivalent to (8): for any x in $\mathbb{R}$,

$$(1-G^-)(x)\bigl(F-\widehat{F_n}\bigr)(x)=(\mathrm{P}_n-\mathrm{P})(J_{\widehat{F_n},x}),$$

where

$$J_{K,x}=I_{K,x}+(1-K)(x)(1-G^-)(x)\int_{-\infty}^{x}1/(1-G^-)(u)\,\mathrm{d}(1/1-K)(u)$$

may be written

$$J_{K,x}(w,\delta)=(1-K)(x)(1-G^-)(x)\mathbf{1}_{]-\infty,x]}(w)\biggl[\frac{\delta}{(1-K)(w)}\times(1-G^-)(w)+\int_{w}^{x}\frac{1}{(1-G^-)(u)}\mathrm{d}\biggl(\frac{1}{(1-K)}\biggr)(u)\biggr].$$

We now consider the class of functions $\mathcal{J}=\{J_{K,x},\ K\in\mathcal{K},x\in\mathbb{R}\}$. This modification will make easier the evaluations of the constants m and M.

Indeed, noticing that on $]-\infty,x]$, both $(1-K)(x)/(1-K)$ and $(1-G^-)(x)/(1-G^-)$ belong to [0, 1] (since K and G^- are nondecreasing maps), one can verify that $0\leqslant J_{K,x}\leqslant 1$ for any $J_{K,x}$ in $\mathcal{J}$.

We now use the following lemma whose proof is postponed to Appendix A. This result derives from Lemma 1 due to van de Geer [24].

LEMMA 2. – *For any probability measure* P, *the* $\mathbb{L}^2(\mathrm{P})$-*entropy with bracketing of the class* $\mathcal{J}$ *satisfies* $H(\varepsilon)\leqslant\gamma/\varepsilon$, *where* γ *is an absolute constant.*

Now the proof of Theorem 2 follows immediately from Corollary 2. □

4.2. Proof of the LIL (Corollary 1)

Let ε be in $]0,1[$ and M_k be $k\|(1-G^-)(\widehat{F_k}-F)\|_\infty/\sqrt{n}$. We deduce from Theorem 2 that, for $\lambda \geqslant C/(2\varepsilon)$,

$$\mathbb{P}\Big(\sup_{1\leqslant k\leqslant n} M_k \geqslant \lambda\Big) \leqslant 2.5\,\mathrm{e}^{-2(1-\varepsilon)\lambda^2}.$$

Let $\Phi(n)=\sqrt{\ln\ln n/2}$ and let $\theta>1$. From the above inequality applied to $\varepsilon<1-1/\theta^2$ and $\lambda=\theta\Phi(n_k)$, we derive that

$$\sum_{k\geqslant 1}\mathbb{P}\Big(\sup_{1\leqslant j\leqslant n_k} M_j \geqslant \theta\Phi(n_k)\Big) < +\infty,$$

where n_k is the integer part of θ^k. From Borel-Cantelli lemma, we deduce that a.s., $\sup_{1\leqslant j\leqslant n_k} M_j < \theta\Phi(n_k)$ for k large enough ($k\geqslant k^*$). Let now n be large enough such that $n_{k-1}<n\leqslant n_k$ for some $k\geqslant k^*$. We have

$$\frac{M_n}{\Phi(n)} \leqslant \sup_{1\leqslant j\leqslant n_k}\frac{M_j}{\Phi(n)} \leqslant \sup_{1\leqslant j\leqslant n_k}\frac{M_j}{\Phi(n_k)}\frac{\Phi(n_k)}{\Phi(n_{k-1})}$$

so since, as easily seen, $\Phi(n_k)/\Phi(n_{k-1})\to 1$ as $n\to+\infty$, we deduce that for any $\theta'>1$, a.s. $M_n/\Phi(n)<\theta'$ for n large enough. This leads to

$$\limsup_{n\to+\infty}\frac{M_n}{\Phi(n)} \leqslant 1 \quad \text{a.s.}$$

This concludes the proof of inequality (4). □

4.3. Proof of Theorem 5

We set

$$\widetilde{\nu_n}(f) = \sup_{1\leqslant k\leqslant n}\sqrt{\frac{k}{n}}\,\nu_k(f).$$

We shall prove exponential inequality (10) by controlling $\sup_{f\in\mathcal{F}}\widetilde{\nu_n}(f)$. The required two sided inequality will follow by considering the class $-\mathcal{F}$ instead of $\mathcal{F}$ and multiplying the one sided probability bound by 2.

4.3.1. Notations

We first introduce some notations. Let k be fixed in $\mathbb{N}$ and ε_k be $2^{-k}\varepsilon_0$ for any k in $\mathbb{N}$.

We consider a covering of $\mathcal{F}$ by a set $\mathcal{N}_{\varepsilon_k}$ of brackets with diameter not larger than ε_k. We assume that $\mathcal{N}_{\varepsilon_k}$ has cardinality not larger than $\mathrm{e}^{H(\varepsilon_k)}$. For any f in $\mathcal{F}$, we denote by $[f_k^L, f_k^U]$ a bracket of $\mathcal{N}_{\varepsilon_k}$ such that $f_k^L \leqslant f \leqslant f_k^U$, and we denote by $\Delta_k(f)$ the function $f_k^U - f_k^L$. For any function $g \in \mathbb{L}^2(\mathrm{P})$, we shall denote by $\|g\|_2$ the quantity $\mathrm{P}^{1/2}(g^2) = (\int g^2\,\mathrm{dP})^{1/2}$.

Let $\Pi_k(f)$ be $f_k^U \wedge M$, where $\wedge$ denotes the minimum, so that the function $f - \Pi_k(f)$ is nonpositive (recall that $m \leqslant f \leqslant M$ for any f in $\mathcal{F}$).

We consider some decreasing sequence of positive real numbers $(u_k)_{k\in\mathbb{N}}$ and, for any f in $\mathcal{F}$, we set $\tau(f) = \min\{k \in \mathbb{N},\ \Delta_k(f) > u_k\} \wedge p$ where p has to be chosen in $\mathbb{N}$.

We denote by Id the identity operator.

4.3.2. Decomposition

We shall use the following decomposition given in Doukhan et al. [8], p. 412.

$$\begin{aligned}\mathrm{Id} = {} & \Pi_0 + (\mathrm{Id} - \Pi_0)\mathbf{1}_{\{\tau=0\}} + (\mathrm{Id} - \Pi_{p-1})\mathbf{1}_{\{\tau\geqslant p\}} \\ & + \sum_{k=1}^{p-1}(\mathrm{Id} - \Pi_{k-1})\mathbf{1}_{\{\tau=k,\Delta_k>u_{k-1}\}} + \sum_{k=1}^{p-1}(\mathrm{Id} - \Pi_k)\mathbf{1}_{\{\tau=k,\Delta_k\leqslant u_{k-1}\}} \\ & + \sum_{k=1}^{p-1}(\Pi_k - \Pi_{k-1})\mathbf{1}_{\{\tau\geqslant k,\Delta_k\leqslant u_{k-1}\}}.\end{aligned}$$

Let Λ be a real number. The following inequality derives from the previous identity.

$$\mathbb{P}^*\Big(\sup_{f\in\mathcal{F}} \widetilde{\nu_n}(f) > \Lambda\Big) \leqslant \sum_{j=1}^{6} \mathbb{P}_j, \tag{11}$$

where, assuming that $\sum_{j=1}^{6} \lambda_j \leqslant \Lambda$,

$$\mathbb{P}_1 = \mathbb{P}\Big(\sup_{f\in\mathcal{F}} \widetilde{\nu_n}\big(\Pi_0(f)\big) > \lambda_1\Big),$$

$$\mathbb{P}_2 = \mathbb{P}^*\Big(\sup_{f\in\mathcal{F}} \widetilde{\nu_n}\big((f - \Pi_0(f))\mathbf{1}_{\{\tau(f)=0\}}\big) > \lambda_2\Big),$$

$$\mathbb{P}_3 = \mathbb{P}^*\Big(\sup_{f\in\mathcal{F}} \widetilde{\nu_n}\big((f - \Pi_{p-1}(f))\mathbf{1}_{\{\tau(f)\geqslant p\}}\big) > \lambda_3\Big),$$

$$\mathbb{P}_4 = \mathbb{P}^*\left(\sup_{f\in\mathcal{F}} \widetilde{\nu_n}\left(\sum_{k=1}^{p-1}\big(f - \Pi_{k-1}(f)\big)\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)>u_{k-1}\}}\right) > \lambda_4\right),$$

$$\mathbb{P}_5 = \mathbb{P}^*\left(\sup_{f\in\mathcal{F}} \tilde{\nu}_n\left(\sum_{k=1}^{p-1}(f - \Pi_k(f))\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)\leqslant u_{k-1}\}}\right) > \lambda_5\right),$$

$$\mathbb{P}_6 = \mathbb{P}\left(\sup_{f\in\mathcal{F}} \tilde{\nu}_n\left(\sum_{k=1}^{p-1}(\Pi_k(f) - \Pi_{k-1}(f))\mathbf{1}_{\{\tau(f)\geqslant k,\Delta_k(f)\leqslant u_{k-1}\}}\right) > \lambda_6\right).$$

The λ_j's will be chosen later on.

4.3.3. Controls of $\mathbb{P}_1$ to $\mathbb{P}_6$

We shall often use the following two arguments.

ARGUMENT 1. – *For any j in $\mathbb{N}$ and any set A,*

$$\tilde{\nu}_n((f - \Pi_j(f))\mathbf{1}_A) \leqslant \sqrt{n}\,\mathrm{P}(\Delta_j\mathbf{1}_A).$$

Proof. – We have $0 \leqslant (\Pi_j(f) - f)\mathbf{1}_A \leqslant \Delta_j\mathbf{1}_A$ so, since $(f - \Pi_j(f))\mathbf{1}_A \leqslant 0$,

$$\begin{aligned}\tilde{\nu}_n((f - \Pi_j(f))\mathbf{1}_A) &= \sqrt{n}[\mathrm{P}_n((f - \Pi_j(f))\mathbf{1}_A) - \mathrm{P}((f - \Pi_j(f))\mathbf{1}_A)]\\ &\leqslant \sqrt{n}\,\mathrm{P}((-f + \Pi_j(f))\mathbf{1}_A),\\ \tilde{\nu}_n((f - \Pi_j(f))\mathbf{1}_A) &\leqslant \sqrt{n}\,\mathrm{P}(\Delta_j\mathbf{1}_A). \quad \square\end{aligned}$$

ARGUMENT 2 (Markov's inequality). – *For any nonnegative random variable Z and any positive real number ζ,*

$$\mathrm{E}(Z\mathbf{1}_{\{Z>\zeta\}}) \leqslant \frac{1}{\zeta}\mathrm{E}(Z^2).$$

Control of $\mathbb{P}_1$. In order to control $\mathbb{P}_1$, we shall use essentially Hoeffding's inequality and a special version of Bernstein's inequality due to Birgé and Massart [3]. These inequalities provide controls in probability for the partial sum S_n of independent and centered variables. In fact, since they rely on Chernoff's bound, it follows from classical martingale arguments (mainly Doob's inequality) that they remain valid for $\sup_{1\leqslant k\leqslant n} S_k$ instead of S_n. The resulting inequalities are stated below without proof (for more details about the combination of Chernoff's bound and martingale arguments see for instance Shorack and Wellner [21], pp. 444–445).

LEMMA 3. – *Let $U_1, \ldots, U_n$ be independent random variables such that $m \leqslant U_i \leqslant M$ for $1 \leqslant i \leqslant n$. Let $a = M - m$, and let $S_k = \sum_{i=1}^{k}(U_i -$*

$\mathrm{E}(U_i))$. *Then, for any positive* λ,

$$\mathbb{P}\Big(\sup_{1\leqslant k\leqslant n} S_k \geqslant n\lambda\Big) \leqslant \exp\Big(-\frac{2n\lambda^2}{a^2}\Big).$$

LEMMA 4. – *Let* $U_1, \ldots, U_n$ *be independent random variables satisfying for some positive constants* δ *and* c *and any* $m \geqslant 2$ *the following moment condition*

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(|U_i^m|) \leqslant \frac{m!}{2}\delta^2 c^{m-2}. \tag{12}$$

Let $S_k = \sum_{i=1}^{k}(U_i - \mathrm{E}(U_i))$, *for* $1 \leqslant k \leqslant n$. *Then, for any positive* λ,

$$\mathbb{P}\Big(\sup_{1\leqslant k\leqslant n} S_k > n\lambda\Big) \leqslant \mathrm{e}^{-nh_{c,\delta}(\lambda)},$$

where

$$h_{c,\delta}(\lambda) = \frac{\delta^2}{c^2}\psi\Big(\frac{c\lambda}{\delta^2}\Big),$$

with $\psi(t) = t + 1 - \sqrt{1+2t}$. *Moreover,*

$$h_{c,\delta}^{-1}(\zeta) = \delta\sqrt{2\zeta} + c\zeta.$$

In what follows, we shall use Lemma 4 repeatedly in the situation where $|U_i| \leqslant b$ and $\mathrm{E}(U_i^2) \leqslant \delta^2$. In that case, inequality (12) holds with $c = b/3$. We shall now control $\mathbb{P}_1$. When f describes the set $\mathcal{F}$, there are at most $\mathrm{e}^{H(\varepsilon_0)}$ functions of the type $\Pi_0(f)$. Since each of these functions satisfies $m \leqslant \Pi_0(f) \leqslant M$, it follows from Lemma 3 that for any $\lambda_1 > 0$

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}} \widetilde{\nu}_n(\Pi_0(f)) \geqslant \lambda_1\Big) \leqslant \mathrm{e}^{H(\varepsilon_0)}\,\mathrm{e}^{-2\lambda_1^2/a^2}. \tag{13}$$

Moreover, we note that

$$|\Pi_0(f) - \mathrm{P}(\Pi_0(f))| \leqslant a = M - m \quad \text{since } m \leqslant \Pi_0(f) \leqslant M$$

and

$$\begin{aligned}
&\|\Pi_0(f) - \mathrm{P}(\Pi_0(f))\|_2 \\
&\quad\leqslant \|\Pi_0(f) - f - \mathrm{P}(\Pi_0(f) - f)\|_2 + \|f - \mathrm{P}(f)\|_2 \\
&\quad\leqslant \varepsilon_0 + \sigma.
\end{aligned}$$

Applying Lemma 4, with $U_i = \Pi_0(f)(Z_i) - \mathrm{P}(\Pi_0(f))$, we obtain

$$\mathbb{P}\big(\widetilde{\nu}_n(\Pi_0(f)) > \lambda_1\big) \leqslant \mathrm{e}^{-h_{c,\varepsilon_0+\sigma}(\lambda_1)},$$

where $c = a/3$. This implies that

$$\mathbb{P}\Big(\widetilde{\nu}_n(\Pi_0(f)) > \lambda_1 + \varepsilon_0\sqrt{2h_{c,\sigma}(\lambda_1)}\Big) \leqslant \mathrm{e}^{-h_{c,\sigma}(\lambda_1)},$$

since $h^{-1}_{c,\varepsilon_0+\sigma}(h_{c,\sigma}(\lambda_1)) = \lambda_1 + \varepsilon_0\sqrt{2h_{c,\sigma}(\lambda_1)}$. Therefore, setting $\lambda_1 = \lambda$, we get

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}} \widetilde{\nu}_n(\Pi_0(f)) > \lambda + \varepsilon_0\sqrt{2h(\lambda)}\Big) \leqslant \mathrm{e}^{H(\varepsilon_0)}\,\mathrm{e}^{-h(\lambda)},$$

where $h(\lambda)$ can be taken as $2\lambda^2/a^2 = h_1(\lambda)$ or $h_{c,\sigma}(\lambda) = h_2(\lambda)$.

Control of $\mathbb{P}_2$. From Argument 1, we get

$$\widetilde{\nu}_n\big((f - \Pi_0(f))\mathbf{1}_{\{\tau(f)=0\}}\big) \leqslant \sqrt{n}\,\mathrm{P}\big(\Delta_0(f)\mathbf{1}_{\{\tau(f)=0\}}\big)$$

and, since $\tau(f) = 0$ implies $\Delta_0(f) > u_0$, we obtain by Argument 2

$$\widetilde{\nu}_n\big((f - \Pi_0(f))\mathbf{1}_{\{\tau(f)=0\}}\big) \leqslant \frac{\sqrt{n}}{u_0}\mathrm{P}\big(\Delta_0^2(f)\big) \leqslant \frac{\sqrt{n}}{u_0}\varepsilon_0^2.$$

Setting $\lambda_2 = \sqrt{n}\,\varepsilon_0^2/u_0$, we conclude

$$\mathbb{P}^*\Big(\sup_{f\in\mathcal{F}} \widetilde{\nu}_n\big((f - \Pi_0(f))\mathbf{1}_{\{\tau(f)=0\}}\big) > \lambda_2\Big) = 0.$$

Control of $\mathbb{P}_3$. From Argument 1, we get

$$\widetilde{\nu}_n\big((f - \Pi_{p-1}(f))\mathbf{1}_{\{\tau(f)\geqslant p\}}\big) \leqslant \sqrt{n}\,\mathrm{P}\big(\Delta_{p-1}(f)\mathbf{1}_{\{\tau(f)\geqslant p\}}\big)$$

and, by Cauchy–Schwarz inequality,

$$\widetilde{\nu}_n\big((f - \Pi_{p-1}(f))\mathbf{1}_{\{\tau(f)\geqslant p\}}\big) \leqslant \sqrt{n}\,\mathrm{P}^{1/2}\big(\Delta_{p-1}^2(f)\big) \leqslant \sqrt{n}\,\varepsilon_{p-1}.$$

Setting $\lambda_3 = \sqrt{n}\,\varepsilon_{p-1}$, we conclude

$$\mathbb{P}^*\Big(\sup_{f\in\mathcal{F}} \widetilde{\nu}_n\big((f - \Pi_{p-1}(f))\mathbf{1}_{\{\tau(f)\geqslant p\}}\big) > \lambda_3\Big) = 0.$$

Control of $\mathbb{P}_4$. From Argument 1, we get

$$\widetilde{\nu}_n\big((f-\Pi_{k-1}(f))\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)>u_{k-1}\}}\big)$$
$$\leqslant \sqrt{n}\,\mathrm{P}\big(\Delta_{k-1}(f)\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)>u_{k-1}\}}\big)$$

and, since $\tau(f)=k$ implies $\Delta_{k-1}(f)\leqslant u_{k-1}$, we get

$$\widetilde{\nu}_n\big((f-\Pi_{k-1}(f))\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)>u_{k-1}\}}\big)\leqslant \sqrt{n}\,\mathrm{P}\big(\Delta_k(f)\mathbf{1}_{\{\Delta_k(f)>u_{k-1}\}}\big).$$

Argument 2 provides us with

$$\widetilde{\nu}_n\big((f-\Pi_{k-1}(f))\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)>u_{k-1}\}}\big)\leqslant \frac{\sqrt{n}}{u_{k-1}}\mathrm{P}\big(\Delta_k^2(f)\big)\leqslant \frac{\sqrt{n}}{u_{k-1}}\varepsilon_k^2.$$

We deduce from the above inequality that, setting $\lambda_4=\sqrt{n}\sum_{k=1}^{p-1}\varepsilon_k^2/u_{k-1}$,

$$\mathbb{P}^*\left(\sup_{f\in\mathcal{F}}\widetilde{\nu}_n\sum_{k=1}^{p-1}\big((f-\Pi_{k-1}(f))\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)>u_{k-1}\}}\big)>\lambda_4\right)=0.$$

Control of $\mathbb{P}_5$. From Argument 1, we get

$$\widetilde{\nu}_n\big((f-\Pi_k(f))\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)\leqslant u_{k-1}\}}\big)\leqslant \sqrt{n}\,\mathrm{P}\big(\Delta_k(f)\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)\leqslant u_{k-1}\}}\big)$$

and, since $\tau(f)=k$ implies $\Delta_k(f)>u_k$,

$$\widetilde{\nu}_n\big((f-\Pi_k(f))\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)\leqslant u_{k-1}\}}\big)\leqslant \sqrt{n}\,\mathrm{P}\big(\Delta_k(f)\mathbf{1}_{\{\Delta_k(f)>u_k\}}\big).$$

Argument 2 provides us with

$$\widetilde{\nu}_n\big((f-\Pi_k(f))\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)\leqslant u_{k-1}\}}\big)\leqslant \frac{\sqrt{n}}{u_k}\mathrm{P}\big(\Delta_k^2(f)\big)\leqslant \frac{\sqrt{n}}{u_k}\varepsilon_k^2.$$

We deduce from the above inequality that, setting $\lambda_5=\sqrt{n}\sum_{k=1}^{p-1}\varepsilon_k^2/u_k$,

$$\mathbb{P}^*\left(\sup_{f\in\mathcal{F}}\widetilde{\nu}_n\sum_{k=1}^{p-1}\big((f-\Pi_k(f))\mathbf{1}_{\{\tau(f)=k,\Delta_k(f)\leqslant u_{k-1}\}}\big)>\lambda_5\right)=0.$$

Control of $\mathbb{P}_6$. Let

$$V_k(f)=\big(\Pi_k(f)-\Pi_{k-1}(f)\big)\mathbf{1}_{\{\tau(f)\geqslant k,\Delta_k(f)\leqslant u_{k-1}\}}.$$

We have in view to use Lemma 4 to control $\widetilde{\nu_n}(V_k(f))$ for each k and then sum up the resulting probability bounds to handle $\mathbb{P}_6$. We can bound from above $|V_k(f)|$ by $2u_{k-1}$ and $\mathrm{P}(V_k(f)^2)$ by $9\,\varepsilon_k^2$. Indeed,

- since $\tau(f) \geqslant k$ implies $\Delta_{k-1}(f) \leqslant u_{k-1}$,

$$\begin{aligned}|V_k(f)| &\leqslant \big(|f - \Pi_k(f)| + |f - \Pi_{k-1}(f)|\big)\mathbf{1}_{\{\tau(f)\geqslant k, \Delta_k(f)\leqslant u_{k-1}\}}\\ &\leqslant \big(\Delta_k(f) + \Delta_{k-1}(f)\big)\mathbf{1}_{\{\Delta_{k-1}(f)\leqslant u_{k-1}, \Delta_k(f)\leqslant u_{k-1}\}}\\ &\leqslant 2u_{k-1};\end{aligned}$$

- by triangular inequality and since $\varepsilon_{k-1} = 2\varepsilon_k$,

$$\begin{aligned}\mathrm{P}\big(V_k(f)^2\big) &\leqslant \mathrm{P}\big[\big(\Delta_k(f) + \Delta_{k-1}(f)\big)^2\big]\\ &\leqslant \big[\mathrm{P}^{1/2}\big(\Delta_k(f)^2\big) + \mathrm{P}^{1/2}\big(\Delta_{k-1}(f)^2\big)\big]^2\\ &\leqslant (\varepsilon_k + \varepsilon_{k-1})^2\\ &\leqslant 9\varepsilon_k^2.\end{aligned}$$

Lemma 4 with $c = 2u_{k-1}/3$ and $\delta = 3\varepsilon_k$ yields for any $\zeta_k > 0$

$$\mathbb{P}\left(\widetilde{\nu_n}\big(V_k(f)\big) > \frac{2u_{k-1}}{3\sqrt{n}}\zeta_k + 3\sqrt{2}\,\varepsilon_k\sqrt{\zeta_k}\right) \leqslant \mathrm{e}^{-\zeta_k}.$$

We now control the quantity $\sup_{f\in\mathcal{F}} \widetilde{\nu_n}(V_k(f))$. We recall that

$$V_k(f) = \big(\Pi_k(f) - \Pi_{k-1}(f)\big)\mathbf{1}_{\{\Delta_0(f)\leqslant u_0, \ldots, \Delta_{k-1}(f)\leqslant u_{k-1}, \Delta_k(f)\leqslant u_{k-1}\}}.$$

When f varies in $\mathcal{F}$, there are at most $\prod_{j=0}^{k} \mathrm{e}^{H(\varepsilon_j)}$ functions $V_k(f)$. Therefore, setting $\mathbb{H}_k = \sum_{j=0}^{k} H(\varepsilon_j)$, we get:

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}} \widetilde{\nu_n}\big(V_k(f)\big) > \frac{2u_{k-1}}{3\sqrt{n}}\zeta_k + 3\sqrt{2}\,\varepsilon_k\sqrt{\zeta_k}\right) \leqslant \mathrm{e}^{\mathbb{H}_k-\zeta_k}.$$

Hence, for any $\zeta_1, \ldots, \zeta_{p-1} > 0$,

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}} \widetilde{\nu_n}\left(\sum_{k=1}^{p-1} V_k(f)\right) > \frac{2}{3\sqrt{n}}\sum_{k=1}^{p-1} u_{k-1}\zeta_k + 3\sqrt{2}\sum_{k=1}^{p-1}\varepsilon_k\sqrt{\zeta_k}\right)$$
$$\leqslant \sum_{k=1}^{p-1} \mathrm{e}^{\mathbb{H}_k-\zeta_k}.$$

Note that H is a nonincreasing function so

$$\mathbb{H}_k = \sum_{j=0}^{k} H(\varepsilon_j) \geqslant (k+1)H(\varepsilon_0).$$

It follows that, if we set $\zeta_k = 2\mathbb{H}_k + h(\lambda)$, we obtain

$$\sum_{k=1}^{p-1} \mathrm{e}^{\mathbb{H}_k - \zeta_k} = \mathrm{e}^{-h(\lambda)} \sum_{k=1}^{p-1} \mathrm{e}^{-\mathbb{H}_k} \leqslant \mathrm{e}^{-h(\lambda)} \sum_{k\geqslant 1} \mathrm{e}^{-(k+1)H(\varepsilon_0)}.$$

Moreover, $H(\varepsilon_0) \geqslant 1$ by definition, so

$$\sum_{k=1}^{p-1} \mathrm{e}^{\mathbb{H}_k - \zeta_k} \leqslant \mathrm{e}^{-h(\lambda)} \sum_{k\geqslant 2} \mathrm{e}^{-k} \leqslant 0.215\, \mathrm{e}^{-h(\lambda)}.$$

Setting $K = \sum_{k\geqslant 2} \mathrm{e}^{-k}$ and

$$\lambda_6 = \frac{2}{3\sqrt{n}} \sum_{k=1}^{p-1} u_{k-1}\zeta_k + 3\sqrt{2} \sum_{k=1}^{p-1} \varepsilon_k \sqrt{\zeta_k},$$

we get at last

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}} \widetilde{\nu_n} \left(\sum_{k=1}^{p-1} V_k(f) \right) > \lambda_6 \right) \leqslant K\, \mathrm{e}^{-h(\lambda)}. \tag{14}$$

4.3.4. Conclusion

Collecting inequalities (11) and (13) to (14) gives, setting $\Lambda = \sum_{j=1}^{6} \lambda_j$,

$$\mathbb{P}^*\left(\sup_{f\in\mathcal{F}} \widetilde{\nu_n}(f) > \Lambda \right) \leqslant \left(\mathrm{e}^{H(\varepsilon_0)} + K \right) \mathrm{e}^{-h(\lambda)}. \tag{15}$$

It remains to bound from above Λ.

$$\begin{aligned}\Lambda - \lambda_1 = {} & \sqrt{n}\,\frac{\varepsilon_0^2}{u_0} + \sqrt{n} \sum_{k=1}^{p-1} \frac{\varepsilon_k^2}{u_{k-1}} + \sqrt{n} \sum_{k=1}^{p-1} \frac{\varepsilon_k^2}{u_k} + \frac{2}{3\sqrt{n}} \sum_{k=1}^{p-1} u_{k-1}\zeta_k \\ & + 3\sqrt{2} \sum_{k=1}^{p-1} \varepsilon_k \sqrt{\zeta_k} + \sqrt{n}\, \varepsilon_{p-1}\end{aligned}$$

$$\leqslant \sqrt{n}\sum_{k=1}^{p}\left(\frac{\varepsilon_k^2}{u_{k-1}}+\frac{\varepsilon_{k-1}^2}{u_{k-1}}\right)+\frac{2}{3\sqrt{n}}\sum_{k=1}^{p-1}u_{k-1}\zeta_k$$

$$+3\sqrt{2}\sum_{k=1}^{p-1}\varepsilon_k\sqrt{\zeta_k}+\sqrt{n}\,\varepsilon_{p-1}$$

$$\leqslant 5\sqrt{n}\sum_{k=1}^{p}\frac{\varepsilon_k^2}{u_{k-1}}+\frac{2}{3\sqrt{n}}\sum_{k=1}^{p}u_{k-1}\zeta_k+3\sqrt{2}\sum_{k=1}^{p-1}\varepsilon_k\sqrt{\zeta_k}+\sqrt{n}\,\varepsilon_{p-1},$$

since $\varepsilon_{k-1}=2\varepsilon_k$. In order to minimize the previous bound, we will now choose the sequence $(u_k)_{k\in\mathbb{N}}$ such that, for $k\geqslant 1$,

$$5\sqrt{n}\,\varepsilon_k^2/u_{k-1}=2u_{k-1}\zeta_k/\big(3\sqrt{n}\big),$$

which is equivalent to $u_{k-1}=\sqrt{15n}\,\varepsilon_k/\sqrt{2\zeta_k}$ (observe that, from the definitions $\varepsilon_k=2^{-k}\varepsilon_0$ and $\zeta_k=2\mathbb{H}_k+h(\lambda)$, the sequence $(u_k)_{k\in\mathbb{N}}$ is indeed decreasing). Replacing u_{k-1}, we get the following upper bound for $\Lambda-\lambda_1$.

$$\Lambda-\lambda_1\leqslant 2\sqrt{\frac{10}{3}}\sum_{k=1}^{p}\varepsilon_k\sqrt{\zeta_k}+3\sqrt{2}\sum_{k=1}^{p-1}\varepsilon_k\sqrt{\zeta_k}+\sqrt{n}\,\varepsilon_{p-1}.$$

We recall that $\sqrt{n}\,\varepsilon_{p-1}=\sqrt{n}\,\varepsilon_0/2^{p-1}$ tends to 0 as $p\to+\infty$. This allows to bound from above $\Lambda-\lambda_1$, for any $\varepsilon>0$, by

$$\left(2\sqrt{\frac{10}{3}}+3\sqrt{2}+\varepsilon\right)\sum_{k=1}^{+\infty}\varepsilon_k\sqrt{\zeta_k}.$$

We will now control $\sum_{k=1}^{+\infty}\varepsilon_k\sqrt{\zeta_k}$. By definition of ζ_k, $\mathbb{H}_k$ and ε_k,

$$\sum_{k=1}^{+\infty}\varepsilon_k\sqrt{\zeta_k}\leqslant\sqrt{2}\sum_{k=1}^{+\infty}\varepsilon_k\sqrt{\sum_{j=0}^{k}H(\varepsilon_j)}+\sqrt{h(\lambda)}\,\varepsilon_0.$$

Moreover,

$$\sum_{k=1}^{+\infty}\varepsilon_k\sqrt{\sum_{j=0}^{k}H(\varepsilon_j)}\leqslant\sum_{k\geqslant 1,\ 0\leqslant j\leqslant k}\varepsilon_k\sqrt{H(\varepsilon_j)}$$

$$\leqslant\sum_{j\geqslant 0}\sqrt{H(\varepsilon_j)}\sum_{k\geqslant j}\varepsilon_k\leqslant 2\sum_{j\geqslant 0}\varepsilon_j\sqrt{H(\varepsilon_j)}\leqslant 4\int_0^{\varepsilon_0}\sqrt{H(x)}\,\mathrm{d}x.$$

Collecting the above evaluations we obtain, for any $\varepsilon > 0$,

$$\Lambda - \lambda_1 \leqslant \left(2\sqrt{\frac{10}{3}} + 3\sqrt{2} + \varepsilon\right)\left(4\sqrt{2}\varphi(\varepsilon_0) + \varepsilon_0\sqrt{h(\lambda)}\right)$$

which leads, recalling that $\lambda_1 = \lambda$, to

$$\Lambda \leqslant \lambda + 7.9\varepsilon_0\sqrt{h(\lambda)} + 44.7\varphi(\varepsilon_0).$$

It follows from inequality (15) that

$$\mathbb{P}^*\left(\sup_{f\in\mathcal{F}} \widetilde{\nu}_n(f) > \lambda + 7.9\varepsilon_0\sqrt{h(\lambda)} + 44.7\varphi(\varepsilon_0)\right) \leqslant (1+K)\,\mathrm{e}^{H(\varepsilon_0)}\,\mathrm{e}^{-h(\lambda)},$$

where $= \sum_{k\geqslant 2}\mathrm{e}^{-k} \leqslant 0.25$. Since $\sigma^2 \leqslant a^2/4$ and $\psi(t) \leqslant t^2/2$, we have that $h(\lambda) \leqslant \lambda^2/2\sigma^2$. This concludes the proof of Theorem 5. □

4.4. Proof of Corollary 2

Let $H(\varepsilon)$ be bounded by γ/ε. By Theorem 5,

$$\mathbb{P}^*\left(\sup_{1\leqslant k\leqslant n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}|\nu_k(f)| > \lambda + C_3\varepsilon_0\lambda + C_4\sqrt{\varepsilon_0}\right) \leqslant 2.5\,\mathrm{e}^{\gamma/\varepsilon_0}\,\mathrm{e}^{-2\lambda^2/a^2},$$

where C_3 and C_4 are absolute constants.

Setting $\zeta = \lambda(1 + C_3\varepsilon_0) + C_4\sqrt{\varepsilon_0}$, it follows that for any $\zeta \geqslant C_4\sqrt{\varepsilon_0}$,

$$\mathbb{P}^*\left(\sup_{1\leqslant k\leqslant n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}|\nu_k(f)| > \zeta\right) \leqslant 2.5\,\mathrm{e}^{\gamma/\varepsilon_0}\exp\left[-\frac{2}{a^2}\left(\frac{\zeta - C_4\sqrt{\varepsilon_0}}{1 + C_3\varepsilon_0}\right)^2\right].$$

Using that $1/(1+t)^2 \geqslant (1-t)^2$ for any $0 \leqslant t \leqslant \sqrt{2}$, we get

$$\begin{aligned}\left(\frac{\zeta - C_4\sqrt{\varepsilon_0}}{1 + C_3\varepsilon_0}\right)^2 &\geqslant \left(\zeta - C_4\sqrt{\varepsilon_0}\right)^2(1 - C_3\varepsilon_0)^2\\ &\geqslant \zeta^2 - 2C_3\zeta^2\varepsilon_0 - 4C_4\sqrt{\varepsilon_0}\,\zeta,\end{aligned}$$

as soon as $\varepsilon_0 < 1/(2C_3)$.

Setting $\varepsilon_0 = 1/\zeta$, there exists absolute constants ζ^* and C_5 such that, for any $\zeta \geqslant \zeta^*$,

$$\mathbb{P}^*\left(\sup_{1\leqslant k\leqslant n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}|\nu_k(f)| > \zeta\right) \leqslant 2.5\,\mathrm{e}^{-2\zeta^2/a^2 + C_5\zeta}.$$

Possibly enlarging C_5, previous inequality holds for any $\zeta > 0$.

This concludes the proof of Corollary 2. □

4.5. Proof of Corollary 3

We claim that the function h_2 satisfies for any positive λ and K, the inequality $h_2(\lambda+K) \geqslant h_2(\lambda)+h_2(K)$. Indeed, since $\psi^{-1}(x) = \sqrt{2x}+x$, it is clear that $\psi^{-1}(x+y) \leqslant \psi^{-1}(x) + \psi^{-1}(y)$. Therefore, the reverse inequality holds for h_2. Let

$$K = h_2^{-1}\left(\left(\frac{\varphi(\theta\sigma)}{\theta\sigma}\right)^2\right) = \sigma\sqrt{2}\,\frac{\varphi(\theta\sigma)}{\theta\sigma} + \frac{a}{3\sqrt{n}}\left(\frac{\varphi(\theta\sigma)}{\theta\sigma}\right)^2.$$

Since $H^{1/2}(\theta\sigma) \leqslant \frac{\varphi(\theta\sigma)}{\theta\sigma}$, we get $K \geqslant h_2^{-1}(H(\theta\sigma))$.

We now apply inequality (10) to $\lambda' = \lambda + K$, $\varepsilon_0 = \theta\sigma$ for some $\theta \in]0, 1]$. We get, possibly enlarging C,

$$\mathbb{P}^*\left(\sup_{1\leqslant k\leqslant n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}\left|\nu_k(f)\right| > \lambda(1+C\theta) + C\frac{\varphi(\theta\sigma)}{\theta}\left(1+\frac{\varphi(\theta\sigma)}{\theta\sigma^2\sqrt{n}}\right)\right)$$
$$\leqslant 2.5\,\mathrm{e}^{-h_2(\lambda)}.$$

Since $\psi(\lambda) \geqslant \frac{\lambda^2}{2\lambda+2}$, we get

$$h_2(\lambda) \geqslant \frac{\lambda^2}{2\sigma^2 + \frac{2a\lambda}{3\sqrt{n}}}.$$

This concludes the proof of Corollary 3. □

5. CONCLUSION

To conclude we would like to mention several questions of interest that derive from the results of this paper.

5.1. Substituting $\widehat{G_n}$ for G

For statistical purposes, it would be desirable to replace the factor $1 - G$ by $1 - \widehat{G_n}$ in the exponential inequality (1). We do not know whether such an inequality holds nonasymptotically. All we can do is to provide an asymptotic evaluation which is analogous to Theorem 3.

THEOREM 6. – *Let $\widehat{F_n}$ be the Kaplan–Meier estimator of F, $\widehat{G_n}$ be the Kaplan–Meier estimator of the distribution function G of the censoring times, and $W_{(n)}$ be the largest observation. Let B^0 be a Brownian bridge. Then, for any positive λ,*

$$\limsup_{n\to+\infty} \mathbb{P}\Big(\sup_{t\leqslant W_{(n)}} \sqrt{n}\,\big|(1-\widehat{G_n})(\widehat{F_n}-F)(t)\big| > \lambda\Big) \leqslant \mathbb{P}\Big(\sup_{t\in[0,1]} \big|B^0(t)\big| > \lambda\Big). \tag{16}$$

Inequality (16) actually allows to build confidence bands for F in the spirit of Hall and Wellner [17] and Gill [14]. Nevertheless, the conservatism of these bands may make them useless under heavy censorship.

5.2. Proof of Theorem 6

Using inequality (6) and the decomposition

$$\sqrt{n}(1-\widehat{G_n})(\widehat{F_n}-F) = \sqrt{n}(1-G)(\widehat{F_n}-F) + \sqrt{n}(G-\widehat{G_n})(\widehat{F_n}-F)$$

we have to show that

$$\sup_{t\leqslant W_{(n)}} \sqrt{n}\,\big|(\widehat{G_n}-G)(\widehat{F_n}-F)\big| \to 0 \quad \text{in probability.}$$

We first notice that the problem is symmetric with respect to F and G. This allows to assume that

$$\tau = \inf\{x \in \mathbb{R},\ (1-G(x)) = 0\}.$$

We shall distinguish two cases: either $G^-(\tau) < 1$, or $G^-(\tau) = 1$.

In the first case, we write for $t \leqslant W_{(n)}$,

$$\sqrt{n}\,\big|(\widehat{G_n}-G)(\widehat{F_n}-F)\big|(t) = \left(\frac{\widehat{G_n}-G}{1-G^-}\right)(t)\sqrt{n}(1-G^-)(\widehat{F_n}-F)(t).$$

It follows from Theorem 2 that the supremum on $\mathbb{R}$ of the process $|\sqrt{n}(1-G^-)(\widehat{F_n}-F)|$ is bounded in probability. Moreover,

$$\sup_{t\leqslant W_{(n)}} \left|\frac{(\widehat{G_n}-G)(t)}{(1-G^-)(t)}\right| \leqslant \frac{1}{1-G^-(\tau)} \sup_{t\leqslant W_{(n)}} \big|(\widehat{G_n}-G)(t)\big|.$$

We conclude with a result due to Wang [27] which ensures that

$$\sup_{t\leqslant W_{(n)}} \left|\left(\widehat{G_n}-G\right)(t)\right| \to 0 \quad \text{in probability.}$$

If $G^-(\tau)=1$, either $F^-(\tau)<1$ and we can argue as above by exchanging F for G, or $F^-(\tau)=1$. If it is so, F and G are continuous at point τ, therefore $\mathbb{P}(W_{(n)}=\tau)=0$, and $G(W_{(n)})<1$ a.s. Let $\sigma<\tau$. Since $F^-(\sigma)<1$, it follows from Remark 3 that $\sup_{t\leqslant\sigma}|(\widehat{G_n}-G)(t)|\to 0$ a.s. Using this result, and similar arguments as above, we show that

$$\sup_{t\leqslant\sigma}\sqrt{n}\left|\left(\widehat{G_n}-G\right)\left(\widehat{F_n}-F\right)\right| \to 0 \quad \text{in probability.}$$

Now,

$$\begin{aligned}
&\sup_{\sigma\leqslant t\leqslant\sigma\vee W_{(n)}} \sqrt{n}\left|\left(\widehat{G_n}-G\right)\left(\widehat{F_n}-F\right)\right| \\
&\quad\leqslant \sup_{\sigma\leqslant t\leqslant\sigma\vee W_{(n)}} \sqrt{n}\left|(1-G)\left(\widehat{F_n}-F\right)\right| \sup_{t\leqslant W_{(n)}}\left|\frac{\widehat{G_n}-G}{1-G}\right|.
\end{aligned}$$

It follows from Gill [12], Lemma 2.6 that $\sup_{t\leqslant W_{(n)}}|(\widehat{G_n}-G)/(1-G)|$ is bounded in probability. Moreover, it follows from the convergence of the process $\sqrt{n}(1-G)(\widehat{F_n}-F)$ towards $((1-F)(1-G)/(1-K))\times B^0(K)$ in $\mathcal{D}(\mathcal{T})$ that

$$\begin{aligned}
&\limsup_{n\to+\infty}\mathbb{P}\Big(\sup_{\sigma\leqslant t\leqslant\tau}\sqrt{n}\left|(1-G)\left(\widehat{F_n}-F\right)\right|>\lambda\Big) \\
&\quad\leqslant \mathbb{P}\left(\sup_{\sigma\leqslant t\leqslant\tau}\left|\frac{(1-F)(1-G)}{1-K}B^0(K)(t)\right|>\lambda\right).
\end{aligned}$$

Since K is a $[0,1]$ valued, continuous and nondecreasing function, we shall consider the case of $\lim_{t\to\tau}K(t)<1$ and the case of this limit equals 1. In the first case, since F and G are continuous at point τ and tend to 1, we get

$$\lim_{t\to\tau}\frac{(1-F)(1-G)}{1-K}(t)=0.$$

In the second case, we get $(B^0(K))(\tau)=0$ a.s. Moreover, the process $B^0(K)$ is continuous. Finally, we obtain in both cases,

$$\mathbb{P}\left(\sup_{\sigma\leqslant t\leqslant\tau}\left|\frac{(1-F)(1-G)}{1-K}B^0(K)(t)\right|>\lambda\right) \xrightarrow[\sigma\to\tau]{\text{a.s.}} 0.$$

This concludes the proof of Theorem 6.

5.3. Normalizing factor

One could wonder whether the normalizing factor $(1-G)$ used in Theorem 1 is adequate or not. If one tries to answer this question, it is interesting first to notice that some weighting is actually necessary if one wants to get a uniform result on the real line since otherwise one really gets into trouble due to the bad behaviour of the Kaplan–Meier estimator in the tail (see Stute [22]). In order to better understand what kind of weighting should be considered, let us analyse the asymptotic behaviour of the Kaplan–Meier process $\sqrt{n}(\widehat{F_n}-F)$. Let

$$C(t)=\int_0^t \frac{\mathrm{d}F(s)}{(1-F^-(s))^2(1-G^-)(s)}$$

and let $K(t)=\frac{C(t)}{1+C(t)}$ if $C(t)<+\infty$ and $K(t)=1$ if $C(t)=+\infty$. It follows from Gill [12] that the process $\sqrt{n}(1-G)^{1/2}(\widehat{F_n}-F)$ converges in $\mathcal{D}(\mathcal{T})$ towards $\{(1-F)(1-G)^{1/2}/(1-K)\}B^0(K)$, where $\mathcal{T}=\{t, G^-(t)<1\}$. One can notice that the variance of the limiting process is uniformly bounded since it is equal to $(1-G)(1-F)^2C$. Therefore, a question arises: is it possible to replace in the exponential inequality (1) the normalizing factor $1-G$ by $(1-G)^{1/2}$. Our method does not allow to do this since it is based on the use of the influence function of the process $\sqrt{n}(\widehat{F_n}-F)$. The influence function has to be normalized by $(1-G^-)$ (or $1-G$), in order to be stabilized on the whole interval. It is not clear at all whether another method could work since it could happen that the size of the supremum norm of the influence function indeed plays a role in an exponential bound like ours.

APPENDIX A

A.1. Proof of Lemma 2

In order to get the $\mathbb{L}^2(\mathrm{P})$-entropy with bracketing of the class $\mathcal{J}$, we will use Lemma 1. We recall that

$$\mathcal{J}=\{J_{K,x},\ K\in\mathcal{K},\ x\in\mathbb{R}\},$$

where $J_{K,x}(w,\delta)$ may be written under the form

$$J_{K,x}(w,\delta)=\big(\delta J^{(1)}_{K,x}(w)+J^{(2)}_{K,x}(w)\big)J^{(3)}_{K,x}(w),$$

if we set

$$J^{(1)}_{K,x}(w) = (1-K)(x)(1-G^-)(x)/\big((1-K)(1-G^-)\big)(w \wedge x),$$

$$J^{(2)}_{K,x}(w) = (1-K)(x)(1-G^-)(x)\int_{w\wedge x}^{x} \big(1/(1-G^-)\big)\,\mathrm{d}\big(1/(1-K)\big),$$

$$J^{(3)}_{K,x}(w) = \mathbf{1}_{]-\infty,x]}(w).$$

We denote by $\mathcal{J}^{(j)}$, $j=1,2,3$, the classes of functions defined by

$$\mathcal{J}^{(j)} = \big\{w \in \mathbb{R} \to J^{(j)}_{K,x}(w),\ K \in \mathcal{K}, x \in \mathbb{R}\big\}.$$

Since the functions $J^{(j)}_{K,x}$ are $[0,1]$ valued and monotone, it follows from Lemma 1 that there exists some constant η such that for $j=1,2,3$, the $\mathbb{L}^2(\mathrm{P})$-entropy with bracketing H_j of the class $\mathcal{J}^{(j)}$ satisfies $H_j(\varepsilon) \leqslant \eta/\varepsilon$. By definition of the entropy, we can find sets of brackets $\mathcal{N}_j$ with diameter not larger than $\varepsilon/3$, covering $\mathcal{J}_j$, defined by

$$\mathcal{N}_j = \big\{[f^L_{j,i}, f^U_{j,i}],\ 1 \leqslant i \leqslant N\big\}, \quad j=1,2,3,$$

where N is bounded by $\exp(3\eta/\varepsilon)$. For each function $J^{(j)}_{K,x}$ in $\mathcal{J}^{(j)}$, $j=1,2,3$, we can find $i \in \{1,2,\ldots,N\}$ satisfying

$$f^L_{j,i} \leqslant J^{(j)}_{K,x} \leqslant f^U_{j,i}.$$

Since the functions $J^{(j)}_{K,x}$ are $[0,1]$ valued function, we can choose also $[0,1]$ valued functions to define the sets of brackets covering the classes $\mathcal{J}^{(j)}$, $j=1,2,3$.

With these sets, we can build a set of brackets with diameter not larger than ε covering the class $\mathcal{J}$. Indeed, for each function $J_{K,x}$ in $\mathcal{J}$ there exist i, i', i'' in $\{1,2,\ldots,N\}$ such that

$$\big(\delta f^L_{1,i} + f^L_{2,i'}\big) f^L_{3,i''} \leqslant J_{K,x} \leqslant \big(\delta f^U_{1,i} + f^U_{2,i'}\big) f^U_{3,i''}.$$

It is easy to verify that the $\mathbb{L}^2(\mathrm{P})$-norm of the function $(\delta f^L_{1,i} + f^L_{2,i'}) f^L_{3,i''} - (\delta f^U_{1,i} + f^U_{2,i'}) f^U_{3,i''}$ is bounded by ε. Hence, we found a set of brackets with diameter not larger than ε covering $\mathcal{J}$ with cardinality bounded by N^3. It follows that the $\mathbb{L}^2(\mathrm{P})$-entropy with bracketing H of the class $\mathcal{J}$ satisfies $H(\varepsilon) \leqslant \gamma/\varepsilon$ for some $\gamma \in \mathbb{R}$. This concludes the proof of Lemma 2.

REFERENCES

[1] R.F. BASS, Law of iterated logarithm for set-indexed partial sum processes with finite variance, *Z. Warscheinlichkeitstheor. Verw. Geb.* 70 (1985) 591–608.

[2] L. BIRGÉ and P. MASSART, Rates of convergence for minimum contrast estimators, *Probab. Theory Related Fields* 97 (1993) 113–150.

[3] L. BIRGÉ and P. MASSART, Minimum contrast estimators on sieves, *Bernoulli* 4 (3) (1998) 329–375.

[4] D. BITOUZÉ, Estimation de fonctionnelles d'une densité à partir d'observations directes ou censurées, Ph.D. Thesis, Laboratoire de modélisation statistique et stochastique, Bât. 425, Université de Paris XI, 91405 Orsay Cédex, France, 1995 (in English).

[5] N. BRESLOW and J. CROWLEY, A large sample study of the life table and product limit estimators under random censorship, *Ann. Statist.* 11 (1974) 49–58.

[6] I.H. DINWOODIE, Large deviations for censored data, *Ann. Statist.* 21 (3) (1993) 1608–1620.

[7] M.D. DONSKER, Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems, *Ann. Math. Statist.* 23 (1952) 277–281.

[8] P. DOUKHAN, P. MASSART and E. RIO, Invariance principles for absolutely regular empirical processes, *Ann. Inst. Henri Poincaré* 31 (2) (1995) 393–427.

[9] R.M. DUDLEY, Central limit theorems for empirical measures, *Ann. Probab.* 6 (6) (1978) 899–929.

[10] A. DVORETZKY, J.C. KIEFER and J. WOLFOWITZ, Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *Ann. Math. Statist.* 33 (1956) 642–669.

[11] A. FÖLDES and L. REJTŐ, A LIL type result for the product limit estimator, *Z. Wahrscheinlichkeitstheor. Verw. Geb.* 56 (1981) 75–86.

[12] R.D. GILL, Large sample behavior of the product limit estimator on the whole line, *Ann. Statist.* 11 (1983) 49–58.

[13] R.D. GILL, Glivenko–Cantelli for Kaplan–Meier, *Math. Methods of Statistics* 3 (1) (1994) 76–87.

[14] R.D. GILL, Lectures on survival analysis, in: *Lectures on Probability Theory, École d'Été de Probabilités de Saint-Flour XXII-1992*, Lecture Notes in Mathematics, Vol. 1581, Springer, Berlin, 1994, pp. 115–241.

[15] R.D. GILL and S. JOHANSEN, A survey of product-integration with a view towards applications in survival analysis, *Ann. Statist.* 6 (1990) 1501–1555.

[16] M.G. GU and T.L. LAI, Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation, *Ann. Probab.* 18 (1) (1990) 160–189.

[17] H.J. HALL and J.A. WELLNER, Confidence bands for a survival curve from censored data, *Biometrika* 67 (1980) 133–143.

[18] E.L. KAPLAN and P. MEIER, Non-parametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* 53 (1958) 562–563, 897–919.

[19] P. MASSART, The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality, *Ann. Probab.* 18 (3) (1990) 1269–1283.

[20] M. OSSIANDER, A central limit theorem under metric entropy with $\mathbb{L}^2$ bracketing, *Ann. Probab.* 15 (1987) 897–919.

[21] G.R. SHORACK and J.A. WELLNER, *Empirical Process with Applications to Statistics*, Wiley, New York, 1986.

[22] W. STUTE, Strong and weak representations of cumulative hazard function and Kaplan–Meier estimators on increasing sets, *J. Statist. Planning and Inference* 42 (1994) 315–329.

[23] W. STUTE and J.L. WANG, The strong law under random censorship, *Ann. Statist.* 21 (3) (1993) 1591–1607.

[24] S. VAN DE GEER, Hellinger consistency of certain nonparametric maximum likelihood estimators, *Ann. Statist.* 21 (1) (1993) 14–44.

[25] M.J. VAN DER LAAN, Efficient and inefficient estimation in semiparametric models, Ph.D. Thesis, Department of Mathematics, University of Utrecht, the Netherlands, 1993.

[26] M.J. VAN DER LAAN, Proving efficiency of NPMLE and identities, Technical Report 44, Group in Biostatistics, University of California, Berkeley, CA 94720, 1994.

[27] J.G. WANG, A note on the uniform consistency of the Kaplan–Meier estimator, *Ann. Statist.* 15 (1987) 1313–1316.