The Annals of Statistics
1987, Vol. 15, No. 3, 1313–1316

A NOTE ON THE UNIFORM CONSISTENCY OF THE KAPLAN–MEIER ESTIMATOR

BY JIA-GANG WANG

Fudan University

Let $\{X_n, n \geq 1\}$ be i.i.d. with $P(X_i \leq u) = F(u)$ and $\{U_n, n \geq 1\}$ be i.i.d. with $P(U_i \leq u) = G(u)$. $\hat{F}_n(t)$ is the Kaplan–Meier estimator based on the censored data $(\tilde{X}_i = X_i \wedge U_i, \delta_i = 1_{(X_i \leq U_i)}, 1 \leq i \leq n)$. In this note, it is shown that for $T_n = \max_{1 \leq i \leq n} \tilde{X}_i$,

$$\operatorname*{pr-lim}_{n \to \infty} \sup_{t \leq T_n} |\hat{F}_n(t) - F(t)| = 0.$$

Hence, the largest interval on which the Kaplan–Meier estimator is uniformly consistent is found.

Let $\{X_n, n \geq 1\}$ be a sequence of independent positive random variables with common distribution function F. Independent of the X_i's, let $\{U_i, i \geq 1\}$ be also a sequence of independent positive random variables with common distribution function G. The censored observations $(\tilde{X}_i, \delta_i)$ are defined by

$$\tilde{X}_i = X_i \wedge U_i, \qquad \delta_i = I\{X_i \leq U_i\},$$

where $\wedge$ denotes minimum and $I\{\cdot\}$ is the indicator random variable of the specified event. In survival analysis F is estimated by the product estimator $\hat{F}_n$, introduced by Kaplan and Meier (1958). Define processes N_n and Y_n on $[0, \infty[$ by

$$N_n(t) = \#\{i \leq n\colon \tilde{X}_i \leq t, \delta_i = 1\} = \#\{i \leq n\colon X_i \leq U_i \wedge t\},$$
$$Y_n(t) = \#\{i \leq n\colon \tilde{X}_i \geq t\} = \#\{i \leq n\colon X_i \wedge U_i \geq t\}.$$

Then the Kaplan–Meier estimator $\hat{F}_n$ is given by

$$\hat{F}_n(t) = 1 - \prod_{s \leq t \wedge T_n} \left\{1 - \frac{\Delta N_n(s)}{Y_n(s)}\right\},$$

where $\Delta N_n(s) = N_n(S) - N_n(s-)$ and $T_n = \max_{1 \leq i \leq n} \tilde{X}_i$.

Let H be the distribution function of $\tilde{X}_i$'s given by $H = 1 - (1 - F)(1 - G)$ and define (possibly infinite) τ_F, τ_G and τ_H by $\tau_F = \sup\{t\colon F(t) < 1\}$, etc.

Gill (1983) proved (with many earlier authors) that if F is continuous, then for each τ such that $H(\tau-) < 1$,

$$\operatorname*{pr-lim}_{n \to \infty} \sup_{t \leq \tau} |\hat{F}_n(t) - F(t)| = 0, \tag{1}$$

Received July 1986; revised January 1987.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62N05, 62P10.
Key words and phrases. Product-limit estimator, Kaplan–Meier estimator, random censoring, uniform consistency, martingales, stochastic integrals.

where pr-lim denotes limit in probability, and he also posed the question whether the supremum in (1) may be taken over $t \le T_n$. This note, motivated by Peterson (1977) and Gill (1983), is concerned with the uniform consistency of $\hat{F}_n$ for general distributions F, G. The main result is the following

THEOREM. *For possibly noncontinuous distributions F and G*

$$\text{pr-}\lim_{n\to\infty} \sup_{t<\tau_H} |\hat{F}_n(t) - F(t)| = 0, \tag{2}$$

$$\text{pr-}\lim_{n\to\infty} \sup_{t\le T_n} |\hat{F}_n(t) - F(t)| = 0, \tag{3}$$

where $T_n = \max_{i\le n} \tilde{X}_i$.

This theorem gives a positive answer to the question posed by Gill (1983). Following Gill (1980), for a (sub-) distribution function F on $[0, \infty[$, the cumulative hazard function Λ is defined by

$$\Lambda(t) = \int_{[0,t]} \frac{dF(s)}{1 - F(s-)}.$$

Thus

$$1 - F(t) = \mathscr{E}(-\Lambda)_t = \prod_{s\le t}(1 - \Delta\Lambda(s))\exp(-\Lambda^c(t)), \tag{4}$$

where Λ^c is the continuous part of Λ. For the cumulative hazard function Λ, its Nelson estimator $\hat{\Lambda}_n$ is defined by

$$\hat{\Lambda}_n(t) = \sum_{s\le t\wedge T_n} \frac{\Delta N(s)}{Y_n(s)} = \int_{[0,t\wedge T_n]} \frac{dN_n(s)}{Y_n(s)}$$

and the Kaplan–Meier estimator of the survival function $1 - F$ is

$$1 - \hat{F}_n(t) = \mathscr{E}(-\hat{\Lambda}_n)_t. \tag{5}$$

LEMMA. *For each $\tau \le \tau_H$, if $F(\tau-) < 1$, then*

$$\text{pr-}\lim_{n\to\infty} \sup_{t<\tau} |\hat{\Lambda}_n(t) - \Lambda(t)| = 0, \tag{6}$$

$$\text{pr-}\lim_{n\to\infty} \sup_{t<\tau} |\hat{F}_n(t) - F(t)| = 0. \tag{7}$$

PROOF. Let $M_n(t) = \hat{\Lambda}_n(t) - \Lambda(t \wedge T_n)$. Then $M_n(t) = M_n(t \wedge T_n)$, $M_n = \{M_n(t), t \ge 0\}$ is a square integrable martingale and its predictable quadratic variation is [cf. Gill (1980)]

$$\langle M_n, M_n\rangle(t) = \int_0^{t\wedge T_n} \frac{1 - \Delta\Lambda(s)}{Y_n(s)}\, d\Lambda(s).$$

Now we will show that

$$\lim_{n\to\infty} \langle M_n, M_n\rangle(\tau-) = 0 \quad \text{a.s.} \tag{8}$$

For given $\varepsilon > 0$, taking $\sigma < \tau$ such that $\Lambda(\tau -) - \Lambda(\sigma) < \varepsilon$, hence, also $H(\sigma) < 1$,

$$\begin{aligned}&\langle M_n, M_n\rangle(\tau -) - \langle M_n, M_n\rangle(\sigma)\\ &\quad = \int_{]\sigma,\tau[} 1_{[0,T_n]}(s)\frac{1 - \Delta\Lambda(s)}{Y_n(s)}\,d\Lambda(s) \le \Lambda(\tau -) - \Lambda(\sigma) < \varepsilon,\end{aligned}$$

and by the Glivenko–Cantelli theorem

$$\begin{aligned}\lim_{n\to\infty} n\langle M_n, M_n\rangle(\sigma) &= \lim_{n\to\infty}\int_0^\sigma \frac{1 - \Delta\Lambda(s)}{\frac{1}{n}Y_n(s)}\,d\Lambda(s)\\ &= \int_0^\sigma \frac{1 - \Delta\Lambda(s)}{1 - H(s -)}\,d\Lambda(s) \quad \text{a.s.}\end{aligned}$$

Therefore, $\lim_{n\to\infty}\langle M_n, M_n\rangle(\sigma) = 0$, a.s. and

$$\limsup_{n\to\infty}\langle M_n, M_n\rangle(\tau -) = \limsup_{n\to\infty}\big(\langle M_n, M_n\rangle(\tau -) - \langle M_n, M_n\rangle(\sigma)\big) \le \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, (8) is true.

Now using Lenglart's inequality (1977) we get

$$\text{pr-}\lim_{n\to\infty}\sup_{t<\tau,\, t\le T_n}|\hat{\Lambda}_n(t) - \Lambda(t)| \le \text{pr-}\lim_{n\to\infty}\sup_{t<\tau}|M_n(t)| = 0.$$

Since also $I(T_n < \tau)(\Lambda(\tau -) - \Lambda(T_n)) \to 0$ a.s. as $n \to \infty$, (6) is true.

For each subsequence $\{\mathscr{E}(-\hat{\Lambda}_{n'})\}$, in view of (6), there exists a subsequence $\{\mathscr{E}(-\hat{\Lambda}_{n''})\}$ such that

$$\lim_{n''\to\infty}\sup_{t<\tau}|\hat{\Lambda}_{n''}(t) - \Lambda(t)| = 0 \quad \text{a.s.}$$

Now using Lemma 2 in Gill (1981) we have

$$\lim_{n''\to\infty}\sup_{t<\tau}|F_{n''}(t) - F(t)| = 0 \quad \text{a.s.}$$

This means that (7) holds. □

Proof of the theorem. For the proof of (2), in view of the lemma it suffices to consider the case $F(\tau_H -) = 1$. For given $\varepsilon > 0$, choose $\tau < \tau_H$ such that $1 > F(\tau -) > 1 - \varepsilon$. Now for $t \in [\tau, \tau_H]$, we have

$$\hat{F}_n(\tau -) \le \hat{F}_n(t) \le 1, \qquad 1 - \varepsilon < F(\tau -) \le F(t) < 1.$$

Therefore,

$$\sup_{\tau\le t\le\tau_H}|\hat{F}_n(t) - F(t)| < \max\{\varepsilon, 1 - \hat{F}_n(\tau -)\}.$$

Combining this, (7) and the arbitrariness of $\varepsilon > 0$, (2) holds.

Using Lemma 2.8 in Gill (1983), it suffices to prove that (3) holds for $H(\tau_H -) = 1$, but in this case we know that $T_n = \max_{i\le n}\tilde{X}_i < \tau_H$. Hence (3) is a consequence of (2). □

From this theorem it is easy to derive the following corollaries.

COROLLARY 1. $\text{pr-lim}_{n\to\infty} \sup_{t\le\tau_H}|\hat{F}_n(t) - F(t)| = 0$ *if and only if* $G(\tau_H -) < 1$ *or* $\Delta F(\tau_H) = 0$.

COROLLARY 2. $\text{pr-lim}_{n\to\infty} \sup_{t<\infty}|\hat{F}_n(t) - F(t)| = 0$ *if and only if one of the following conditions holds*: (i) $\tau_F < \tau_G$; (ii) $\tau_F = \tau_G$ *and* $G(\tau_H -) < 1$; (iii) $\tau_F = \tau_G$ *and* $\Delta F(\tau_H) = 0$.

In a recent book, Shorack and Wellner (1986) discuss the a.s. version of the same problem. However, there are some difficulties with their proof and in using their method, and only the following result has actually been established.

PROPOSITION. *For possibly noncontinuous distributions F and G, if* $H(\tau_H -) < 1$ *or* $F(\tau_H -) = 1$, *then*

$$\lim_{n\to\infty} \sup_{t<\tau_H} |\hat{F}_n(t) - F(t)| = 0 \quad a.s.,$$

$$\lim_{n\to\infty} \sup_{t\le T_n} |\hat{F}_n(t) - F(t)| = 0 \quad a.s.$$

Therefore, whether the a.s. version of our theorem holds is still an open problem.

REFERENCES

GILL, R. D. (1980). Censoring and stochastic integrals. *Mathematical Centre Tracts* **124**. Mathematisch Centrum, Amsterdam.

GILL, R. D. (1981). Testing with replacement and the product limit estimator. *Ann. Statist.* **9** 853–860.

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

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

LENGLART, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **13** 171–179.

PETERSON. A. V. (1977). Expressing the Kaplan–Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* **72** 854–858.

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

INSTITUTE OF MATHEMATICS
FUDAN UNIVERSITY
SHANGHAI
THE PEOPLE'S REPUBLIC OF CHINA