Lifetime Data Anal (2007) 13:481–496
DOI 10.1007/s10985-007-9055-z

On an exponential bound for the Kaplan–Meier estimator

Jon A. Wellner

Received: 10 July 2007 / Accepted: 30 July 2007 / Published online: 31 August 2007

Abstract We review limit theory and inequalities for the Kaplan–Meier Kaplan and Meier (J Am Stat Assoc 53:457–481, 1958) product limit estimator of a survival function on the whole line $\mathbb{R}$. Along the way we provide bounds for the constant in an interesting inequality due to Biotouzé et al. (Ann Inst H Poincaré Probab Stat 35:735–763, 1999), and provide some numerical evidence in support of one of their conjectures.

Keywords Random censoring · Right censoring · Product limit estimator · Inequalities · Universal limit theorems · Uniform limit theorems · Confidence bands · Empirical distribution

1 Introduction: the Kaplan–Meier estimator

Suppose that $X_1, \ldots, X_n, \ldots$ are independent and identically distributed with distribution function F, and suppose that $Y_1, \ldots, Y_n$ are independent and identically distributed with distribution function G. Suppose that we only observe $W_i = (Z_i, \delta_i)$ where

$$Z_i = X_i \wedge Y_i, \qquad \delta_i = 1_{[X_i \le Y_i]}.$$

Supported in part by NSF grant DMS-0503822 and by NI-AID grant 2R01 AI291968-04.

J. A. Wellner (✉)
Department of Statistics, University of Washington, 354322, Seattle, WA 98195-4322, USA
e-mail: jaw@stat.washington.edu

Typically in survival analysis the X_i's represent lifetimes and the Y_i's represent censoring times. The above model for (right-)censored data is sometimes called the *random right censoring model*.

Define the empirical sub-distribution functions of the uncensored and censored Z's by

$$\mathbb{H}_n^{uc}(z) = \frac{1}{n}\sum_{i=1}^{n} \delta_i 1_{[Z_i \le z]},$$
$$\mathbb{H}_n^{c}(z) = \frac{1}{n}\sum_{i=1}^{n} (1-\delta_i) 1_{[Z_i \le z]},$$
$$\mathbb{H}_n(z) = \mathbb{H}_n^{uc}(z) + \mathbb{H}_n^{c}(z),$$

and define $\widehat{\Lambda}_n$ by

$$\widehat{\Lambda}_n(x) = \int_{(-\infty,x]} \frac{1}{1-\mathbb{H}_n(z-)} d\mathbb{H}_n^{uc}(z).$$

Then the Kaplan–Meier (or product limit) estimator $\widehat{F}_n$ of the distribution function F is defined by

$$1 - \widehat{F}_n(x) = \prod_{z \le x}(1 - \Delta\widehat{\Lambda}_n(z)) = \prod_{z: Z_i \le x} \left(\frac{n - R_i}{n - R_i + 1}\right)^{\delta_i}$$

where R_i is the rank of $(Z_i, 1-\delta_i)$ in the set $\{(Z_j, 1-\delta_j), j = 1, \ldots, n\}$ with lexicographical ordering.

For the empirical distribution function $\mathbb{F}_n$ of the X_i's, we know from Dvoretzky et al. (1956) and Massart (1990) that

$$P(\sqrt{n}\|\mathbb{F}_n - F\|_\infty > \lambda) \le 2\exp(-2\lambda^2) \tag{1}$$

for all $\lambda > 0$. In the spirit of the exponential bound (1), Bitouzé et al. (1999) established the following exponential bound for the supremum distance between $\widehat{F}_n$ and F weighted by $1 - G$:

Theorem 1 *Let $\widehat{F}_n$ be the Kaplan–Meier estimator of the distribution function F. There is an absolute constant C such that, for any $\lambda > 0$,*

$$P(\sqrt{n}\|(1-G)(\widehat{F}_n - F)\|_\infty > \lambda) \le 2.5\exp(-2\lambda^2 + C\lambda). \tag{2}$$

As noted by Bitouzé et al. (1999),

> "Of course, it would be desirable to compute C, but unfortunately, our techniques are not sharp enough to do that efficiently."

One goal of this paper is to provide the following bound for the constant C:

Proposition 1 *The constant C in* (2) *satisfies* $C \leq 11709$.

In the course of the proof of this proposition, we also provide explicit bounds on the constants involved in the bracketing entropy for bounded monotone functions given in Theorem 2.7.5 of van der Vaart and Wellner (1996); see Theorem 2. This bound on the numerical constants in that theorem may be of independent interest.

Bitouzé et al. (1999) also say (p. 741) that:

> "We do not know whether the inequality
>
> $$P(\sqrt{n}\|(1-G)(\widehat{F}_n - F)\|_\infty > \lambda) \leq 2\exp(-2\lambda^2) \tag{3}$$
>
> holds or not for the Kaplan–Meier estimator. Indeed, this is a natural question since on the one hand we know ... that it holds asymptotically, and on the other hand that it is valid for all n in the non censored case (see Massart (1990)). Inequality (2) does not of course imply (3), but can be considered as a first step toward such a result."

Of course the bound provided by Proposition 1 is absurdly large, especially in view of the numerical evidence in favor of (3), but it gives a strong indication of the limitations of the current method of proof. Significant improvements of the current bound will likely involve the use of local entropy as discussed in van de Geer (1993, 2000), or improvements of the bracketing entropy bound obtained here in Theorem 2, or both.

Here is some numerical evidence in support of (3). The following two scenarios and plots have been selected from a list of about thirty different scenarios. The first scenario is F=exponential(1), G = Uniform $[0, 3]$, $n = 1000$, and $m = 1000$ Monte-Carlo replications. The dashed line shows the conjectured bound $2\exp(-2\lambda^2)$ (3), while the solid line is the empirical survival function (or one minus the empirical distribution function) for the $m = 1000$ computed values of $\sqrt{n}\|(1-G)(\widehat{F}_n - F)\|_\infty$.

In the second scenario, F = exponential(1), $1 - G(t) = (1 - t/2)^{.01}$, $n = 1000$, and $m = 1000$. The dashed line shows the conjectured bound $2\exp(-2\lambda^2)$ (3), while the solid line is the empirical survival function (or one minus the empirical distribution function) for the $m = 1000$ computed values of $\sqrt{n}\|(1-G)(\widehat{F}_n - F)\|_\infty$ (Figs. 1,2).

2 Bracketing entropy for monotone functions

Let $\mathcal{F}$ denote a class of uniformly bounded monotone functions on the real line; without loss we assume that $f : \mathbb{R} \mapsto [0, 1]$ for all $f \in \mathcal{F}$.

Theorem 2 *The class $\mathcal{F}$ of monotone functions $f : \mathbb{R} \mapsto [0, 1]$ satisfies*

$$\log N_{[\,]}(\epsilon, \mathcal{F}, L_r(Q)) \leq \frac{K_r}{\epsilon}$$

for every probability measure Q, every $r \geq 1$, and a constant

$$K_r = 2C_r(2^{r-1}(2^r + C_r))^{1/r}\log(2L_r^2)$$

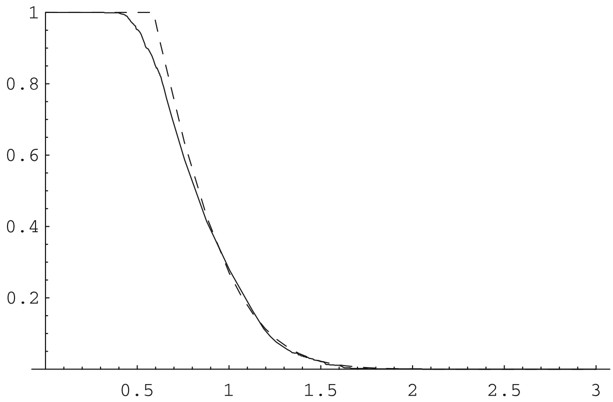

Fig. 1 Empirical survival function (*solid*) and conjectured bound (*dashed*) $n = 1000$, $F =$ exponential(1), $G =$ Uniform(0, 3)

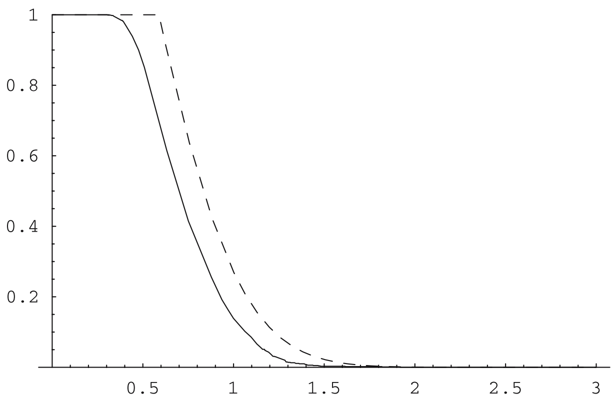

Fig. 2 Empirical survival function (*solid*) and conjectured bound (*dashed*) $n = 1000$, $F =$ exponential(1), $1 - G(t) = (1 - t/2)^{.01}$

where $C_r = 4/(1-(1/2)^{1/(r+1)})^2$ *and* $L_r = 2(2^{1/r}+1)$. *In particular* $C_2 \approx 93.9861$ *and* $K_2 \approx 10110.43$.

Proof This is proved in Section 2.7.2 of van der Vaart and Wellner (1996 pp. 159–162). The constant follows from that proof by careful book-keeping; see Sect. 5 for details. □

3 Limit theory for the Kaplan–Meier estimator

In this section we give a very brief review of known limit theory for the Kaplan–Meier estimator $\widehat{F}_n$, with emphasis on results that hold on either all $\mathbb{R}$ or large subintervals of $[0, \tau_H]$ where, for any distribution function F, let $\tau_F \equiv \inf\{t : F(t) = 1\}$.

3.1 Donsker theorems

The original weak convergence theorem for the Kaplan–Meier estimator established by Breslow and Crowley(1974) showed that

$$\sqrt{n}(\widehat{F}_n - F) \Rightarrow (1-F)\mathbb{B}(C) \quad \text{in } D[0,T] \tag{4}$$

for fixed $T < \tau_H$ where $\mathbb{B}$ is a standard Brownian motion process on $[0,\infty)$ and

$$C(t) \equiv \int_{[0,t]} \frac{1}{(1-H_-)^2} dH^{uc} = \int_{[0,t]} \frac{1}{(1-F_-)^2(1-G_-)} dF.$$

It was noted by Hall and Wellner (1980) that (4) could be rephrased as

$$\sqrt{n}\frac{1-K}{1-F}(\widehat{F}_n - F) \Rightarrow \mathbb{B}^0(K) \quad \text{in } D[0,T] \tag{5}$$

where $\mathbb{B}^0$ is a standard Brownian bridge process on $[0, 1]$ and $K \equiv C/(1+C)$.

The weak convergence result in (4) was improved to a strong approximation result by Burke et al. (1981) as follows: consider intervals of the form $[0, T_n]$ with $T_n \nearrow \tau_H$ satisfying $1 - H(T_n) \geq (2\delta \log n/n)^{1/2}$ where $\delta > 0$. Then $\widehat{F}_n$ can be defined on a common probability space with a (sequence of) Brownian motion process(es) $\mathbb{B}_n$ so that

$$Pr\left(\sup_{0\leq t\leq T_n} |\sqrt{n}(\widehat{F}_n(t) - F(t)) - (1-F(t))\mathbb{B}_n(t)| \geq r_n\right) \leq Cn^{-\delta}$$

where C is a constant, $b_n \equiv 1/(1 - H(T_n))$, and

$$r_n = O(\max\{n^{-1/3}b_n^2(\log n)^{3/2},\ n^{-1/2}b_n^4 \log n,\ n^{-3/2}b_n^6(\log n)^2\}).$$

In a different direction, Gill (1983) showed that

$$\left\{\sqrt{n}\frac{1-K}{1-F}(\widehat{F}_n - F)\right\}^{Z_{(n)}} \Rightarrow \mathbb{B}^0(K) \quad \text{in } D[0,\tau_H] \tag{6}$$

where $Z_{(n)} \equiv \max_{1\leq i\leq n} Z_i$ and for any process W the stopped process $W^T \equiv W(\cdot \wedge T)$. Moreover, Gill (1983) showed that if

$$\int_0^{\tau_H} \frac{dF}{1-G_-} < \infty, \tag{7}$$

then

$$\{\sqrt{n}(\widehat{F}_n - F)\}^{Z_{(n)}} \Rightarrow (1-F)\mathbb{B}(C) \quad \text{in } D[0,\tau_H] \tag{8}$$

and

$$\left\{\sqrt{n}\frac{1-\widehat{K}}{1-\widehat{F}}(\widehat{F}_n - F)\right\}^{Z_{(n)}} \Rightarrow \mathbb{B}^0(K) \quad \text{in } D[0,\tau_H]. \tag{9}$$

These conclusions were strengthened by Ying(1989) who showed that stopping of the process at $Z_{(n)}$ is not needed; e.g.,

$$\sqrt{n}\frac{1-K}{1-F}(\widehat{F}_n - F) \Rightarrow \mathbb{B}^0(K) \quad \text{in } D[0,\tau_H], \tag{10}$$

and if (7) holds, then

$$\sqrt{n}(\widehat{F}_n - F) \Rightarrow (1-F)\mathbb{B}(C) \quad \text{in } D[0,\tau_H] \tag{11}$$

and

$$\sqrt{n}\frac{1-\widehat{K}}{1-\widehat{F}}(\widehat{F}_n - F) \Rightarrow \mathbb{B}^0(K) \quad \text{in } D[0,\tau_H]. \tag{12}$$

On the other hand Chen and Ying (1996) showed that (9) can fail if (7) does not hold, providing a counterexample to a conjecture of Hall and Wellner (1980). See Gillespie and Fisher (1979), Nair (1981), Csörgő and Horváth (1986), Hollander and Peña (1989), Hollander et al. (1997), and McKeague and Zhao (2006) for other related work on confidence bands.

3.2 Glivenko–Cantelli theorems

After incorrect proofs of Glivenko–Cantelli type theorems for the Kaplan–Meier estimator by Gill (1980) and Shorack and Wellner (1986), Wang (1987) succeeded in establishing the first correct (and clean) theorem. Wang showed that

$$\sup_{t<\tau_H} |\widehat{F}(t) - F(t)| \to_p 0, \quad \text{and} \tag{13}$$

$$\sup_{t<Z_{(n)}} |\widehat{F}(t) - F(t)| \to_p 0 \tag{14}$$

always hold, and hence that

$$\sup_{t\le\tau_H} |\widehat{F}(t) - F(t)| \to_p 0 \tag{15}$$

if and only if $G(\tau_H) < 1$ or $\Delta F(\tau_H) = F\{\tau_H\} = 0$. Wang (1987) also pointed out an error in Shorack and Wellner (1986) and gave a correct statement of the result which could be proved by the methods used by Shorack and Wellner. Stute and Wang (1993) showed that (15) also holds with $\rightarrow_{a.s.}$ under the same conditions. Gill (1994) provided an alternative proof of the result of Stute and Wang (1993). Stute (1994a) obtained a weighted Glivenko–Cantelli type theorem by showing that if $\psi = \psi_1 + \psi_2$ is a U-type weight function where ψ_1 is $\downarrow$ and ψ_2 is $\uparrow$, then if

$$\int_0^{\tau_H} \psi dF < \infty, \qquad \int_0^{\tau_H} \psi_2^2 dF < \infty,$$
$$\int \frac{dF}{1-G} < \infty, \qquad \text{and} \qquad \int C^{1/2} dF < \infty,$$

it follows that

$$\sup_{t \le \tau_H} \psi(t)|\widehat{F}_n(t) - F(t)| \rightarrow_p 0.$$

Stute (1994a) also indicated how this could be extended to an almost sure result.

3.3 Univeral approximation theorems on increasing sets

Gu and Lai (1990) and Gu (1991) apparently initiated the study of

$$\sup_{t \le Z_{(n-k_n)}} \frac{|\widehat{F}_n(t) - F(t)|}{1 - F(t)} \tag{16}$$

for $k_n \in \{1, \ldots, n-1\}$. Stute (1994b) continued this study and showed that if $(n - k_n)/n \uparrow$ and $k_n/\log n \rightarrow \infty$, then for each $\epsilon > 0$

$$\sup_{t \le Z_{(n-k_n)}} |\widehat{\Lambda}_n(t) - \Lambda(t)| = O\left(\frac{(\log n)^{1+\epsilon}}{k_{2n}} + \frac{(\log n)^{(1+\epsilon)/2}}{\sqrt{k_n}}\right) \equiv O(\gamma_n)$$

almost surely, and

$$\sup_{t \le Z_{(n-k_n))}} |\widehat{F}_n(t) - F(t)| = O(\gamma_n).$$

The results of Stute (1994b) were improved and sharpened by Csörgő (1996) who showed that

$$\sup_{t \le Z_{(n-k_n)}} |\widehat{\Lambda}_n(t) - \Lambda(t)| = \begin{cases} O\left(\frac{\sqrt{\log n}}{\sqrt{k_{2n}}}\right) & \text{almost surely} \\ O_p\left(\frac{1}{\sqrt{k_n}}\right) & \text{in probability.} \end{cases} \tag{17}$$

and similarly with the left side in (17) replaced by the supremum in (16). The almost sure part of Csörgő's bounds were further improved by Giné and Guillou (1999) who showed that the $\sqrt{\log n}$ in the almost sure bounds can be replaced by $\sqrt{\log \log n}$.

Alternative derivations of limit theory for the Kaplan–Meier estimator $\widehat{F}_n$ under the assumption of constant censoring times have been given by Meier (1975) and Pollard (1990). It would be interesting to try to extend the results of these authors to the whole line or $[0, \tau_H]$ with some appropriate re-definition of $[0, \tau_H]$. For a treatment of some rate of convergence results in the encompassing framework of "variable censoring" in which each censoring time Y_i is allowed to have its own distribution G_i, see Csörgő and Horváth (1983, sect. 2, pp. 417–419).

4 Proof of Proposition 1

From Theorem 5 of Bitouzé et al. (1999), the proof on p. 756, and rephrasing in the notation of van der Vaart and Wellner (1996), it follows that

$$P^*\left(\sup_{1\le k\le n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}|\mathbb{G}_k(f)| > \lambda + c_1\epsilon_0\frac{\lambda}{\sigma} + c_2\varphi(\epsilon_0)\right) \le 2.5e^{H(\epsilon_0)}\exp(-2\lambda^2/a^2) \tag{18}$$

where $c_1 = 7.9/\sqrt{2}$, $c_2 = 44.7$, $0 < \epsilon_0 < 1$, $\sup_{f\in\mathcal{F}} Var_P(f) \le \sigma^2 \le a^2/4$, and

$$\varphi(t) = \int_0^t \sqrt{\log N_{[\,]}(x, \mathcal{F}, L_2(P))}\,dx \equiv \int_0^t \sqrt{H(x)}\,dx.$$

In the context of Bitouzé et al. (1999), Corollary 2,

$$\varphi(t) \le \int_0^t \sqrt{\gamma/x}dx = 2\sqrt{\gamma}t^{1/2}$$

where $\gamma \equiv K_2 \approx 10110.43$ by Theorem 2, so the inequality in (18) becomes

$$P^*\left(\sup_{1\le k\le n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}|\mathbb{G}_k(f)| > \lambda(1 + c_1\epsilon_0) + c_3\sqrt{\epsilon_0}\right) \le 2.5e^{\gamma/\epsilon_0}\exp(-2\lambda^2/a^2) \tag{19}$$

where $c_3 = 2\sqrt{\gamma}c_2 = 2\sqrt{K_2}c_2 = 2\cdot 44.7\cdot\sqrt{K_2}$.

Set $\xi \equiv \lambda(1 + c_1\epsilon_0) + c_3\sqrt{\epsilon_0}$; then for $\xi \ge c_3\sqrt{\epsilon_0}$,

$$P^*\left(\sup_{1\le k\le n}\sup_{f\in\mathcal{F}}\sqrt{\frac{k}{n}}|\mathbb{G}_k(f)| > \xi\right) \le 2.5e^{\gamma/\epsilon_0}\exp\left(-\frac{2}{a^2}\left(\frac{\xi - c_3\sqrt{\epsilon_0}}{1 + c_1\epsilon_0}\right)^2\right). \tag{20}$$

Note that for $\epsilon_0 < 1/c_1$ we have

$$\begin{aligned}\left(\frac{\xi - c_3\sqrt{\epsilon_0}}{1 + c_1\epsilon_0}\right)^2 &\geq (\xi - c_3\sqrt{\epsilon_0})^2(1 - c_1\epsilon_0)^2 \\ &\geq \xi^2 - 2c_1\epsilon_0\xi^2 - 2c_3\sqrt{\epsilon_0}\xi(1 - c_1\epsilon_0)^2 \\ &\geq \xi^2 - 2c_1\epsilon_0\xi^2 - 2c_3\sqrt{\epsilon_0}\xi.\end{aligned}$$

Taking $\epsilon_0 = b/\xi \leq 1/(2c_1)$ for a constant b (which will be chosen appropriately) so that $\xi \geq 2bc_1$, the bound becomes

$$\begin{aligned}&2.5\exp\left(\frac{\gamma\xi}{b} + \frac{4c_1\xi}{a^2}b + \frac{4c_3}{a^2}\sqrt{b\xi}\right)\exp(-2\xi^2/a^2) \\ &= 2.5\exp\left(\frac{2}{a}(4\gamma c_1)^{1/2}\xi + \frac{4c_3}{a^2}\frac{\gamma^{1/4}}{(4c_1/a^2)^{1/4}}\sqrt{\xi}\right)\exp(-2\xi^2/a^2) \\ &\quad \text{by choosing} \quad b = a\sqrt{\gamma/(4c_1)} \quad \text{so that} \\ &\quad \xi \geq 2ac_1\sqrt{\gamma/(4c_1)} = a\sqrt{\gamma c_1}, \\ &\quad \text{and hence} \quad \sqrt{\xi} \geq \sqrt{a}(\gamma c_1)^{1/4}, \\ &\leq 2.5\exp\left(\frac{4}{a}(\gamma c_1)^{1/2}\xi + \frac{c_3}{c_1}\frac{4c_1/a^2}{(4c_1/a^2)^{1/4}}\gamma^{1/4}\sqrt{\xi}\right)\exp(-2\xi^2/a^2) \\ &\leq 2.5\exp\left(\frac{4}{a}(\gamma c_1)^{1/2}\xi + \frac{c_3}{c_1}(4c_1/a^2)^{3/4}\frac{\sqrt{\xi}}{\sqrt{a}(\gamma c_1)^{1/4}}\gamma^{1/4}\sqrt{\xi}\right)\exp(-2\xi^2/a^2) \\ &= 2.5\exp\left(\frac{4}{a}(\gamma c_1)^{1/2}\xi + \frac{c_3}{c_1}(4c_1/a^2)^{3/4}\frac{1}{\sqrt{a}(c_1)^{1/4}}\xi\right)\exp(-2\xi^2/a^2) \\ &= 2.5\exp\left\{\left(\frac{4}{a}(\gamma c_1)^{1/2} + \frac{2^{3/2}c_3}{a^2\sqrt{c_1}}\right)\xi\right\}\exp(-2\xi^2/a^2).\end{aligned}$$

In our particular problem (which corresponds to Theorem 2 of Bitouzé et al. (1999)), $a = 1$ (from Bitouzé et al. (1999), proof of Theorem 2, pp. 745–746), and $\gamma = K_2$ is given by Theorem 2 above, so we calculate

$$\frac{4}{a}(\gamma c_1)^{1/2} + \frac{2^{3/2}c_3}{a^2\sqrt{c_1}} \approx 11708.11.$$

Note that the bound of (2) holds trivially for $\lambda = \xi \leq 5055.2$ since the right side with the above choice of C is ≥ 1 for all such values of λ; thus the exponential bound holds trivially for $\lambda = \xi < 2bc_1 \approx 237.65$. This completes the proof of Proposition 1. □

5 Proof of Theorem 2

Proof Let $\mathcal{G}$ denote the collection of all distribution functions concentrated on [0, 1] (i.e., corresponding to the distributions of random variables V with $V \in [0, 1]$ with probability 1). As shown in van der Vaart and Wellner (1996), we can reduce to finding a bound for $\log N_{[\,]}(\epsilon, \mathcal{G}, L_r(\lambda))$ for the uniform probability measure λ on [0, 1] (i.e., Lebesgue measure).

We say F^L is a left $\epsilon-$bracket for $F \in \mathcal{G}$ if $F^L(x) \le F(x)$ for all $x \in [0, 1]$, and $\|F^L - F\|_{\lambda,r} \le \epsilon$; similarly, F^R is a right $\epsilon-$bracket for $F \in \mathcal{G}$ if $F^R(x) \ge F(x)$ for all $x \in [0, 1]$, and $\|F^R - F\|_{\lambda,r} \le \epsilon$. To obtain an upper bound for the bracketing number for the class $\mathcal{G}$, the key is to define left and right $\epsilon-$brackets and count the number require to cover $\mathcal{G}$. As in the proof of Theorem 2.7.5 in van der Vaart and Wellner (1996) (which is based on van de Geer (1991) who, in turn, used results and methods of Birman and Solomjak (1967)), the idea here is based on different levels of partitioning of the unit interval [0, 1]. At each level of partitioning, the class $\mathcal{G}$ is partitioned into finitely many subsets. Then we construct left and right brackets, so that they have size (proportional to) ϵ, and find an upper bound for the total number of brackets.

Fix $\epsilon > 0$. Let $c = 2^{-1/r}$. Fix $F \in \mathcal{G}$. Let $\mathcal{P}_0$ be the (trivial) partition of the unit interval $0 = u_0^{(0)} < u_1^{(0)} = 1$. The i-th partition $\mathcal{P}_i$ of [0, 1] is of the form

$$0 = u_0^{(i)} < u_1^{(i)} < \cdots < u_{n^{(i)}}^{(i)} = 1.$$

Given a partition $\mathcal{P}_i$, define

$$\epsilon_i \equiv \epsilon_i(F) = \max_{1 \le j \le n^{(i)}} \left(\left[F(u_j^{(i)}) - F(u_{j-1}^{(i)}) \right] (u_j^{(i)} - u_{j-1}^{(i)})^{1/r} \right).$$

To form the next partition $\mathcal{P}_{i+1} \supset \mathcal{P}_i$, find all the intervals in $(u_{j-1}^{(i)}, u_j^{(i)}]$ in the partition $\mathcal{P}_i$ that satisfy

$$\left[F(u_j^{(i)}) - F(u_{j-1}^{(i)}) \right] (u_j^{(i)} - u_{j-1}^{(i)})^{1/r} \ge c\epsilon_i. \tag{21}$$

Then, the partition $\mathcal{P}_{i+1}$ is obtained from $\mathcal{P}_i$ by dividing all the intervals $[u_{j-1}^{(i)}, u_j^{(i)})$ that satisfy (21) into two halves of equal length. The partition $\mathcal{P}_{i+1}$ is obtained by forming $S_j^{(i+1)} = (u_{j-1}^{(i+1)}, u_j^{(i+1)}]$, where $j \in \{1, 2, \ldots, n^{(i+1)}\}$.

Let $(u_{j-1}^{(i+1)}, u_j^{(i+1)}]$ be an interval in the partition $\mathcal{P}_{i+1}$ that is contained in the interval $(u_{j-1}^{(i)}, u_j^{(i)}]$ in partition $\mathcal{P}_i$. If (21) is not satisfied for an interval $(u_{j-1}^{(i)}, u_j^{(i)}]$, then

$$\begin{aligned} &\left[F(u_j^{(i+1)}) - F(u_{j-1}^{(i+1)}) \right] (u_j^{(i+1)} - u_{j-1}^{(i+1)})^{1/r} \\ &= \left[F(u_j^{(i)}) - F(u_{j-1}^{(i)}) \right] (u_j^{(i)} - u_{j-1}^{(i)})^{1/r} \\ &\le \epsilon_{i+1} \le c\epsilon_i. \end{aligned}$$

If (21) is satisfied for $[u_{j-1}^{(i)}, u_j^{(i)})$ in $\mathcal{P}_i$, then $2(u_j^{(i+1)} - u_{j-1}^{(i+1)}) = u_j^{(i)} - u_{j-1}^{(i)}$, and

$$F(u_j^{(i+1)}) - F(u_{j-1}^{(i+1)}) \le F(u_j^{(i)}) - F(u_{j-1}^{(i)}).$$

It follows that

$$\left[F(u_j^{(i+1)}) - F(u_{j-1}^{(i+1)})\right](u_j^{(i+1)} - u_{j-1}^{(i+1)})^{1/r} \\ \leq c\left[F(u_j^{(i)}) - F(u_{j-1}^{(i)})\right](u_j^{(i)} - u_{j-1}^{(i)})^{1/r}.$$

This shows again that $\epsilon_{i+1} \leq c\epsilon_i$.

On the other hand, let $(u_{j-1}^{(i)}, u_j^{(i)}]$ be an interval in the partition $\mathcal{P}_i$ that satisfies (21) and

$$\epsilon_i = \left[F(u_j^{(i)}) - F(u_{j-1}^{(i)})\right](u_j^{(i)} - u_{j-1}^{(i)})^{1/r}.$$

It follows from the definition of ϵ_{i+1} that

$$\begin{aligned} 2\epsilon_{i+1} \geq & \left[F\left(\frac{u_j^{(i)} + u_{j-1}^{(i)}}{2}\right) - F\left(u_{j-1}^{(i)}\right)\right]\left(\frac{u_j^{(i)} + u_{j-1}^{(i)}}{2} - u_{j-1}^{(i)}\right)^{1/r} \\ & + \left[F\left(u_j^{(i)}\right) - F\left(\frac{u_j^{(i)} + u_{j-1}^{(i)}}{2}\right)\right]\left(u_j^{(i)} - \frac{u_j^{(i)} + u_{j-1}^{(i)}}{2}\right)^{1/r} \\ = & \, c\left[F(u_j^{(i)}) - F(u_{j-1}^{(i)})\right]\left(u_j^{(i)} - u_{j-1}^{(i)}\right)^{1/r} = c\epsilon_i. \end{aligned}$$

Hence, we have

$$\epsilon_{i+1} \leq c\epsilon_i \leq 2\epsilon_{i+1}. \tag{22}$$

Let $n_i \equiv n_i(F) = n^{(i)}$ be the number of intervals in the partition $\mathcal{P}_i$. Let $s_i = n_{i+1} - n_i$ be the number of intervals in $\mathcal{P}_i$ that are divided to obtain $\mathcal{P}_{i+1}$. It follows from the definitions of s_i, ϵ_i and Hölder's inequality (with $p = (r+1)/r$ and $q = r + 1$) that

$$\begin{aligned} s_i\,(c\epsilon_i)^{\frac{r}{r+1}} & \leq \sum_{j=1}^{n^{(i)}}\left[F(u_j^{(i)}) - F(u_{j-1}^{(i)})\right]^{\frac{r}{r+1}}(u_j^{(i)} - u_{j-1}^{(i)})^{\frac{1}{r+1}} \\ & \leq \left[\sum_{j=1}^{n^{(i)}}\left(F(u_j^{(i)}) - F_k(u_{j-1}^{(i)})\right)\right]^{\frac{r}{r+1}}\left[\sum_{j=1}^{n^{(i)}}\left(u_j^{(i)} - u_{j-1}^{(i)}\right)\right]^{\frac{1}{r+1}} \\ & = 1. \end{aligned}$$

Therefore

$$s_i \leq (c\epsilon_i)^{-\frac{r}{r+1}}.$$

Consequently,

$$\sum_{j=1}^{i} n_j = i + \sum_{j=1}^{i} j s_{i-j} \le \left(i + \sum_{j=1}^{i} j (c\epsilon_{i-j})^{-\frac{r}{r+1}} \right) \le 2 \sum_{j=1}^{i} j (c\epsilon_{i-j})^{-\frac{r}{r+1}}$$

since

$$\sum_{j=1}^{i} j (c\epsilon_{i-j})^{-\frac{r}{r+1}} \ge i c^{-\frac{r}{r+1}} \epsilon_0^{-\frac{r}{r+1}} = i 2^{1/(r+1)} \cdot 1 \ge i \, .$$

Note that, by (22),

$$c\epsilon_{i-j} \ge \epsilon_{i-j+1} \ge \frac{\epsilon_{i-j+2}}{c} \ge \frac{\epsilon_{i-j+3}}{c^2} \ge \cdots \ge \frac{\epsilon_i}{c^{j-1}}.$$

This implies that

$$\left(c\epsilon_{i-j}\right)^{-\frac{r}{r+1}} \le c^{\frac{(j-1)r}{r+1}} \epsilon_i^{-\frac{r}{r+1}}.$$

Thus,

$$\sum_{j=1}^{i} n_j \le 2 \sum_{j=1}^{i} j c^{\frac{(j-1)r}{r+1}} \epsilon_i^{-\frac{r}{r+1}} = 2 \sum_{j=1}^{i} j \left(\frac{1}{2}\right)^{\frac{j-1}{r+1}} \epsilon_i^{-\frac{r}{r+1}} \equiv 2 T_i \epsilon_i^{-\frac{r}{r+1}}.$$

Then, we have

$$\begin{aligned} T_i - \left(\frac{1}{2}\right)^{\frac{1}{r+1}} T_i &= 1 + \left(\frac{1}{2}\right)^{\frac{1}{r+1}} + \left(\frac{1}{2}\right)^{\frac{2}{r+1}} + \cdots + \left(\frac{1}{2}\right)^{\frac{i-1}{r+1}} - i \left(\frac{1}{2}\right)^{\frac{i}{r+1}} \\ &= \frac{1 - \left(\frac{1}{2}\right)^{\frac{i}{r+1}}}{1 - \left(\frac{1}{2}\right)^{\frac{1}{r+1}}} - i \left(\frac{1}{2}\right)^{\frac{i}{r+1}} \le \frac{1}{1 - \left(\frac{1}{2}\right)^{\frac{1}{r+1}}} \, . \end{aligned}$$

It follows that

$$T_i \le \left[1 - \left(\frac{1}{2}\right)^{\frac{1}{r+1}} \right]^{-2} \equiv \frac{1}{2} C'(r) \, .$$

Hence,

$$\sum_{j=1}^{i} n_j \le C'(r) \epsilon_i^{-\frac{r}{r+1}}. \tag{23}$$

The above partitions $\mathcal{P}_0 \subset \mathcal{P}_1 \subset \cdots$ are generated by a fixed function $F \in \mathcal{G}$. Two monotone functions F and G are said to be equivalent at level i if their partitions up to the ith level are the same. As we continue the partitioning process, the class $\mathcal{G}$ is equivalently partitioned into a finite number of subsets. Let L denote the first such level of partition that $\epsilon_L(F)^r \leq \epsilon^{r+1}$ for every F in that subset of $\mathcal{G}$. For each subset and i, define

$$\tilde{\epsilon}_i = \sup_F \epsilon_i(F),$$

where F ranges over the subset of $\mathcal{G}$ in the final level partition. Note that the number of intervals $n_0 \leq n_1 \leq \cdots \leq n_i$ only depends on the sequence of partitions $\mathcal{P}_0 \subset \mathcal{P}_1 \subset \cdots \subset \mathcal{P}_i$. Thus, the ϵ_i in the inequality (23) can be replaced by $\tilde{\epsilon}_i$. To obtain the upper bound for the bracketing number $N_{[\,]}(\epsilon, \mathcal{G}, L_r(\lambda))$, we first find an upper bound for the total number of different final partitions. By definition of the L-th level partition, we know that there exists a function $G \in \mathcal{G}$ such that $\epsilon_{L-1}(G)^r > \epsilon^{r+1}$ at the $(L-1)$th level partition. This implies that $\epsilon_{L-1}(G)^{-\frac{r}{r+1}} \leq 1/\epsilon$. Then it follows from (23) that

$$\sum_{j=1}^{L-1} n_j \leq C'(r)\epsilon_{L-1}^{-\frac{r}{r+1}} \leq \frac{C'(r)}{\epsilon}.$$

Since $n_L \leq 2n_{L-1}$, then we have

$$1 = n_0 \leq n_1 \leq \cdots \leq n_L \leq \frac{C}{\epsilon}, \tag{24}$$

where $C \equiv C(r) = 2C'(r) = 4\left[1 - (1/2)^{1/(r+1)}\right]^{-2}$.

To count the number of sequences $n_0 \leq n_1 \leq \cdots \leq n_L \leq C/\epsilon$, note that this is equivalent to picking L numbers out of $\{1, 2, 3, \ldots, \lfloor C/\epsilon \rfloor\}$, and this is easily seen to be is $\binom{\lfloor C/\epsilon \rfloor}{L}$, which is bounded by $2^{C/\epsilon}$. For a given sequence $n_0 \leq n_1 \leq \cdots \leq n_L$, the number of ways to obtain $\mathcal{P}_{i+1}$ from $\mathcal{P}_i$ is $\binom{n_i}{s_i}$ which is bounded by 2^{n_i}. Thus, the total number of different final partitions of the form $\mathcal{P}_0 \subset \mathcal{P}_1 \subset \cdots \subset \mathcal{P}_L$ when F ranges over $\mathcal{G}$ is bounded by

$$2^{C/\epsilon} 2^{n_1} \cdots 2^{n_{L-1}} \leq 2^{2C/\epsilon}$$

since $\sum_{i=1}^{L-1} n_i \leq C/\epsilon$.

We now define the bracket $[F_l, F_r]$ for a fixed function $F \in \mathcal{G}$ in the final stage partition $\mathcal{P}_L$. For $u \in [0, 1]$, let

$$F_l(u) = \sum_{j=1}^{n^{(L)}} \left\{ \epsilon \left\lfloor F\left(u_{j-1}^{(L)}\right) / \epsilon \right\rfloor 1_{S_j^{(L)}}(u) \right\},$$

and

$$F_r(u) = \sum_{j=1}^{n^{(L)}} \left\{ \epsilon \left\lceil F\left(u_j^{(L)}\right)/\epsilon \right\rceil 1_{S_j^{(L)}}(u) \right\}.$$

For $u \in S_j^{(L)}$, it follows from (21), the definitions of the brackets, and $\tilde{\epsilon}_L$ that

$$\begin{aligned} F_r(u) - F_l(u) &\leq \epsilon \left[\frac{F\left(u_j^{(L)}\right)}{\epsilon} + 1 \right] - \epsilon \left[\frac{F\left(u_{j-1}^{(L)}\right)}{\epsilon} - 1 \right] \\ &= 2\epsilon + F\left(u_j^{(L)}\right) - F\left(u_{j-1}^{(L)}\right) \\ &\leq 2\epsilon + \left\{ \frac{\epsilon_L}{\left(u_j^{(L)} - u_{j-1}^{(L)}\right)^{1/r}} \right\} \\ &\leq 2\epsilon + \left\{ \frac{\tilde{\epsilon}_L}{\left(u_j^{(L)} - u_{j-1}^{(L)}\right)^{1/r}} \right\}. \end{aligned}$$

Thus, we have, using the inequality $|x+y|^r \leq 2^{r-1}\{|x|^r + |y|^r\}$,

$$\begin{aligned} &\|F_r(u) - F_l(u)\|_{\lambda,r}^r \\ &\leq \sum_{j=1}^{n^{(L)}} \left| 2\epsilon + \tilde{\epsilon}_L \left[\frac{1}{(u_j^{(L)} - u_{j-1}^{(L)})^{1/r}} \right] \right|^r \lambda(S_j^{(L)}) \\ &\leq \sum_{j=1}^{n^{(L)}} 2^{r-1} \left\{ (2\epsilon)^r + (\tilde{\epsilon}_L)^r \left[\frac{1}{u_j^{(L)} - u_{j-1}^{(L)}} \right] \right\} \lambda(S_j^{(L)}) \\ &= 2^{r-1} \left\{ (2\epsilon)^r \sum_{j=1}^{n^{(L)}} \lambda(S_j^{(L)}) + (\tilde{\epsilon}_L)^r \sum_{j=1}^{n^{(L)}} \left[\frac{\lambda(S_j^{(L)})}{u_j^{(L)} - u_{j-1}^{(L)}} \right] \right\} \\ &= 2^{r-1} \left\{ (2\epsilon)^r + (\tilde{\epsilon}_L)^r n^{(L)} \right\} \\ &\leq 2^{r-1} \left\{ (2\epsilon)^r + \epsilon^{r+1} \left[\frac{C}{\epsilon} \right] \right\} = 2^{r-1}(2^r + C)\epsilon^r \end{aligned} \tag{25}$$

by (24). Here we have also used the fact that λ is the uniform distribution, and hence $u_j^{(L)} - u_{j-1}^{(L)} = \lambda(S_j^{(L)})$. The inequality (25) implies that the brackets $[F_l, F_r]$ have the right size.

Now we count how many brackets can be constructed on the final stage partition $\mathcal{P}_L$ when F ranges over the subset of $\mathcal{G}$ for a given sequence of partitions $\mathcal{P}_0 \subset \mathcal{P}_1 \subset \cdots \subset \mathcal{P}_L$. Let

$$l_j = \left\lfloor \frac{F(u_j)}{\epsilon} \right\rfloor, \qquad r_j = \left\lceil \frac{F(u_j)}{\epsilon} \right\rceil.$$

As argued in van der Vaart and Wellner (1996, p. 162), the number of left brackets for a given final partition is at most $L^{2C/\epsilon}$ with $L = 2(1/c + 1) = 2(2^{1/r} + 1)$, and similarly for the number of right brackets. Thus we conclude that

$$N_{[\,]}\left(\left[2^{r-1}(2^r + C)\right]^{1/r} \epsilon, \mathcal{G}, L_r(\lambda)\right) \\ \leq 2^{2C/\epsilon} \times L^{2C/\epsilon} \times L^{2C/\epsilon} = (2L^2)^{2C/\epsilon},$$

or, equivalently

$$\log N_{[\,]}\left(\left[2^{r-1}(2^r + C)\right]^{1/r} \epsilon, \mathcal{G}, L_r(\lambda)\right) \leq \frac{2C \log(2L^2)}{\epsilon}.$$

That is,

$$\log N_{[\,]}(\epsilon, \mathcal{G}, L_r(\lambda)) \leq \frac{K}{\epsilon},$$

where $K = K(r) \equiv [2C \log(2L^2)] \left[2^{r-1}(2^r + C)\right]^{1/r}$. Combining this with the conclusion of the reduction step at the beginning of the proof yields the stated bound. □

Acknowledgements I owe thanks to John Crowley for stimulating my interest in the Kaplan–Meier estimator during conversations in Seattle in 1974–1975 concerning his work with Norman Breslow on the asymptotic theory of the Kaplan–Meier estimator. I also owe thanks to Richard Gill for his many probing questions and challenges about censored data over the years, and to Shuguang Song for help with the computation of the constants in Theorem 2.

References

Birman MŠ, Solomjak MZ (1967) Piecewise polynomial approximations of functions of classes $W_p{}^{\alpha}$. Mat Sb (NS) 73(115):331–355

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

Breslow N, Crowley J (1974) A large sample study of the life table and product limit estimates under random censorship. Ann Stat 2:437–453

Burke MD, Csörgő S, Horváth L (1981) Strong approximations of some biometric estimates under random censorship. Z Wahrsch Verw Gebiete 56:87–112

Chen K, Ying Z (1996) A counterexample to a conjecture concerning the Hall-Wellner band. Ann Stat 24:641–646

Csörgő S (1996) Universal Gaussian approximations under random censorship. Ann Stat 24:2744–2778

Csörgő S, Horváth L (1983) The rate of strong uniform consistency for the product-limit estimator. Z Wahrsch Verw Gebiete 62:411–426

Csörgő S, Horváth L (1986) Confidence bands from censored samples. Canad J Stat 14:131–144

Dvoretzky A, Kiefer J, Wolfowitz J (1956) Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann Math Stat 27:642–669

Gill RD (1980) Censoring and stochastic integrals, vol. 124 of Mathematical Centre Tracts. Mathematisch Centrum, Amsterdam

Gill R (1983) Large sample behaviour of the product-limit estimator on the whole line. Ann Stat 11:49–58

Gill RD (1994) Glivenko–Cantelli for Kaplan–Meier. Math Methods Stat 3:76–87

Gillespie MJ, Fisher L (1979) Confidence bands for the Kaplan–Meier survival curve estimate. Ann Stat 7:920–924

Giné E, Guillou A (1999) Laws of the iterated logarithm for censored data. Ann Probab 27:2042–2067

Gu M (1991) The Chung–Smirnov law for the product-limit estimator under random censorship. Chin. Ann Math Ser B 12:96–105. A Chinese summary appears in Chin Ann Math Ser A 12(1):119

Gu MG, Lai TL (1990) Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation. Ann Probab 18:160–189

Hall WJ, Wellner JA (1980) Confidence bands for a survival curve from censored data. Biometrika 67:133–143

Hollander M, McKeague IW, Yang J (1997) Likelihood ratio-based confidence bands for survival functions. J Am Stat Assoc 92:215–226

Hollander M, Peña E (1989) Families of confidence bands for the survival function under the general random censorship model and the Koziol-Green model. Can J Stat 17:59–74

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457–481

Massart P (1990) The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. Ann Probab 18:1269–1283

McKeague IW, Zhao Y (2006) Width-scaled confidence bands for survival functions. Stat Probab Lett 76:327–339

Meier, P (1975) Estimation of a distribution function from incomplete observations. In Perspectives in probability and statistics (papers in honour of M. S. Bartlett on the occasion of his 65th birthday). Applied Probability Trust, Univ. Sheffield, Sheffield, pp. 67–87

Nair VN (1981) Plots and tests for goodness of fit with randomly censored data. Biometrika 68:99–103

Pollard D (1990) Empirical processes: theory and applications. NSF-CBMS regional conference series in probability and statistics, 2, Institute of Mathematical Statistics, Hayward, CA

Shorack GR, Wellner JA (1986) Empirical processes with applications to statistics. Wiley series in probability and mathematical statistics: probability and mathematical statistics. Wiley, New York

Stute W (1994a) Convergence of the Kaplan–Meier estimator in weighted sup-norms. Stat Probab Lett 20:219–223

Stute W (1994b) Strong and weak representations of cumulative hazard function and Kaplan–Meier estimators on increasing sets. J Stat Plan Inference 42:315–329

Stute W, Wang J-L (1993) The strong law under random censorship. Ann Stat 21:1591–1607

van de Geer S (1991) The entropy bound for monotone functions. Tech Rep Report 91-10, University of Leiden

van de Geer S (1993) Hellinger-consistency of certain nonparametric maximum likelihood estimators. Ann Stat 21:14–44

van de Geer SA (2000) Applications of empirical process theory, vol. 6 of Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge

van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer Series in Statistics, Springer-Verlag, New York. With applications to statistics

Wang JG (1987) A note on the uniform consistency of the Kaplan–Meier estimator. Ann Stat 15:1313–1316

Ying Z (1989) A note on the asymptotic properties of the product-limit estimator on the whole line. Stat Probab Lett 7:311–314