The Annals of Probability
1999, Vol. 27, No. 4, 2042–2067

LAWS OF THE ITERATED LOGARITHM FOR CENSORED DATA

BY EVARIST GINÉ[1] AND ARMELLE GUILLOU[2]

University of Connecticut and Université Paris VI

First- and second-order laws of the iterated logarithm are obtained for both the Nelson–Aalen and the Kaplan–Meier estimators in the random censorship model, uniform up to a large order statistic of the censored data. The rates for the first-order processes are exact except for constants. The LIL for the second-order processes (where one subtracts a linear, empirical process, term from the difference between the original process and the estimator), uniform over fixed intervals, is also proved. Somewhat surprisingly, there is a certain degree of proof unification for fixed and variable intervals in the second-order results for the Nelson–Aalen estimator. No assumptions are made on the distribution of the censoring variables and only continuity of the distribution function of the original variables is assumed for the results on the Kaplan–Meier estimator.

1. Introduction. Let X, X_i, $i \in \mathbb{N}$, be independent and identically distributed (i.i.d.) random variables with common distribution function (df) F and let $\Lambda(x)$ be its cumulative hazard function. Let Y, Y_i, $i \in \mathbb{N}$, be a second i.i.d. sequence, independent of the first, and let $Z = X \wedge Y$, $\delta = \mathbb{1}_{X \le Y}$, $Z_i = X_i \wedge Y_i$, $\delta_i = \mathbb{1}_{X_i \le Y_i}$, $i \in \mathbb{N}$. We denote by H the df of Z and by $\tau_H = \inf\{x: H(x) = 1\}$, the supremum of the support of H. Let $\Lambda_n(x)$, $-\infty < x < \tau_H$, denote the Nelson–Aalen estimator of $\Lambda(x)$, which is in terms of (Z_i, δ_i), $i = 1, \dots, n$, [Nelson (1972) and Aalen (1976)] and let $\hat{F}_n(x)$, $-\infty < x < \tau_H$, be the Kaplan–Meier (1958) product limit estimator of $F(x)$, also in terms of (Z_i, δ_i). (See Section 2 for definitions.)

The accuracy of the approximation of Λ by Λ_n uniformly in $(-\infty, T]$ or in $(-\infty, T_N]$ and that of F by $\hat{F}_n$ have been widely studied as they have obvious statistical interest. We refer to the interesting articles of Csörgő (1996) and Stute (1994) for detailed accounts on the history of this subject. In particular, these two authors study the a.s. asymptotic behavior of $\Lambda_n(x) - \Lambda(x)$ and $(\hat{F}_n(x) - F(x))/(1 - F(x))$ uniformly over the data driven intervals $(-\infty, Z_{n(1-\varepsilon_n), n}]$, where $Z_{j,n}$ is the jth order statistic of $Z_1, \dots, Z_n$ and $\{\varepsilon_n\}$ is a nonincreasing sequence such that $n\varepsilon_n \ge \log n$ and $n\varepsilon_n$ is a positive integer. The idea of replacing a fixed end point T by $Z_{n(1-\varepsilon_n), n}$ comes from Stute (1994), and Csörgő (1996) contains an extensive study of the subject, in

Received June 1999.
[1] Supported in part by NSF Grant DMS-96-25457.
[2] Research carried out at the Departments of Mathematics and Statistics, University of Connecticut, Storrs.
AMS 1991 *subject classifications*. Primary 62G05, 60F15; secondary 62G30, 62G20.
Key words and phrases. Random censorship, law of the iterated logarithm, U-statistics, U-processes, cumulative hazard function, product limit estimator.

particular refining Stute's results. We refrain from adding to these authors' comments on the statistical relevance of this type of results.

Csörgő (1996) proved, among many other results, that, assuming F continuous,

$$\sup_{x \le Z_{n(1-\varepsilon_n), n}} \left| \frac{\hat{F}_n(x) - F(x)}{1 - F(x)} \right| = O_P\left(\frac{1}{\sqrt{n\varepsilon_n}} \right)$$

for all $n\varepsilon_n \in \{1, 2, \dots, n-1\}$ and that

$$\sup_{x \le Z_{n(1-\varepsilon_n), n}} \left| \frac{\hat{F}_n(x) - F(x)}{1 - F(x)} \right| = O\left(\sqrt{\frac{\log n}{n\varepsilon_n}} \right) \quad \text{a.s.}$$

if $n\varepsilon_n \ge \log n$; Csörgő also showed that the first result is best possible and conjectured that the second should have $\log n$ replaced by $\log\log n$. Our first goal in this article is to prove this conjecture. In Theorem 7 and Remark 3 below we show that, assuming F continuous, and $n\varepsilon_n \ge 9\log n$, $n\varepsilon_n \in \mathbb{N}$,

$$(1.1) \qquad \lim_{n\to\infty} \sqrt{\frac{n\varepsilon_{2n}}{d_n \log\log n}} \sup_{x \le Z_{n(1-\varepsilon_n), n}} \left| \frac{\hat{F}_n(x) - F(x)}{1 - F(x)} \right| = 0 \quad \text{a.s.},$$

where $\{d_n\}$ is any nondecreasing sequence such that $\sum 1/(kd_{2^k} \log k) < \infty$ [e.g., $d_n = (\log\log\log n)^{1+\delta}$] and that, if we further assume $n\varepsilon_{2n} \ge Cd_n \log n$ for some $C > 0$ and some d_n as stated, then

$$(1.2) \qquad \limsup_{n\to\infty} \sqrt{\frac{n\varepsilon_{2n}}{\log\log n}} \sup_{x \le Z_{n(1-\varepsilon_n), n}} \left| \frac{\hat{F}_n(x) - F(x)}{1 - F(x)} \right| < \infty \quad \text{a.s.}$$

(Of course, this lim sup is a constant by the 0–1 law.) The same rates apply to $\sup_{x \le Z_{n(1-\varepsilon_n), n}} |\Lambda_n(x) - \Lambda(x)|$ (Theorem 6 and Remark 3 below). The rate (1.2) is best possible [Claim in Csörgő (1996), page 2749]. Our theorems are stated in terms of variable nonrandom end points T_n because, as observed by Stute and by Csörgő (see also Remark 3 below), the random case reduces almost trivially to this case. There has been a considerable amount of work on the LIL for $(\hat{F}_n - F)/(1 - F)$ uniform over fixed intervals $(-\infty, T]$, $T < \tau_H$, and we refer to Csörgő and Horváth (1983) for the final result; except for constants, these results are recovered by (1.2).

A second, related, goal here is that of obtaining "second-order laws of the iterated logarithm" for the Nelson–Aalen and the Kaplan–Meier estimators, uniform over fixed as well as variable intervals. To describe these results, we recall that the statistic $\Lambda_n(x)$, which is a symmetric statistic in the variables $(X_1, Y_1), \dots, (X_n, Y_n)$, admits a von Mises type development of the form

$$(1.3) \quad \Lambda_n(x) = \Lambda(x) + L_n(x) + Q_n(x) + D_n(x), \qquad -\infty < x < \tau_H,$$

where $\{L_n(x)\colon -\infty < x < \tau_H\}$ is a centered empirical process (the linear term), $\{Q_n(x)\colon -\infty < x < \tau_H\}$ is a canonical V-process of order 2 (the quadratic term), and $D_n(x)$ is a remainder [Stute (1994)], and that this development induces a similar one for the product limit estimator $\hat{F}_n(x)$ via a decomposition of Breslow and Crowley (1974). The term "second-order results" refers to result on the behavior of $\Lambda_n - \Lambda - L_n$ and $(\hat{F}_n - F)/(1-F) - L_n$. Our results of this kind, which are also optimal up to multiplicative constants, asserts that

$$\limsup_n \frac{n(1-H(T-))}{\log\log n} \sup_{x\le T} |\Lambda_n(x) - \Lambda(x) - L_n(x)| \le C_1 \quad \text{a.s.} \tag{1.4}$$

and that, for F continuous,

$$\limsup_n \frac{n(1-H(T-))}{\log\log n} \sup_{x\le T} \left| \frac{\hat{F}_n(x) - F(x)}{1-F(x)} - L_n(x) \right| \le C_2 \quad \text{a.s.} \tag{1.5}$$

for all $T < \tau_H$, where C_i are finite universal constants (respectively, Theorems 2 and 3 below). Concerning the uniformity of the second-order processes over the random intervals $(-\infty, Z_{n(1-\varepsilon_n),n}]$, we recover Csörgő's (1996) rates for the Nelson–Aalen estimator (Theorem 4 below) and improve his rates for the product limit estimator (Theorem 8). It is perhaps interesting to mention here that our proofs of the second-order results for the Nelson–Aalen estimator for fixed intervals and for variable intervals are basically the same; actually, in a sense, the result for variable intervals is a corollary from the one for fixed intervals. This makes us believe that the rates obtained for the second-order processes over variable intervals are not best possible (although they are good enough to imply sharp results for the first-order processes).

In order to obtain the first-order results, we combine the second-order results for the Nelson–Aalen estimator with a law of the iterated logarithm for the linear term,

$$\sup_{x\le Z_{n(1-\varepsilon_n),n}} |L_n(x)| \tag{1.6}$$

(Theorem 5). Perhaps the main technical improvement with respect to previous work is contained in our proof of this result, which relies on recent important theorems of Montgomery-Smith (1993) and Talagrand (1996) but is, otherwise, quite standard.

The strategy of proof of the second-order results is also relatively new and it closely follows the method of Stute (1993) for truncated data as modified by Arcones and Giné (1995). We decompose $D_n(x)$ from (1.3) as the product of the empirical quotient,

$$D_{n,1}(x) = \begin{cases} \displaystyle\max_{i\le n:\, Z_i\le x} \frac{1-H(Z_i-)}{1-H_n(Z_i-)}, & \text{if } \{i\le n: Z_i\le x\} \ne \emptyset, \\ 0, & \text{otherwise,} \end{cases} \tag{1.7}$$

and an almost degenerate V-process of order 3, say $D_{n,2}(x)$, and then we use Hoeffding's decomposition on the V-processes $Q_n(x)$ and $D_{n,2}(x)$ and apply the LIL for canonical U-statistics and U-processes [Arcones and Giné (1995)] to these components.

In order to show that the sequence $\{D_{n,1}(T_n)\}$ is a.s. bounded for very general T_n, we use an exponential bound from Shorack and Wellner (1986) based on Doob's maximal inequality. The bounded LIL for U-statistics and processes in Arcones and Giné (1995) is in terms of bounds for moments of the suprema of the normalized sums [an idea originated in Pisier (1975) and this allows for a unified treatment of both fixed and variable T. In the end, the proofs reduce to estimating second moments of appropriate kernels, and for this we follow Stute (1994).

Section 2 collects preliminary material, Section 3 is devoted to the LIL uniform over fixed intervals and variable intervals are treated in Section 4.

2. Preparatory material. We begin with some additional notation from Stute (1994): we set $\tilde{H}(x) = \Pr\{Z \le x, \delta = 1\}$, $-\infty < x < \tau_H$, and define H_n and $\tilde{H}_n$ to be the empirical counterparts of H and $\tilde{H}$, respectively, that is,

$$H_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{Z_i \le x} \quad \text{and} \quad \tilde{H}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{Z_i \le x,\, \delta_i = 1\}}, \qquad n \in \mathbb{N},$$

for $-\infty < x \le \tau_H$. The obvious facts that $d\tilde{H} \le dH$ and $d\tilde{H}_n \le dH_n$ will be used without further mention. We should recall that, with this notation and the notation set up in the Introduction,

$$\Lambda(x) = \int_{-\infty}^{x} \frac{d\tilde{H}}{1 - H_-}, \quad \Lambda_n(x) = \int_{-\infty}^{x} \frac{d\tilde{H}_n}{1 - H_{n-}}$$

and

$$1 - \hat{F}_n(x) = \prod_{j=1}^{n}\left[1 + \frac{\delta_{j,n}\mathbb{I}_{Z_{j,n} \le x}}{n - j + 1}\right]$$

for all $x \in (-\infty, \tau_H)$, where $\delta_{j,n} = \delta_k$ iff $Z_{j,n} = Z_k$ and H_-, H_{n-} are the left continuous versions of H and H_n, respectively.

We also require the following notation from U-statistics. The symmetrization sh of a function $h\colon S^r \mapsto \mathbb{R}$ where S is any set is defined as

$$sh(x_1, \dots, x_r) = \frac{1}{r!}\sum h(x_{\tau_1}, \dots, x_{\tau_r})$$

with the sum extended over all the permutations τ of $1, \dots, r$. Given a probability measure P on measurable space $(S, \mathscr{S})$ and a P^r-integrable function of r variables symmetric in its entries, $h(x_1, \dots, x_r)$ on S^r, its Hoeffding projections $\pi_m h$, $1 \le m \le r$, are defined as

$$\pi_m h(x_1, \dots, x_m) = (P - \delta_{x_1}) \times \cdots \times (P - \delta_{x_m}) \times P^{r-m} h$$

and $\pi_0 h = P^r h$, where we are using functional notation for integrals and δ_x is unit mass at x [then, e.g., for functions of two variables $h(x, y)$, we have $\pi_2 h(x, y) = h(x, y) - \int h(x, y)\, dP(x) - \int h(x, y)\, dP(y) + \int\int h(x, y)\, dP(x)\, dP(y)$]; $\pi_m h$ is a canonical or totally centered kernel, that is, $\int \pi_m h(x_1, \dots, x_m)\, dP(x_i) = 0$ for all $1 \le i \le m$. Analogous definitions can also be made for nonsymmetric functions: π_m is defined exactly in the same way for $m = 0$ and $m = r$, but there are more than one $\pi_m h$ if $0 < m < r$ and the definitions in this case become unnecessarily cumbersome.

As Stute (1994) observes,

$$\Lambda_n(x) - \Lambda(x) = L_n(x) + Q_n(x) + D_n(x), \qquad -\infty < x < \tau_H, \tag{2.1}$$

where

$$L_n(x) = \int_{-\infty}^{x} \frac{d\big(\tilde{H}_n(y) - \tilde{H}(y)\big)}{1 - H(y-)} + \int_{-\infty}^{x} \frac{H_n(y-) - H(y-)}{(1 - H(y-))^2}\, d\tilde{H}(y) \tag{2.2}$$

is the linear term in the von Mises decomposition of Λ_n,

$$\begin{aligned} Q_n(x) = {} & -\frac{1}{n^2} \sum_{1 \le i \ne j \le n} (\pi_2 h_x)\big((X_i, Y_i), (X_j, Y_j)\big) \\ & - \frac{1}{n} \int_{-\infty}^{x} \frac{d\tilde{H}_n(y)}{(1 - H(y-))^2} + \frac{1}{n} \int_{-\infty}^{x} \frac{d\tilde{H}_n(y)}{1 - H(y-)} \\ & - \frac{1}{n} \int_{-\infty}^{x} \frac{H_n(y-) - H(y-)}{(1 - H(y-))^2}\, d\tilde{H}(y) \\ := {} & -\frac{R_n(x)}{n^2} - R_{n,1}(x) + R_{n,3}(x) - R_{n,4}(x) \end{aligned} \tag{2.3}$$

is the quadratic term, with

$$h_x(\mathbf{x}_1, \mathbf{x}_2) = \mathbb{1}_{\{x_2 \wedge y_2 \ge x_1,\, x_1 \le y_1,\, x_1 \le x\}} / (1 - H(x_1 -))^2$$

(or its symmetrization; it does not matter) and

$$D_n(x) = \int_{-\infty}^{x} \frac{(H_n(y-) - H(y-))^2}{(1 - H(y-))^2 (1 - H_n(y-))}\, d\tilde{H}_n(y) \tag{2.4}$$

is the remainder term. [There is no term $R_{n,2}$ in (2.3) in order to conform as much as possible with Csörgő's (1996) notation.] Partly inspired by Stute (1993), we bound D_n as follows:

$$\begin{aligned} \sup_{x \le T} |D_n(x)| = D_n(T) \le D_{n,1}(T) & \int_{-\infty}^{T} \frac{(H_n(y-) - H(y-))^2}{(1 - H(y-))^3} \\ & \times dH_n(y) \\ := D_{n,1}(T) & D_{n,2}(T), \end{aligned} \tag{2.5}$$

where $D_{n,1}$ is as defined by (1.7). Now $D_{n,2}(T)$ is a V-statistic of degree 3 which is degenerate in two of the three coordinates. In order to analyze $D_{n,2}(T)$, it is natural (although somewhat tedious) to decompose it into its

U-statistic components, and then further decompose these, via Hoeffding's decomposition, into canonical U-statistics of different orders. If we do this, we obtain

$$
\begin{aligned}
D_{n,2}(T) &= \frac{1}{n^3} \sum_{1\le i\ne j\ne k\le n} \big(\mathbb{1}_{Z_i<Z_k} - H(Z_k-)\big)\big(\mathbb{1}_{Z_j<Z_k} - H(Z_k-)\big) \\
&\qquad\qquad \times \frac{\mathbb{1}_{(Z_k\le T)}}{(1-H(Z_k-))^3} \\
&\quad + \frac{1}{n^3} \sum_{1\le i\ne k\le n} \big(\mathbb{1}_{Z_i<Z_k} - H(Z_k-)\big)^2 \frac{\mathbb{1}_{(Z_k\le T)}}{(1-H(Z_k-))^3} \\
&\quad + \frac{2}{n^3} \sum_{1\le i\ne j\le n} \big(\mathbb{1}_{Z_i<Z_j} - H(Z_j-)\big)\big(-H(Z_j-)\big) \\
&\qquad\qquad \times \frac{\mathbb{1}_{(Z_j\le T)}}{\big(1-H(Z_j-)\big)^3} \\
&\quad + \frac{1}{n^3} \sum_{1\le i\le n} H^2(Z_i-) \frac{\mathbb{1}_{(Z_i\le T)}}{(1-H(Z_i-))^3} \\
&= \frac{1}{n^3} \sum_{1\le i\ne j\ne k\le n} (\pi_3 f)(Z_i, Z_j, Z_k) + \frac{1}{n^2} \\
&\quad \times \sum_{1\le i\ne j\le n} \int_{-\infty}^{T} \frac{\big(\mathbb{1}_{Z_i<y} - H(y-)\big)\big(\mathbb{1}_{Z_j<y} - H(y-)\big)}{(1-H(y-))^3}\, dH(y) \\
&\qquad + \frac{1}{n^3} \sum_{1\le i\ne k\le n} (\pi_2 g)(Z_i, Z_k) \\
&\qquad + \frac{1}{n^2} \sum_{1\le i\le n} \Bigg[\int_{-\infty}^{T} \frac{\big(\mathbb{1}_{Z_i<y} - H(y-)\big)^2}{(1-H(y-))^3}\, dH(y) \\
&\qquad - \mathbb{E}\bigg(\int_{-\infty}^{T} \frac{\big(\mathbb{1}_{Z_i<y} - H(y-)\big)^2}{(1-H(y-))^3}\, dH(y)\bigg)\Bigg] \\
&\qquad + \frac{1}{n^2} \sum_{1\le k\le n} \Bigg[\frac{H(Z_k-)\,\mathbb{1}_{Z_k<T}}{(1-H(Z_k-))^2} \\
&\qquad - \int_{-\infty}^{T} \frac{H(y-)}{(1-H(y-))^2}\, dH(y)\Bigg]
\end{aligned} \tag{2.6}
$$

$$
\begin{aligned}
&+\frac{1}{n}\int_{-\infty}^{T}\frac{H(y-)}{(1-H(y-))^{2}}\,dH(y)\\
&+\frac{2}{n^{3}}\sum_{1\le i\ne j\le n}(\pi_{2}h)(Z_i,Z_j)\\
&+\frac{2}{n^{2}}\sum_{1\le i\le n}\int_{-\infty}^{T}\frac{\big(\mathbb{1}_{Z_i<y}-H(y-)\big)(-H(y-))}{(1-H(y-))^{3}}\,dH(y)\\
&+\frac{1}{n^{3}}\sum_{1\le i\le n}\Bigg[\frac{H^{2}(Z_i-)\mathbb{1}_{Z_i<T}}{(1-H(Z_i-))^{3}}\\
&\qquad-\int_{-\infty}^{T}\frac{H^{2}(y-)}{(1-H(y-))^{3}}\,dH(y)\Bigg]\\
&+\frac{1}{n^{2}}\int_{-\infty}^{T}\frac{H^{2}(y-)}{(1-H(y-))^{3}}\,dH(y)\\
:=&\sum_{r=1}^{10}D_{n,2,r}(T),
\end{aligned}
$$

where

$$
\begin{aligned}
f(z_1,z_2,z_3)&:=\big(\mathbb{1}_{z_1<z_3}-H(z_3-)\big)\big(\mathbb{1}_{z_2<z_3}-H(z_3-)\big)\\
&\qquad\times\frac{\mathbb{1}_{z_3\le T}}{(1-H(z_3-))^{3}},\\
g(z_1,z_2)&:=\big(\mathbb{1}_{z_1<z_2}-H(z_2-)\big)^{2}\frac{\mathbb{1}_{z_2\le T}}{(1-H(z_2-))^{3}},\\
h(z_1,z_2)&:=\big(\mathbb{1}_{z_1<z_2}-H(z_2-)\big)(-H(z_2-))\\
&\qquad\times\frac{\mathbb{1}_{z_2\le T}}{(1-H(z_2-))^{3}}.
\end{aligned}
\tag{2.7}
$$

The terms R_n/n^2 and $R_{n,i}$, $i=1,3,4$, are all U-processes of different degrees (1 and 2), $D_{n,2,i}$, $i\ne 6,10$, are canonical U-statistics of different degrees (1, 2, 3) and $D_{n,2,6}$ and $D_{n,2,10}$ are constants. We will handle all the U-statistics and U-processes in the same way, by applying to each of them a version of the LIL for degenerate U-statistics and U-processes consisting of maximal inequalities for moments.

Regarding $D_{n,1}$ in (1.7) [and (2.5)], we will apply to it the following exponential inequality from Shorack and Wellner [(1986), pages 416 and 417, (a) and (6), implemented in analogy with Stute (1993), page 151].

LEMMA 1. *For all $\lambda \geq 1$, we have*

$$\Pr\{D_{n,1}(T) \geq \lambda\} \leq \exp\left(n(1 - H(T-))\left[\frac{\log \lambda}{\lambda} - 1 + \frac{1}{\lambda}\right]\right). \tag{2.8}$$

The U-process and U-statistic maximal inequalities will be in terms of the variances of the suprema of the respective kernels over x, and our estimates for these variances will be based on the following observation of Stute (1994).

LEMMA 2. *For any distribution function L and real number $r > 1$, we have*

$$\int_{-\infty}^{T} \frac{dL(y)}{(1 - L(y-))^{r}} \leq \frac{r}{(r-1)(1 - L(T-))^{r-1}}. \tag{2.9}$$

Finally, in order to describe the maximal inequalities in question it is convenient to recall two definitions from empirical process theory. Given a metric or pseudometric space (T, d), and $\varepsilon > 0$, the ε-covering number of (T, d), $N(T, d, \varepsilon)$ is defined as the minimal number of open d-balls of radius ε and centers in T required to cover T. A class of measurable functions $\mathscr{H}$ on a measure space $(S, \mathscr{S})$ is *a Vapnik–Červonenkis* (*VC*) class of functions with respect to the envelope H if there exists a function H measurable and everywhere finite with $|h| \leq H$ for all $h \in \mathscr{H}$ and numbers A and v finite, such that

$$N\big(\mathscr{H}, \|\cdot\|_{L_2(P)}, \varepsilon\|H\|_{L^2(P)}\big) \leq \left(\frac{A}{\varepsilon}\right)^{v}$$

for all $\varepsilon \in (0,1)$ and for all probability measures P on $(S, \mathscr{S})$ for which $\int H^2\, dP < \infty$. This definition is similar to Nolan and Polard's (1987) definition of Euclidean classes. We also say that the class $\mathscr{H}$ is *measurable* if it can be parametrized by a complete separable metric space Θ and the map $(\theta, x) \mapsto h_\theta(x)$ is jointly measurable.

The classes of functions we will use have a very simple structure and they will all be obviously measurable. To show that they are VC it will suffice to apply the next lemma.

LEMMA 3. (a) *If $\mathscr{H}$ is finite, then $\mathscr{H}$ is VC with respect to* $\max\{|h|: h \in \mathscr{H}\}$.

(b) *If $\mathscr{H} = \{h_x: x \in J\}$ where J is a subset of $\mathbb{R}$ and $0 \leq h_x(s) \leq h_y(s)$ for all $x < y$, $x, y \in J$ and $s \in S$, then $\mathscr{H}$ is VC for $H = \sup\{|h|: h \in \mathscr{H}\}$.*

(c) *If $\mathscr{H}_1$ and $\mathscr{H}_2$ are VC, respectively, for H_1 and H_2, then $\{h_1 + h_2: h_i \in \mathscr{H}_i,\ i = 1, 2\}$ and $\{h_1 - h_2: h_i \in \mathscr{H}_i,\ i = 1, 2\}$ are VC with respect to $(H_1^2 + H_2^2)^{1/2}$.*

(d) *If $\mathscr{H}$ is a class of functions on $(S^r, \mathscr{S}^r)$ with respect to an envelope H which is P^r-square integrable, P being a probability measure on $(S, \mathscr{S})$, and if π_m denotes the mth Hoeffding projection with respect to P, then the class*

$\{\pi_m sh\colon h \in \mathscr{H}\}$ is VC with respect to an envelope K such that $\|K\|_{L^2(P^m)} \le c_r \|H\|_{L^2(P^r)}$, $1 \le m \le r$, where c_r is a constant that depends on r only.

PROOF. (a) is trivial, (c) and (d) are essentially Corollaries 17 and 21 in Nolan and Pollard (1987) and (b) is well known but we indicate its proof. Given P such that $\int H^2\,dP < \infty$ and $\varepsilon \in (0,1)$ (we assume w.l.o.g. that ε^{-2} is an integer), let $I_k = [(k-1)\varepsilon^2, k\varepsilon^2]$, $k = 1,\dots,\varepsilon^{-2}$, and let us pick up, for each k for which it is possible, a number $x_k \in J$ such that $\int (h_{x_k}/H)^2\,dP \in I_k$. Then, for every x there is k such that $\int (h_x/H)^2\,dP \in I_k$. For any given x let k be such a number; then the fact that either $h_x \le h_{x_k}$ or $h_{x_k} \le h_x$ and both are nonnegative, implies

$$\int \left(\frac{h_{x_k} - h_x}{H}\right)^2 dP \le \left| \int \left(\frac{h_{x_k}}{H}\right)^2 dP - \int \left(\frac{h_x}{H}\right)^2 dP \right| \le \varepsilon^2,$$

showing that $N(\mathscr{H}, \|\cdot\|_{L^2(P)}, \varepsilon \|H\|_{L^2(P)}) \le 2\varepsilon^{-2}$. $\square$

The following result comes from Pisier (1975) for $m = 1$ and from Arcones and Giné (1995) for $m > 1$. We use the notation of Arcones and Giné, except that I_n^m, necessarily for $n \ge m$, stands here for the set of all vectors $(i_1,\dots,i_m) \subset \{1,\dots,n\}^m$ such that $i_k \ne i_l$ if $k \ne l$. In particular, $\|\phi(h)\|_{\mathscr{H}} := \sup_{h\in\mathscr{H}} |\phi(h)|$ for any functional ϕ on $\mathscr{H}$.

LEMMA 4. *Let $(S, \mathscr{S}, P)$ be a probability space, X_i, $i \in \mathbb{N}$, i.i.d.(P), S-valued random variables, and let $\mathscr{H}$ be a measurable Vapnik–Červonenkis class of functions $h\colon S^r \to \mathbb{R}$, with respect to an envelope H. Assume $\mathbb{E}H^2 < \infty$. Then, for all $m \le r$:*

(a) *There exist constants C_p depending only on $\mathscr{H}$ and r such that*

$$\mathbb{E}\left[\sup_n \frac{1}{(n \log\log n)^{m/2}} \left\| \sum_{(i_1,\dots,i_m)\in I_n^m} \pi_m sh(X_{i_1},\dots,X_{i_m}) \right\|_{\mathscr{H}}\right]^p \le C_p (\mathbb{E}H^2)^{p/2}$$

for all $0 < p < 2$, and

(b) For all $N < \infty$,

$$\mathbb{E}\left[\max_{n\le N} \frac{1}{N^{m/2}} \left\| \sum_{(i_1,\dots,i_m)\in I_n^m} \pi_m sh(X_{i_1},\dots,X_{i_m}) \right\|_{\mathscr{H}}\right]^2$$

$$\le C\mathbb{E}\left[\frac{1}{N^{m/2}} \left\| \sum_{(i_1,\dots,i_m)\in I_N^m} \pi_m sh(X_{i_1},\dots,X_{i_m}) \right\|_{\mathscr{H}}\right]^2 \le C_2 \mathbb{E}H^2,$$

where C and C_2 are finite constants that depend only on $\mathscr{H}$ and r.

REMARK 1. (i) If H is bounded (which will be our case), then there is also a bound in (a) for $p > 2$, namely $C_p \mathbb{E}H^p$, and a slightly more complicated bound for $p = 2$ [this is due to Doob's maximal inequalities for tail probabilities (for $p < 2$) and for moments (for $p \geq 2$), used in the proof: see Arcones and Giné (1995), proof of Theorem 2.5]. (ii) Because the sums above are over parallellepipeds (I_n^m) rather than tetrahedra ($1 \leq i_1 < \cdots < i_m \leq n$), there is no need to symmetrize h and we can take any of the logical possibilities for $\pi_m h$. (iii) More important, we note that, under the hypotheses of Lemma 4,

$$\limsup_{n \to \infty} \frac{1}{(n \log\log n)^{m/2}} \left\| \sum_{(i_1, \dots, i_m) \in I_n^m} \pi_m h(X_{i_1}, \dots, X_{i_m}) \right\|_{\mathscr{H}} \leq C_1 \mathbb{E}(H^2)^{1/2} \quad \text{a.s.} \tag{2.10}$$

because, this lim sup being invariant under permutations of the variables $\{X_i\}$, it is a.s. a constant by the Hewitt–Savage zero–one law and, by Lemma 4, this constant cannot exceed the stated bound [actually, this bound is not best possible: a compact LIL is also satisfied, Arcones and Giné (1995), and the best possible bound is the supremum of the limit set].

Part (a) of Lemma 4 is Theorem 2.5 in Arcones and Giné (1995), and part (b) follows from decoupling [de la Pẽna and Montgomery-Smith (1995)], randomization, iterated Lévy's inequalities [Lemma 2.4, Arcones and Giné (1995)] and an entropy bound for measurable VC classes [Lemma 2.2 in Arcones and Giné (1995)]. This is implicit in the proof of Theorem 2.5 there. The proof of (b) is a subset of that of (a). Then (b) extends an inequality in Stute (1994) to general VC collections of kernels and to general m.

For the reader's convenience, this section ends with the statements of the above-mentioned results of Montgomery-Smith and of Talagrand.

LEMMA 5 [Montgomery-Smith (1993)]. *Let $(S, \mathscr{S})$ be a measurable space and let Z_i, $i \in \mathbb{N}$, be i.i.d. S-valued random variables. Let $\mathscr{F}$ be a countable class of measurable real functions on S. Then*,

$$\Pr\left\{ \max_{1 \leq k \leq n} \left\| \sum_{i=1}^{k} f(Z_i) \right\|_{\mathscr{F}} > t \right\} \leq 9 \Pr\left\{ \left\| \sum_{i=1}^{n} f(Z_i) \right\|_{\mathscr{F}} > \frac{t}{30} \right\} \tag{2.11}$$

for all $n \in \mathbb{N}$ and $t > 0$.

LEMMA 6 [Talagrand (1996)]. *Let Z_i be n independent random variables taking values in a measurable space $(S, \mathscr{S})$, $n \in \mathbb{N}$, and let $\mathscr{F}$ be a countable class of measurable real functions on S. Set $W = \sup_{f \in \mathscr{F}} \sum_{i=1}^n f(Z_i)$ and define $U := \sup_{f \in \mathscr{F}} \|f\|_\infty$ and $V := \mathbb{E}[\sup_{f \in \mathscr{F}} \sum_{i=1}^n f^2(Z_i)]$. Then, for each $t > 0$, we have*

$$\Pr\{|W - \mathbb{E}W| \geq t\} \leq K \exp\left(-\frac{1}{K} \frac{t}{U} \log\left(1 + \frac{tU}{V}\right)\right), \tag{2.12}$$

where $K > 0$ is a universal constant.

We are now prepared to prove the announced laws of the iterated logarithm.

3. Moment bounds and the LIL uniform on $(-\infty, T]$. Our main object here is to prove the following theorem (and its corollaries, which we also denote as theorems).

THEOREM 1. *For $T < \tau_H$, $0 < p < 2$, and universal constants C_p, $C < \infty$,*

$$\text{(a)}\quad \mathbb{E}\Bigg[\sup_n \Bigg\{\bigg(\frac{n}{\log\log n}(1-H(T-))\bigg) \wedge \bigg(\frac{n^{3/2}}{\sqrt{\log\log n}}(1-H(T-))^{3/2}\bigg)\Bigg\}\sup_{x\le T}|Q_n(x)|\Bigg]^p \le C_p,$$

$$\text{(b)}\quad \mathbb{E}\Bigg[\max_{n\le N} \Bigg\{\bigg(\frac{n^2}{N}(1-H(T-))\bigg) \wedge \bigg(\frac{n^2}{\sqrt{N}}(1-H(T-))^{3/2}\bigg)\Bigg\}\sup_{x\le T}|Q_n(x)|\Bigg]^2 \le C,$$

$$\text{(c)}\quad \mathbb{E}\Bigg[\sup_n \Bigg\{\bigg(\frac{n}{\log\log n}(1-H(T-))\bigg) \wedge \bigg(\frac{n^{3/2}}{(\log\log n)^{3/2}}(1-H(T-))^{3/2}\bigg) \wedge \bigg(\frac{n^2}{\log\log n}(1-H(T-))^2\bigg) \wedge \bigg(\frac{n^{5/2}}{(\log\log n)^{1/2}}(1-H(T-))^{5/2}\bigg)\Bigg\} D_{n,2}(T)\Bigg]^p \le C_p,$$

$$\text{(d)}\quad \mathbb{E}\Bigg[\max_{n\le N} \Bigg\{\bigg(\frac{n^2}{N}(1-H(T-))\bigg) \wedge \bigg(\frac{n^3}{N}(1-H(T-))^2\bigg) \wedge \bigg(\frac{n^3}{N^{3/2}}(1-H(T-))^{3/2}\bigg) \wedge \bigg(\frac{n^3}{\sqrt{n}}(1-H(T-))^{5/2}\bigg)\Bigg\} D_{n,2}(T)\Bigg]^2 \le C.$$

We note that the pth moments, $p > 2$, in (a) and (c) are also finite if we let the powers of $1 - H(T-)$ be larger [due to the bound $\mathbb{E}H^p$, for $p > 2$, instead of $(\mathbb{E}H^2)^{p/2}$, in Lemma 4].

PROOF. First we prove (a) and (b): we show that these bounds are satisfied by each of the terms in the decomposition (2.3) of Q_n. Then $R_{n,3}$ requires no consideration since

$$\sup_{x \le T} R_{n,3}(x) \le \frac{1}{n(1 - H(T-))} \int_{-\infty}^{T} d\tilde{H}_n \le \frac{1}{n(1 - H(T-))}.$$

Since $d\tilde{H}_n \le dH_n$, Lemma 2 gives

$$\begin{aligned}
\sup_{x \le T} R_{n,1}(x) &= \frac{1}{n} \int_{-\infty}^{T} \frac{d\tilde{H}_n(y)}{(1 - H(y-))^2} \le \frac{1}{n} \int_{-\infty}^{T} \frac{dH_n(y)}{(1 - H(y-))^2} \\
&= \frac{1}{n} \int_{-\infty}^{T} \frac{d(H_n(y) - H(y))}{(1 - H(y-))^2} + \frac{1}{n} \int_{-\infty}^{T} \frac{dH(y)}{(1 - H(y-))^2} \\
&\le \frac{1}{n^2} \sum_{i=1}^{n} \left(\frac{\mathbb{1}_{Z_i < T}}{(1 - H(Z_i -))^2} - \mathbb{E} \frac{\mathbb{1}_{Z < T}}{(1 - H(Z-))^2} \right) \\
&\quad + \frac{2}{n(1 - H(T-))}.
\end{aligned}$$

We can now apply Lemma 4 to the first sum (for $\mathscr{H}$ equal to a single function and $m = 1$); since by Lemma 2,

$$\mathbb{E}\left(\frac{\mathbb{1}_{Z < T}}{(1 - H(Z-))^2} \right)^2 = \int_{-\infty}^{T} \frac{dH(y)}{(1 - H(y-))^4} \le \frac{4}{3(1 - H(T-))^3},$$

Lemma 4 gives

$$\mathbb{E}\left[\sup_{n} \left\{ \left(\frac{n^{3/2}}{\sqrt{\log\log n}} (1 - H(T-))^{3/2} \right) \wedge \left(n(1 - H(T-)) \right) \right\} \sup_{x \le T} R_{n,1}(x) \right]^p$$
$$< C_p$$

for $p < 2$, and

$$\mathbb{E}\left[\max_{n \le N} \left\{ \left(\frac{n^2}{\sqrt{N}} (1 - H(T-))^{3/2} \right) \wedge \left(\frac{n^2}{N} (1 - H(T-)) \right) \right\} \sup_{x \le T} R_{n,1}(x) \right]^2 < C,$$

showing that the conclusions (a) and (b) of the theorem hold for the component $R_{n,1}$ of Q_n. Using Lemma 2 once more,

$$\sup_{x \le T} |R_{n,4}(x)| \le \frac{1}{n}\Big(\sup_{x \le T} |H_n(x-) - H(x-)|\Big)\frac{2}{1 - H(T-)},$$

and we can apply Lemma 4 to the kernels $\{\mathbb{1}_{Z<x}\colon x \le T\}$ with envelope $H \equiv 1$ and $m = 1$ by Lemma 3(b). In this case, the multiplicative constants in (a) and (b) for $R_{n,4}$ are, respectively,

$$\frac{n^{3/2}(1 - H(T-))}{\sqrt{\log\log n}} \quad \text{and} \quad \frac{n^2(1 - H(T-))}{N^{1/2}},$$

which are larger than those predicated for Q_n in the statement of the theorem because $1 - H(T-) < 1$. Finally, we consider R_n/n^2. Since the functions h_x are increasing in x, Lemma 3(b) shows that the class of kernels $\{h_x\colon -\infty < x \le T\}$ is VC with respect to the envelope h_T and then Lemma 3(d) implies that the class $\{\pi_2 sh_x\colon -\infty < x \le T\}$ is also VC with respect to an envelope whose second P^2-moment is dominated by a constant times the second P^2-moment of h_T. Since, moreover, the map $(x, \mathbf{x}_1, \mathbf{x}_2) \to \pi_2 sh_x(\mathbf{x}_1, \mathbf{x}_2)$ is jointly measurable, we can apply Lemma 4 to this class (see the second remark below Lemma 4 regarding the irrelevance of symmetrizing). Computations similar to those above [Stute (1994), page 324] yield

$$\mathbb{E} h_T^2 \le \frac{C}{(1 - H(T-))^2}$$

and, therefore, Lemma 4 for $m = r = 2$ gives

$$\mathbb{E}\left[\sup_n \frac{n}{\log\log n}(1 - H(T-)) \sup_{x \le T}\left|\frac{R_n(x)}{n^2}\right|\right]^p < C_p, \qquad 0 < p < 2$$

and

$$\mathbb{E}\left[\max_{n \le N} \frac{n^2}{N}(1 - H(T-)) \sup_{x \le T}\left|\frac{R_n(x)}{n^2}\right|\right]^2 < C.$$

Parts (a) and (b) of Theorem 1 are thus proved.

Next we prove parts (c) and (d) using the decomposition (2.6) for $D_{n,2}$. Note that

$$D_{n,2,6}(T) \le \frac{2}{n(1 - H(T-))} \quad \text{and} \quad D_{n,2,10}(T) \le \frac{3}{2n^2(1 - H(T-))^2}$$

by Lemma 2, both within the range of (c) and (d). For $D_{n,2,r}(T)$, $r \ne 6, 10$, we will apply Lemma 4 with $\mathscr{H}$ consisting of a single function and $m = r = 1, 2, 3$. We only need to evaluate the variances of the kernels involved; only the computation of one or two of these variances will be given in full size since they are all estimated by application of Lemma 2 and the only extra ingredient used in some of these estimates is the well-known generalized Minkowski inequality. Here they are, in order:

$$\mathbb{E} f^2(Z_1, Z_2, Z_3) \le \int_{-\infty}^{T} \frac{dH(y)}{(1 - H(y-))^4} \le \frac{4}{3(1 - H(T-))^3},$$

$$\mathbb{E}\left[\int_{-\infty}^{T} \frac{(\mathbb{1}_{Z_1 < y} - H(y-))(\mathbb{1}_{Z_2 < y} - H(y-))}{(1 - H(y-))^3}\, dH(y)\right]^2$$

$$\le \left[\int_{-\infty}^{T} \frac{\left(\mathbb{E}\left[(\mathbb{1}_{Z_1 < y} - H(y))^2 (\mathbb{1}_{Z_2 < y} - H(y-))^2\right]\right)^{1/2}}{(1 - H(y-))^3}\, dH(y)\right]^2$$

$$= \left[\int_{-\infty}^{T} \frac{dH(y)}{(1 - H(y-))^2}\right]^2 \le \frac{4}{(1 - H(T-))^2},$$

$$\mathbb{E} g^2(Z_1, Z_2) \le C \int_{-\infty}^{T} \frac{dH(y)}{(1 - H(y-))^5} \le \frac{\tilde{C}}{(1 - H(T-))^4},$$

$$\mathbb{E}\left[\int_{-\infty}^{T} \frac{(\mathbb{1}_{Z_1 < y} - H(y-))^2}{(1 - H(y-))^3}\, dH(y)\right]^2 \le \left(\int_{-\infty}^{T} \frac{dH(y)}{(1 - H(y-))^{5/2}}\right)^2$$

$$\le \frac{25}{9(1 - H(T-))^3},$$

$$\mathbb{E}\left(\frac{H(Z-)\, \mathbb{1}_{Z < T}}{(1 - H(Z-))^2}\right)^2 \le \frac{4}{3(1 - H(T-))^3},$$

$$\mathbb{E} h^2(Z_1, Z_2) \le \mathbb{E} \int_{-\infty}^{T} (\mathbb{1}_{Z_1 < y} - H(y-))^2 \frac{dH(y)}{(1 - H(y-))^6}$$

$$\le \frac{5}{4(1 - H(T-))^4},$$

$$\mathbb{E}\left(\int_{-\infty}^{T} \frac{|\mathbb{1}_{Z < y} - H(y-)|}{(1 - H(y-))^3}\, dH(y)\right)^2 \le \frac{25}{9(1 - H(T-))^3}$$

and

$$\mathbb{E}\left(\frac{\mathbb{1}_{Z < T}}{(1 - H(T-))^3}\right)^2 = \int_{-\infty}^{T} \frac{dH(y)}{(1 - H(y-))^6} \le \frac{6}{5(1 - H(T-))^5}.$$

Applying Lemma 4 for $\mathscr{H}$ consisting of a single function, and using these bounds for the second moments of the kernels, we obtain (c) and (d). $\square$

As a consequence of Theorem 1, here is the LIL for the remainder term after linearization of the Nelson–Aalen estimator, uniform over $(-\infty, T]$.

THEOREM 2. *There is a universal constant $C < \infty$ such that*

$$\limsup_n \frac{n(1 - H(T-))}{\log\log n} \sup_{x \le T} |\Lambda_n(x) - \Lambda(x) - L_n(x)| \le C \quad a.s. \tag{3.1}$$

for all $T < \tau_H$.

PROOF. If we take $\lambda = 3$ in (2.8), we obtain

$$\sum_n \Pr(D_{n,1}(T) \ge 3) \le \sum_n \exp(-Cn(1 - H(T-))) < \infty$$

for $C = 1 - ((1 + \log 3)/3) > 0$, so that

$$\limsup_n D_{n,1}(T) < 3 \quad \text{a.s.}$$

The Hewitt–Savage zero–one law and, respectively, (a) and (c) in Theorem 1, imply

$$\limsup_n \frac{n}{\log\log n}(1 - H(T-)) \sup_{x \le T} |Q_n(x)| \le C_1 \quad \text{a.s.}$$

and

$$\limsup_n \frac{n}{\log\log n}(1 - H(T-)) D_{n,2}(T) \le C_1 \quad \text{a.s.}$$

Now, the result follows because, by (2.1) and (2.5),

$$|\Lambda_n(x) - \Lambda(x) - L_n(x)| \le |Q_n(x)| + D_{n,1}(T) D_{n,2}(T), \qquad -\infty < x \le T.$$

$\square$

This theorem can in fact be proved in an easier way; however, once we have estimated the second moments of all the components, it seem worthwhile to obtain it as a consequence of the stronger statements (a) and (c) in Theorem 1, as well as the easy Lemma 1.

REMARK 2. (Optimality of the LIL in Theorem 2 and a compact LIL). Let $T < \tau_H$. Then, the class of functions $h_x(z) = \mathbb{I}_{z \ge x}/(1 - H(x-))$, $-\infty < x \le T$, is a uniformly bounded measurable VC class [by Lemma 3(b) and (c)] and therefore, for instance by the Vapnik–Červonenkis law of large numbers

[Vapnik–Červonenkis (1981) but also, e.g., by Lemma 4(a)], we have

$$\lim_{n\to\infty}\sup_{x\le T}\left|\frac{1-H_n(x-)}{1-H(x-)}-1\right|=0,\quad \text{a.s.},$$

hence, also

$$\lim_{n\to\infty}\sup_{x\le T}\left|\frac{1-H(x-)}{1-H_n(x-)}-1\right|=0\quad \text{a.s.} \tag{3.2}$$

Because of this we can replace $1-H_n(x-)$ by $1-H(x-)$ in the definition (2.4) of $D_n(x)$ without changing the pertinent lim sup. This observation and the bounded LIL in Lemma 4(a) allow us to conclude, by (2.1), (2.3), (2.4) and a variation on (2.6), that

$$\begin{aligned}
&\limsup_n \frac{n}{\log\log n}\sup_{x\le T}|\Lambda_n(x)-\Lambda(x)-L_n(x)|\\
&\quad=\limsup_n \frac{n}{\log\log n}\sup_{x\le T}\left|\frac{1}{n^2}\sum_{I_n^2}\left[-(\pi_2 h_x)\big((X_i,Y_i),(X_j,Y_j)\big)\right.\right.\\
&\qquad\qquad\left.\left.+\int_{-\infty}^{x}\frac{\big(\mathbb{1}_{Z_i<y}-H(y-)\big)\big(\mathbb{1}_{Z_j<y}-H(y-)\big)}{(1-H(y-))^3}\,d\tilde{H}(y)\right]\right|.
\end{aligned}$$

The process within the absolute value signs is a canonical (or degenerate) U-process over a measurable VC class of kernels (see, e.g., Lemma 3 above). Thus, we can apply to it the compact law of the iterated logarithm in Arcones and Giné [(1995), Theorem 4.7], and conclude that this lim sup is a.s. strictly positive as long as the cdf H is nondegenerate on $(-\infty, T]$. That is, there exists $c>0$ depending on X, Y, and T, such that

$$\limsup_n \frac{n}{\log\log n}\sup_{x\le T}|\Lambda_n(x)-\Lambda(x)-L_n(x)|=c\quad\text{a.s.},$$

showing that the LIL in Theorem 2 is optimal (in its dependence on n). It is easy to obtain an expression for c from the aforementioned compact LIL for U-processes.

The bounds (a)–(d) of Theorem 1 can also be used for $T=T_n$ variable, and (a), (c), may be better or worse than (b), (d), depending on the sequence T_n. We will illustrate this observation in the next section for the important case considered by Stute (1994) and Csörgő (1996). Now we show how Theorem 2 translates into an LIL for the product limit estimator $\hat{F}_n$ of F as follows.

THEOREM 3. *Assuming the distribution function F of X to be continuous, and with $L_n(x)$, $-\infty<x<\tau_H$, $n\in\mathbb{N}$, as in* (2.2), *there exists a universal constant C such that*

$$\limsup_n \frac{n(1-H(T-))}{\log\log n}\sup_{x\le T}\left|\frac{\hat{F}_n(x)-F(x)}{1-F(x)}-L_n(x)\right|\le C\quad a.s. \tag{3.3}$$

for all $T<\tau_H$.

PROOF. It is well known that $1 - F(x) = \exp(-\Lambda(x))$ if F is continuous. In this case, the classical expansion of Breslow and Crowley (1974), as described in Csörgő [(1996), pages 2769–2770], is as follows: for any real function h and for all $x < Z_{n,n}$,

$$\left|\frac{\hat{F}_n(x) - F(x)}{1 - F(x)} - h(x)\right| \leq \left|(\Lambda_n(x) - \Lambda(x)) - h(x)\right| + \left|R_{n,6}(x)\right|, \tag{3.4}$$

where

$$\begin{aligned} R_{n,6}(x) = {} & \left[\tfrac{1}{2}\left|\Lambda_n(x) - \Lambda(x)\right|^2 + \left|l_n(x)\right| \exp(\left|l_n(x)\right|)\right] \\ & \times \exp(\left|\Lambda_n(x) - \Lambda(x)\right|) \end{aligned} \tag{3.5}$$

and

$$l_n(x) = -\log\bigl(1 - \hat{F}_n(x)\bigr) - \Lambda_n(x). \tag{3.6}$$

If we take $h(x) = L_n(x)$ in (3.4), it follows from this expansion, by application of Theorem 2 to the first term at the right side of the obvious identity,

$$\Lambda_n(x) - \Lambda(x) = \left[\Lambda_n(x) - \Lambda(x) - L_n(x)\right] + L_n(x),$$

that we only have to obtain appropriate rates for each of the two sequences,

$$\sup_{x \leq T} \left|L_n(x)\right| \quad \text{and} \quad \sup_{x \leq T} \left|\log\bigl(1 - \hat{F}_n(x)\bigr) + \Lambda_n(x)\right|. \tag{3.7}$$

L_n is the sum of two centered empirical processes (canonical U-processes of degree 1) that can be estimated using Lemma 4, as in Theorem 1. Each of the two summands in the definition (2.2) of L_n can be estimated just as in Theorem 1, as follows. The class of functions $l_x(z) := \mathbb{I}_{z \leq x}/(1 - H(z-))$, $-\infty < x \leq T$, is measurable; it is VC with envelope l_T by Lemma 3(b), and $\mathbb{E} l_T^2 \leq 2/(1 - H(T-))$ by Lemma 2. Then, Lemma 4(a) for $m = 1$ gives

$$\begin{aligned} \mathbb{E}\left(\sup_n \sqrt{\frac{n(1 - H(T-))}{\log\log n}} \sup_{x \leq T} \left|\int_{-\infty}^{x} \frac{d\bigl(\tilde{H}_n(y) - \tilde{H}(y)\bigr)}{1 - H(y-)}\right|\right)^p & \\ \leq C_p, \qquad p < 2. & \end{aligned} \tag{3.8}$$

The second summand in (2.2) is a centered empirical process over the class of functions

$$k_x(z) = \int_{-\infty}^{x} \frac{\mathbb{I}_{z \geq y}}{(1 - H(y-))^2}\, d\tilde{H}(y), \qquad -\infty < x \leq T,$$

which is also measurable, and is VC by Lemma 3(b). Since

$$\begin{aligned} \mathbb{E}\left[\int_{-\infty}^{T} \frac{\mathbb{I}_{Z \geq y}}{(1 - H(y-))^2}\, dH(y)\right]^2 & \leq \left(\int_{-\infty}^{T} \frac{dH}{(1 - H_-)^{3/2}}\right)^2 \\ & \leq \frac{9}{1 - H(T-)} \end{aligned} \tag{3.9}$$

by generalized Minkowski and Lemma 2, it follows from Lemma 4(a) that

$$\mathbb{E}\left(\sup_n \sqrt{\frac{n(1-H(T-))}{\log\log n}} \times \sup_{x\le T}\left|\int_{-\infty}^{x} \frac{H_n(y-)-H(y-)}{(1-H(y-))^2}\,d\tilde{H}(y)\right|\right)^p \le C_p, \qquad p<2. \tag{3.10}$$

Using the bounds (3.8) and (3.10) in (2.2) results in a bound for L_n which, by the Hewitt–Savage zero–one law, implies the LIL for L_n uniform over $(-\infty, T]$,

$$\limsup_n \sqrt{\frac{n(1-H(T-))}{\log\log n}} \sup_{x\le T}|L_n(x)| \le \tilde{C} \quad \text{a.s.} \tag{3.11}$$

for some universal constant $\tilde{C}$.

For the second term in (3.7), $\sup_{x\le T}|\log(1-\hat{F}_n(x)) + \Lambda_n(x)|$, we use Lemma 1 in Breslow and Crowley (1974), which asserts that if $x \le Z_{n,n}$ then, with probability 1,

$$0 < -\log\big(1-\hat{F}_n(x)\big) - \Lambda_n(x) \le \frac{H_n(x-)}{n(1-H_n(x-))}. \tag{3.12}$$

Since T is eventually a.s. smaller than $Z_{n,n}$ we can apply inequality (3.12) and obtain

$$\begin{aligned}\sup_{x\le T}&\left|\log\big(1-\hat{F}_n(x)\big)+\Lambda_n(x)\right| \\ &\le \frac{1}{n(1-H(T-))}\,\frac{1-H(T-)}{1-H_n(T-)} \\ &= O\left(\frac{1}{n(1-H(T-))}\right) \quad \text{a.s.}\end{aligned} \tag{3.13}$$

by the law of large numbers (3.2).

Now the theorem follows by substituting the limits (3.1), (3.11) and (3.13) into equations (3.4)–(3.6) [note that the exponentials in these identities tend to 1, by (3.13), (3.11) and Theorem 2]. □

By directly using the expansion in Breslow and Crowley (1974) [instead of its consequence (3.4)] and Remark 2 above, it can be shown that the lim sup in (3.3) is a.s. a strictly positive constant, that is, that Theorem 3 is optimal.

4. Logarithmic laws uniform on $(-\infty, T_n]$ and applications to first-order LILs. As in the previous section we begin with a result for the Nelson–Aalen estimator. In what follows we denote the quantile function of Z by H^{-1}, that is, $H^{-1}(x) = \inf[z\colon H(z) \ge x]$ for $x \in (0,1)$, and recall that $H(H^{-1}(x)-) \le x \le H(H^{-1}(x))$.

THEOREM 4 [Csörgő (1996), Stute (1994)]. *If $T_n = H^{-1}(1 - \varepsilon_n/8)$ and $\{\varepsilon_n\}$ is a nonincreasing sequence of positive numbers satisfying*

$$n\varepsilon_n \geq 9\log n \tag{4.1}$$

for all $n > n_0$ and some $n_0 < \infty$, then

$$\lim_{n\to\infty} \frac{n\varepsilon_{2n}}{(c_n \log n)^{1/2}} \sup_{x \leq T_n} |\Lambda_n(x) - \Lambda(x) - L_n(x)| = 0 \quad a.s., \tag{4.2}$$

where $\{c_n\}$ is any nondecreasing slowly varying sequence of positive numbers such that $\Sigma 1/(kc_{2^k}) < \infty$ [such as $c_n = (\log\log n)^{1+\varepsilon}$, $c_n = (\log\log n)(\log\log\log n)^{1+\varepsilon}$, etc.].

PROOF. The definition of T_n implies that

$$1 - H(T_n -) \geq \frac{\varepsilon_n}{8}. \tag{4.3}$$

If $N = 2^{k+1}$ and $2^k < n \leq 2^{k+1}$, inequalities (4.1) and (4.3) imply that

$$\left(\frac{n^2}{N}(1 - H(T_n -))\right) \wedge \left(\frac{n^2}{\sqrt{N}}(1 - H(T_n -))^{3/2}\right) = \frac{n^2}{N}(1 - H(T_n -))$$

for all n large enough; since moreover T_n is nondecreasing (as ε_n is nonincreasing), Theorem 1(b) then gives

$$\mathbb{E}\left(\max_{2^k < n \leq 2^{k+1}} \frac{n^2(1 - H(T_{2^{k-1}} -))}{2^{k+1}} \sup_{x \leq T_n} |Q_n(x)|\right)^2 \leq C,$$

where we note that $n^2(1 - H(T_{2^{k+1}} -))/2^{k+1} \geq n\varepsilon_{2n}/16$ (for $n > 2^k$). Hence, if $\eta_k \to 0$ is such that $\Sigma 1/(\eta_k^2 kc_{2^k})$ converges, then, by Chebyshev,

$$\sum_k \Pr\left\{\max_{2^k < n \leq 2^{k+1}} n\varepsilon_{2n} \sup_{x \leq T_n} |Q_n(x)| > \eta_k \sqrt{kc_{2^k}}\right\} < \infty,$$

and therefore, by Borel–Cantelli,

$$\lim_{n\to\infty} \frac{n\varepsilon_{2n}}{\sqrt{c_n \log n}} \sup_{x \leq T_n} |Q_n(x)| = 0 \quad \text{a.s.}$$

Similarly, Theorem 1(d) implies

$$\lim_{n\to\infty} \frac{n\varepsilon_{2n}}{\sqrt{c_n \log n}} D_{n,2}(T_n) = 0 \quad \text{a.s.}$$

Finally, if we take λ so that $1 - (1/\lambda) - (\log \lambda/\lambda) > 8/9$ in Lemma 1, for instance, $\lambda = 50$, we have, by (4.3) and Lemma 1, that

$$\sum_n \Pr\{D_{n,1}(T_n) \geq 50\} \leq \sum_n \exp\left(-\frac{9}{8}\left(1 - \frac{1}{50} - \frac{\log 50}{50}\right)\log n\right) < \infty,$$

showing that $\limsup_{n\to\infty} D_{n,1}(T_n) < \infty$ a.s. The theorem now follows from (2.1) and (2.5). □

REMARK 3. The situation considered in the previous theorem is important because assumption (4.1) on ε_n implies that the variable end points $T_n = H^{-1}(1-\varepsilon_n/8)$ are eventually a.s. larger than $H_n^{-1}(1-\varepsilon_n)$, hence than $Z_{n(1-\varepsilon_n),\,n}$, so that then *one can choose* T_n *based on the observations* [Stute (1994) and Csörgő (1996)].

Here is a quick way to see this: if we set $T_n = H^{-1}(1-p\varepsilon_n)$ for some $0 < p < 1$ and let "Bin(r, q)" stand for a *binomial* (r, q) *random variable*, then the properties of quantile functions imply [Stute (1994), Lemma 2.4] that $\Pr\{H_n^{-1}(1-\varepsilon_n) > H^{-1}(1-p\varepsilon_n)\} \le \Pr\{\mathrm{Bin}(n, p\varepsilon_n) > n\varepsilon_n\}$, which is dominated by $(enp\varepsilon_n/n\varepsilon_n)^{n\varepsilon_n} = (ep)^{n\varepsilon_n}$ [Giné and Zinn (1984), Remark 4.7]; if $p = 1/8$ and $n\varepsilon_n \ge \log n$ then the series $\sum(ep)^{n\varepsilon_n}$ converges.

The proofs of Theorems 1, 2 and 4 provide a unified treatment for the cases of fixed and variable, even data driven, intervals, $(-\infty, T]$ and $(-\infty, Z_{n(1-\varepsilon_n),\,n}]$, and basically establish the Csörgő–Stute results on data driven intervals for the remainder term after linearization of the Nelson–Aalen estimator as a corollary of the result for fixed T. [In fact, if we wanted to obtain only Stute's (1994) result, which is weaker than Csörgő's (1996), then we could as well have used (a) and (c) in Theorem 1, since the extra $\log\log n$ does not change the rate in Stute's paper.] It can be argued that Theorem 4 (variable intervals) is somewhat more elementary than Theorem 2 (fixed intervals): the estimates in (b) and (d) of Theorem 1 (used for Theorem 4) are simpler than the estimates (a) and (c) there (used for Theorem 2). On the other hand, Theorem 2 is essentially best possible whereas we do not know if Theorem 4 is.

The way L_n is handled in the proof of Theorem 3 (fixed intervals for the Kaplan–Meier product limit estimator) is suboptimal in the sense that there are more precise (exponential) bounds for empirical processes than those used in that proof. This was also observed by Stute [(1994), Lemma 2.8] for the first summand in (2.2) and by Csörgő [(1996), proof of Proposition 4] for the second. Here we will derive a law of the iterated logarithm for $\sup_{x \le T_n}|L_n(x)|$, T_n as in Theorem 4, that will have several consequences, both for the second-order law of the iterated logarithm for the product limit estimator and the first-order law of the iterated logarithm, thereby improving some results in Csörgő (1996). This is achieved by combining Montgomery-Smith's maximal inequality (Lemma 5) with Talagrand's exponential inequality (Lemma 6).

THEOREM 5. *Let* $L_n(x)$, $x < \tau_H$, $n \in \mathbb{N}$, *be the linear term of* $\Lambda_n(x) - \Lambda(x)$, *as defined in equation* (2.2), *and let* ε_n *and* T_n, $n \in \mathbb{N}$, *be as in Theorem* 4. *Then*,

$$\limsup_{n\to\infty} \sqrt{\frac{n\varepsilon_{2n}}{\log\log n}} \sup_{x \le T_n} |L_n(x)| < \infty \quad a.s. \tag{4.4}$$

PROOF. We only give the details of the proof of (4.4) for the second summand in the definition (2.2) of L_n, since the proof for the first summand is completely analogous. To this end we define

$$f_x(z) = \int_{-\infty}^{x} \frac{\mathbb{1}_{z \geq y}}{(1 - H(y-))^2}\, d\tilde{H}(y), \qquad x, z \leq T_n,$$

and set $\mathscr{G} = \mathscr{G}_n = \{f_x, -f_x \colon -\infty < x \leq T_n\}$. Then, we can rewrite the second term at the right of (2.2) as

$$\begin{aligned} W_n &:= \sup_{x \leq T_n} \left| \int_{-\infty}^{x} \frac{H_n(y-) - H(y-)}{(1 - H(y-))^2}\, d\tilde{H}(y) \right| \\ &= \frac{1}{n} \sup_{f \in \mathscr{G}_n} \sum_{i=1}^{n} (f(Z_i) - \mathbb{E} f(Z)) \qquad (4.5) \\ &= \frac{1}{n} \left\| \sum_{i=1}^{n} (f(Z_i) - \mathbb{E} f(Z)) \right\|_{\mathscr{G}_n}. \end{aligned}$$

Obviously, there is a countable set A in $\mathbb{R}$ such that W_n can be obtained as the sup of the absolute values of the same integrals over $x \in A \cap (-\infty, T_n]$ and, to be strict, we should define $\mathscr{G}_n$ as the set of functions f_x and $-f_x$ for $x \in A \cap (-\infty, T_n]$, so that $\mathscr{G}_n$ is countable and we can apply to it the theorems of Montgomery-Smith and Talagrand.

Montgomery-Smith's maximal inequality in Lemma 5 shows that, for $k \geq 4$ and all $u > 0$,

$$\begin{aligned} &\Pr\left\{ \max_{2^{k-1} < n \leq 2^k} \sqrt{\frac{n \varepsilon_{2n}}{\log\log n}}\, W_n > u \right\} \\ &\quad \leq \Pr\left\{ \max_{2^{k-1} < n \leq 2^k} \left\| \sum_{i=1}^{n} (f_x(Z_i) - \mathbb{E} f_x(Z)) \right\|_{\mathscr{G}_{2^k}} > u \sqrt{\frac{2^{k-1} \log\log 2^{k-1}}{\varepsilon_{2^k}}} \right\} \qquad (4.6) \\ &\quad \leq 9 \Pr\left\{ \left\| \sum_{i=1}^{2^k} (f_x(Z_i) - \mathbb{E} f_x(Z)) \right\|_{\mathscr{G}_{2^k}} > \frac{u}{60} \sqrt{\frac{2^k \log\log 2^k}{\varepsilon_{2^k}}} \right\}. \end{aligned}$$

To ease notation, we set $2^k = N$.

We now apply Talagrand's inequality from Lemma 6 with $\mathscr{F} = \{f - \mathbb{E} f(Z) \colon f \in \mathscr{G}_N\}$. For this, we estimate the quantities involved [in what follows $\mathbb{E} f$ should be understood as $\mathbb{E} f(Z)$]:

1. By Lemma 3(b) and (d), we can apply Lemma 4(b), which, using the estimate (3.9), and then (4.3), gives

$$\mathbb{E} W \leq C_2 N^{1/2} \left(\mathbb{E} f_{T_N}^2\right)^{1/2} \leq \frac{3 C_2 N^{1/2}}{(1 - H(T_N -))^{1/2}} \leq C \left(\frac{N}{\varepsilon_N}\right)^{1/2} \qquad (4.7)$$

   for a universal constant $C < \infty$.

2. Since $|f_x - \mathbb{E} f_x| \le \int_{-\infty}^{x} dH/(1-H_-)^2$, Lemma 2 and (4.3) give

$$U = \sup_{x \le T_N} |f_x - \mathbb{E} f_x| \le \int_{-\infty}^{T_N} \frac{dH}{(1-H_-)^2} \le \frac{2}{1 - H(T_N -)} \le \frac{16}{\varepsilon_N}. \tag{4.8}$$

3. Again by (3.9) and (4.3),

$$\begin{aligned} V &\le N \mathbb{E} \| f_x - \mathbb{E} f_x \|^2_{\mathscr{G}_N} \le 4N \mathbb{E} f^2_{T_N} \\ &\le 4N \mathbb{E} \left(\int_{-\infty}^{T_N} \frac{\mathbb{1}_{Z > y}}{(1 - H(y-))^2} \, dH(y) \right)^2 \le 288 \left(\frac{N}{\varepsilon_N} \right). \end{aligned} \tag{4.9}$$

Then, taking $L = u/60 - C$ with u smooth such that $L > 0$, Talagrand's inequality (2.12) and the estimates (4.7)–(4.9) give, by monotonicity of the exponent in (2.12),

$$\begin{aligned} \Pr&\left\{ \left\| \sum_{i=1}^{2^k} (f_x(Z_i) - \mathbb{E} f_x(Z)) \right\|_{\mathscr{G}_{2^k}} > \frac{u}{60} \sqrt{\frac{2^k \log\log 2^k}{\varepsilon_{2^k}}} \right\} \\ &= \Pr\left\{ W > \frac{u}{60} \sqrt{\frac{N \log\log N}{\varepsilon_N}} \right\} \\ &\le \Pr\left\{ |W - \mathbb{E} W| > L \sqrt{\frac{N \log\log N}{\varepsilon_N}} \right\} \\ &\le K \exp\left(-\frac{1}{K} \frac{L\sqrt{N\varepsilon_N \log\log N}}{16} \right. \\ &\qquad \left. \times \log\left[1 + \frac{4L}{57} \sqrt{\frac{\log\log N}{N\varepsilon_N}} \right] \right) \end{aligned} \tag{4.10}$$

[see (4.5) for the first identity]. Since, by (4.1), $\log\log N/(N\varepsilon_N) \le 1/9$, if we set $L = rK$ for r to be chosen below, using the fact that the graph of $y = \log(1 + x)$, $x \le \lambda$, is above the chord joining $(0,0)$ and $(\lambda, \log(1+\lambda))$, it follows that the last term in (4.10) is dominated by

$$K \exp\left(-\frac{3r \log(1 + 4rK/171)}{16} \log\log 2^k \right), \tag{4.11}$$

which, for r large enough, is the general term of a convergent series. Combining (4.10) and (4.11) with (4.6) and applying Borel–Cantelli, we obtain

$$\limsup_{n\to\infty} \sqrt{\frac{n\varepsilon_{2n}}{\log\log n}} W_n \le 60(rK + C) \quad \text{a.s.}$$

To complete the proof, one submits the first summand in (2.2) to the same treatment: taking $f_t(x, y) = \mathbb{1}_{\{x \wedge y \le t,\, x \le y\}}/(1 - H(x \wedge y -))$, $-\infty < t \le T_n$, it follows rather trivially that the corresponding parameters $\mathbb{E}W$, U and V

enjoy the same bounds as those above, except for multiplicative constants, and that this gives (4.4) for the first summand in the definition (2.2) of L_n. We skip the details in order to avoid repetition. □

As a corollary of Theorems 4 and 5, we obtain an improvement on the first-order law of the logarithm for the Nelson–Aalen estimator given in Csörgő [(1996), Theorem 1], and fall short by $(\log\log\log n)^{1/2+\delta}$ from proving one of the conjectures on page 2749 of the same article, or, what is the same, the conjecture is settled under a mild additional condition on $\{\varepsilon_n\}$.

THEOREM 6. *For $n \in \mathbb{N}$, let ε_n and T_n be as in Theorem 4 and let $c_n = d_n \log\log n$ for a sequence $d_n \nearrow \infty$ such that $\sum[kd_{2^k}\log k]^{-1} < \infty$. Then we have*

$$\lim_{n\to\infty}\sqrt{\frac{n\varepsilon_{2n}}{d_n \log\log n}}\sup_{x\le T_n}|\Lambda_n(x)-\Lambda(x)| = 0 \quad a.s. \tag{4.12}$$

If, moreover, for some $C > 0$, $n_0 < \infty$ and d_n as stated,

$$n\varepsilon_{2n} \ge Cd_n \log n \tag{4.13}$$

for all $n \ge n_0$, then

$$\limsup_{n\to\infty}\sqrt{\frac{n\varepsilon_{2n}}{\log\log n}}\sup_{x\le T_n}|\Lambda_n(x)-\Lambda(x)| < \infty \quad a.s. \tag{4.14}$$

PROOF. Direct application of Theorems 4 and 5 gives

$$\begin{aligned}
&\sup_{x\le T_n}|\Lambda_n(x)-\Lambda(x)| \\
&\quad\le \sup_{x\le T_n}|\Lambda_n(x)-\Lambda(x)-L_n(x)| + \sup_{x\le T_n}|L_n(x)| \\
&\quad= o\left(\frac{(c_n\log n)^{1/2}}{n\varepsilon_{2n}}\right) + O\left(\left[\frac{\log\log n}{n\varepsilon_{2n}}\right]^{1/2}\right) \quad \text{a.s.}
\end{aligned} \tag{4.15}$$

Since $\sqrt{n\varepsilon_{2n}} > 3\sqrt{\log n}/\sqrt{2}$ by (4.1), we have

$$\frac{[d_n(\log\log n)(\log n)]^{1/2}}{n\varepsilon_{2n}} \le \frac{\sqrt{2}}{3}\left[\frac{d_n\log\log n}{n\varepsilon_{2n}}\right]^{1/2},$$

and inequality (4.15) gives (4.12) (note $d_n \to \infty$). Under hypothesis (4.13),

$$\frac{[d_n(\log\log n)(\log n)]^{1/2}}{n\varepsilon_{2n}} \le \frac{1}{\sqrt{C}}\left[\frac{\log\log n}{n\varepsilon_{2n}}\right]^{1/2},$$

and (4.14) also follows from (4.15) in this case. □

The following theorem improves the first-order law of the logarithm in Theorem 2 of Csörgő (1996) and comes close to proving a second conjecture on page 2749 of the same article (the comments about Theorem 6 apply here as well).

THEOREM 7. *Assuming F is continuous and letting, for $n \in \mathbb{N}$, ε_n and T_n be as in Theorem* 4, *and $c_n = d_n \log\log n$ be as in Theorem* 6, *we have*

$$\lim_{n\to\infty} \sqrt{\frac{n\varepsilon_{2n}}{d_n \log\log n}} \sup_{x\le T_n} \left| \frac{\hat{F}_n(x) - F(x)}{1-F(x)} \right| = 0 \quad a.s. \tag{4.16}$$

and, under hypothesis (4.13), *also*

$$\limsup_{n\to\infty} \sqrt{\frac{n\varepsilon_{2n}}{\log\log n}} \sup_{x\le T_n} \left| \frac{\hat{F}_n(x) - F(x)}{1-F(x)} \right| < \infty \quad a.s. \tag{4.17}$$

PROOF. From the decomposition (3.4)–(3.6) with $h(x) = 0$ it follows that the rate of $\sup_{x\le T_n}|(\hat{F}_n(x) - F(x))/(1-F(x))|$ is the largest of $\sup_{x\le T_n}|l_n(x)|$ and $\sup_{x\le T_n}|\Lambda_n(x) - \Lambda(x)|$. We can apply the Breslow–Crowley inequality (3.12) to the first sequence and obtain, as in (3.13),

$$\begin{aligned} \sup_{x\le T_n} |l_n(x)| &= \sup_{x\le T_n} \left|\log\bigl(1-\hat{F}_n(x)\bigr) + \Lambda_n(x)\right| \\ &\le \frac{8}{n\varepsilon_n} \sup_{x\le T_n} \left| \frac{1-H(T_n)}{1-H_n(T_n-)} \right| \quad \text{a.s.} \end{aligned} \tag{4.18}$$

Now, if we apply Prohorov's inequality, or, for convenience, Talagrand's exponential inequality for a single function to $n^{-1}\sum_{i=1}^n I_{Z_i\ge T_n}/(1-H(T_n))$, with $U = V = (n\varepsilon)^{-1}$, we obtain

$$\Pr\left\{ \left| \frac{1-H_n(T_n-)}{1-H(T_n-)} - 1 \right| > C\sqrt{\frac{\log\log n}{n\varepsilon_{2n}}} \right\} \le K\exp\{-2\log\log n\}$$

for some finite constant C and all n. Then, proceeding as in the proof of Theorem 5, but using this inequality instead of (4.10), we obtain that

$$\left| \frac{1-H_n(T_n-)}{1-H(T_n-)} - 1 \right| = O\left(\sqrt{\frac{\log\log n}{n\varepsilon_{2n}}} \right) \to 0 \quad \text{a.s.}$$

Replacing this estimate into (4.18) shows that $\sup_{x\le T_n}|l_n(x)|$ is a.s. of the order of $(n\varepsilon_{2n})^{-1}$, dominated by the rate prescribed for the Nelson–Aalen estimator in Theorem 6. Theorem 7 then follows from this and Theorem 6. □

We note that Theorems 6 and 7, concretely the limits (4.14) and (4.17), contain bounded LILs for fixed T, best possible up to constants, both for the Nelson–Aalen and the Kaplan–Meier estimators; in particular, we recover, up to constants, part of Theorem 1 [concretely the limit (1.5)] in Csörgő and Horváth (1983). It is clear that the methods from Section 3, including Remark 2, give the Csörgő–Horváth results for fixed T and best constants, as well as the "compact" or Strassen-type LIL, with a proof that dos not depend on strong approximations.

Finally, we consider the second-order law of the logarithm for the product limit estimator and improve on the rates in Theorem 1.4 from Stute (1994) and in part of Proposition 5 from Csörgő (1996).

THEOREM 8. *Assuming F continuous and letting c_n, ε_n and T_n, $n \in \mathbb{N}$, be as in Theorem* 4, *we have*

$$\lim_{n\to\infty} \frac{n\varepsilon_{2n}}{\sqrt{c_n \log n}} \sup_{x \le T_n} \left| \frac{\hat{F}_n(x) - F(x)}{1 - F(x)} - L_n(x) \right| = 0 \quad a.s. \tag{4.19}$$

If d_n is as in Theorem 6 *and* (ε_n, T_n *as above*) *then*

$$\begin{aligned} &\limsup_{n\to\infty} \frac{n\varepsilon_{2n}}{d_n \log\log n} \\ &\quad \times \sup_{x \le T_n} \left| \frac{\hat{F}_n(x) - F(x)}{1 - F(x)} - (\Lambda_n(x) - \Lambda(x)) \right| = 0 \quad a.s. \end{aligned} \tag{4.20}$$

and if, *moreover*, *the sequence* $\{\varepsilon_n\}$ *satisfies condition* (4.13), *then*

$$\lim_{n\to\infty} \frac{n\varepsilon_{2n}}{\log\log n} \sup_{x \le T_n} \left| \frac{\hat{F}_n(x) - F(x)}{1 - F(x)} - (\Lambda_n(x) - \Lambda(x)) \right| < \infty \quad a.s. \tag{4.21}$$

PROOF. The theorem follows immediately by combining the rates in Theorem 4, Theorem 6 [including (4.15)] and (4.18) with the decomposition (3.4)–(3.6), respectively, for $h(x) = L_n(x)$ and $h(x) = \Lambda_n(x) - \Lambda(x)$. □

REMARK 4. We note that, by Remark 3, the nonrandom end points T_n in Theorems 4 to 8 can be replaced by the random $Z_{n(1-\varepsilon_n), n}$, and it is with these random endpoints that Stute and Csörgő state their results.

REFERENCES

AALEN, O. O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3** 15–27.

ARCONES, M. A. and GINÉ E. (1995). On the law of the iterated logarithm for canonical U-statistics and processes. *Stochastic Process. Appl.* **58** 217–245.

BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

CSÖRGŐ, S. (1996). Universal Gaussian approximations under random censorship. *Ann. Statist.* **24** 2744–2778.

CSÖRGŐ, S. and HORVÁTH, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Z. Wahrsch. Verw. Gebiete* **62** 411–426.

DE LA PEÑA, V. and MONTGOMERY-SMITH, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.* **23** 806–816.

GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–998.

KAPLAN, E. L. and MEIER, P. (1958). Non-parametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

MONTGOMERY-SMITH, S. J. (1993). Comparison of sums of independent identically distributed random vectors. *Probab. Math. Statist.* **14** 281–285.

NELSON, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14** 945–966.

NOLAN, D. and POLLARD, D. (1987). *U*-processes: rates of convergence. *Ann. Statist.* **15** 780–799.

PISIER, G. (1975). Le théorème limite central et la loi du logarithme itéré dans les espaces de Banach. In *Séminaire Maurey-Schwartz, Exposés III et IV*. Ecole Polytechnique, Paris.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

STUTE, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.* **21** 146–156.

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

TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563.

VAPNIK, V. N. and ČERVONENKIS, A. JA. (1981). Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory Probab. Appl.* **26** 532–553.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF CONNECTICUT
STORRS, CONNECTICUT 06269
E-MAIL: gine@uconnvm.uconn.edu

UNIVERSITÉ PARIS VI
L.S.T.A., TOUR 45-55 E3
4 PLACE JUSSIEU
75252 PARIS CEDEX 05
FRANCE
E-MAIL: guillou@ccr.jussieu.fr