The Annals of Statistics
1992, Vol. 20, No. 1, 535–544

SOME INEQUALITIES ABOUT THE KAPLAN–MEIER ESTIMATOR

BY SONG YANG

Texas Tech University

In this paper we consider the product-limit estimator of the survival distribution function in the context of independent but nonidentically distributed censoring times. An upper bound on the mean square increment of the stopped Kaplan–Meier process is obtained. Also, a representation is given for the ratio of the survival distribution function to the product-limit estimator as the product of a bounded process and a martingale. From this representation bounds on the mean square of the ratio and on the tail probability of the sup norm of the ratio are derived.

1. Introduction. The product-limit (PL) estimator of the survival distribution function due to Kaplan and Meier (1958) and its variants [e.g., Susarla and Van Ryzin (1980)] are often used in analyzing randomly censored data. While there are many results in the study of censored data, the approach by Gill (1980), via counting process and martingale techniques, has one advantage in that it allows a more successful treatment of the estimator's behavior on the entire support set of the underlying distribution function [e.g., Gill (1983) and Yang (1991)]. Very often some sort of Taylor expansion of the functional involved is used in large sample inferences and some bound on the mean square increment (MSI) of the PL process can be used in proving the negligibility of the higher order terms. In some other cases, the ratio of the survival distribution function to the PL estimator needs to be analyzed [Koul, Susarla and Van Ryzin (1981), Blum and Susarla (1980)]. First of all, in Section 2 of this paper we generalize a result on the MSI of the Kaplan–Meier process due to Yang (1991). Then in Section 3 we give a representation of the above-mentioned ratio as the product of a bounded process and a martingale, from which bounds on the mean square of the ratio and on the tail probability of the sup norm of the ratio are derived. These results are obtained by solving certain inhomogeneous Volterra integral equations [Volterra (1887)] and Gronwall inequalities [Gronwall (1919)], which are derived from a representation result of Gill (1980); see also Gill and Johansen (1989) for a survey of relevant results.

2. The mean square increment. Let $(U_1, C_1), \dots, (U_n, C_n)$ be n pairs of independently distributed nonnegative random variables, where $U_1, \dots, U_n$ have a common cumulative distribution function (c.d.f.) F and for each $i = 1, \dots, n$, C_i is independent of U_i and has (sub) c.d.f. G_i. The observable

Received March 1990; revised June 1991.

AMS 1980 *subject classifications*. Primary 62E20, 62G99; secondary 62M99.

Key words and phrases. Product-limit estimator, martingale, Volterra integral equation, Gronwall's inequality.

data consist of $(U_i \wedge C_i, [U_i \le C_i])$, $i = 1, \dots, n$, where $\wedge$ denotes the minimum and $[A]$ is the indicator function of a set A, and one is interested in the survival distribution function $1 - F$ of the life times $U_1, \dots, U_n$. Let $X_i = U_i \wedge C_i$, $\delta_i = [U_i \le C_i]$ for $i = 1, \dots, n$ and define

$$H_n(t) = n^{-1}\#\{i: X_i \le t\}, \qquad 0 \le t < \infty,$$

$$H_n^1(t) = n^{-1}\#\{i: X_i \le t, \delta_i = 1\}, \qquad 0 \le t < \infty.$$

The PL estimator $1 - F_n$ of $1 - F$ is given by

$$(1) \qquad 1 - F_n(t) = \prod_{s \le t}\left(1 - \frac{\Delta H_n^1(s)}{\overline{H}_{n-}(s)}\right), \qquad 0 \le t < \infty,$$

where $\Delta f(x) = f(x_+) - f(x_-)$ for any function f and $\overline{D} = 1 - D$ for any (sub) c.d.f. D.

Denote $1 - n^{-1}\sum_{i=1}^n \overline{G}_i$ by G and $EH_n = 1 - \overline{F}\overline{G}$ by H. Define $\tau_H = \sup\{t: H(t) < 1\}$ and so on and $\Lambda(t) = \int_0^t (1/\overline{F}_-)\, dF$, $0 \le t < \tau_F$, where as in the sequel $\int_0^t = \int_{[0,t]}$ and $\int_s^t = \int_{(s,t]}$ for $s > 0$. Define the Kaplan–Meier (KM) process Z_n by

$$Z_n = n^{1/2}(F_n - F)/\overline{F} \quad \text{on } [0, \tau_F),$$

and the stopped KM processes Z_{nk} by $Z_{nk}(t) = Z_n(t \wedge X_{(n-k)})$, $k = 0, 1, \dots, n-1$, where $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ are the order statistics of the X_i's. Also let $J_{nk}(t) = [t \le X_{(n-k)}]$, $0 \le t < \infty$. Now we are ready to state our result on the mean square increment bound.

THEOREM 1. *Suppose $\Delta F(\tau_F) = 0$ and $F(0) = G_i(0) = 0$, $i = 1, \dots, n$. Let r be a positive real. Then when either $k \le r$ or F is continuous, we have for $0 \le s < t < \tau_H$ that*

$$(2) \qquad E(Z_{nk}(t) - Z_{nk}(s))^2 \le D(t) - D(s),$$

where

$$D(x) = \{2(1 + 1/r)\}/\overline{F}^{(1+r)/(k+1)}(x)\int_0^x \overline{F}^{(1+r)/(k+1)}/(\overline{H}_-\overline{F})\, dF.$$

REMARK. Although the above result is for general k, the cases $k = 0, 1$ will usually suffice in application.

Consider the case when the G_i's are identical and F is continuous. In this particular case, the above result for $k = 0$ and $r = 1$ is obtained in Lemma 2.5 of Yang (1991). The result for $k = 0$, $0 < r < 1$, improves the bound given in that lemma and the result for $k = 1$, $0 < r < 1$, gives a sharp bound. In order to see this, consider the (uncensored) case when $G_i \equiv 0$, $i = 1, \dots, n$. Now F_n reduces to the empirical distribution function and a direct calculation gives

$EZ_n^2 = F/\bar{F}$, while (2) gives, for $0 < r < 1$,

$$EZ_{n0}^2 \le \{2(1+r)/r^2\}(1-\bar{F}^r)/\bar{F}^{1+r} \le \{2(1+r)/r^2\}F/\bar{F}^{1+r},$$

with an extra explosive factor $1/\bar{F}^r$ and

$$\begin{aligned} EZ_{n1}^2 &\le \{4(1+r)/(r(1-r))\}(\bar{F}^{(r-1)/2}-1)/\bar{F}^{(r+1)/2} \\ &\le \{4(1+r)/(r(1-r))\}F/\bar{F}. \end{aligned}$$

Notice that the sharp bound is obtained at the expense of discarding only one observation. For applications of Theorem 1, see the discussion after the proof of Theorem 1.

PROOF OF THEOREM 1. By the assumption we have $X_{(n)} \le \max U_i < \tau_F$ a.s. Hence from Lemma 3.2.1(iv) of Gill (1980) we obtain that, for $k = 0, 1, \dots, n-1$, Z_{nk} is a locally square integrable mean 0 martingale on $[0, \infty)$, with predictable variation process

$$\begin{aligned} \langle Z_{nk}\rangle(t) &= \int_0^t \left(\bar{F}_{n-}/\bar{F}\right)^2 J_{nk}/\bar{H}_{n-}(1-\Delta\Lambda)\,d\Lambda \\ &= \int_0^t \left(1 - n^{-1/2}Z_{nk-}\right)^2 J_{nk}/\left(\bar{H}_{n-}\bar{F}\right)dF. \end{aligned} \tag{3}$$

Note that $[X_i \ge t]$, $i = 1, \dots, n$, are independent Bernoulli random variables. Hence, by Hoeffding (1956), we have that

$$Ef\left(n\bar{H}_{n-}\right) \le \sum_{i=0}^{n} f(i)\binom{n}{i}\bar{H}_-^i H_-^{n-i}$$

for any function f such that

$$f(i+2) - 2f(i+1) + f(i) > 0, \qquad i = 0, 1, \dots, n-2.$$

Taking $f(i) = 1/i[i > 0] + 2[i = 0]$ and using $n/i \le 2(n+1)/(i+1)$, we get, on $[0, \tau_H)$,

$$\begin{aligned} E\left(J_{nk}/\bar{H}_{n-}\right) &\le nEf\left(n\bar{H}_{n-}\right) \\ &\le \sum_{i=1}^{n} n/i\binom{n}{i}\bar{H}_-^i H_-^{n-i} + 2nH_-^n \\ &\le 2\sum_{i=1}^{n}\binom{n+1}{i+1}\bar{H}_-^i H_-^{n-i} + 2nH_-^n \\ &\le 2/\bar{H}_- \sum_{i=1}^{n+1}\binom{n+1}{i}\bar{H}_-^i H_-^{n+1-i} \le 2/\bar{H}_-. \end{aligned} \tag{4}$$

Now denote $EZ_{nk}^2(t) = E\langle Z_{nk}\rangle(t)$ by $L_k(t)$. Notice that $EZ_{nk-}^2 \le L_{k-}$ by

Fatou's lemma and $J_{nk}/(n\overline{H}_{n-}) \le J_{nk}/(k+1)$, thus on $[0, \tau_H)$,

$$(5) \qquad E\frac{Z_{nk-}J_{nk}}{\sqrt{n}\,\overline{H}_{n-}} \le \sqrt{EZ^2_{nk-}E\left(J_{nk}\left(n\overline{H}^2_{n-}\right)^{-1}\right)} \le \sqrt{\frac{2L_{k-}}{(k+1)\overline{H}_-}}.$$

Using (4), (5) and expanding the square in (3), we get, for $0 < t < \tau_H$ and positive real r,

$$(6) \qquad \begin{aligned} L_k(t) &\le \int_0^t \alpha\, d\beta + 2\int_0^t \sqrt{\frac{\alpha L_{k-}}{k+1}}\, d\beta + \frac{1}{k+1}\int_0^t L_{k-}\, d\beta \\ &\le \left(1 + \frac{1}{r}\right)\int_0^t \alpha\, d\beta + \frac{1+r}{k+1}\int_0^t L_{k-}\, d\beta, \end{aligned}$$

where $\alpha = 2/\overline{H}_-$, $d\beta = \overline{F}^{-1} dF$. To obtain the last inequality we have used

$$(7) \qquad 2AB = 2\frac{A}{\sqrt{r}}\sqrt{r}\,B \le \frac{1}{r}A^2 + rB^2.$$

Similarly we have, for $s < t < \tau_H$,

$$(8) \qquad \begin{aligned} E\left(Z^2_{nk}(t) - Z^2_{nk}(s)\right) &= E(\langle Z_{nk}\rangle(t) - \langle Z_{nk}\rangle(s)) \\ &\le \left(1 + \frac{1}{r}\right)\int_s^t \alpha\, d\beta + \frac{1+r}{k+1}\int_s^t L_{k-}\, d\beta. \end{aligned}$$

From (6) it is easily seen, using Theorem 10 of Gill and Johansen (1989), that L_k can be bounded above by the solution of the corresponding integral equation. This solution can be expressed in terms of the product integral of $(1 + r)/(k + 1)\, d\beta$. Here we will use an integration by parts formula to give the bound an explicit integral form. Without loss of generality, we will assume r to be a rational. The result for a general positive real r can be obtained by passing to the limit for a sequence of rationals r_i decreasing to r.

Since r is a positive rational, there exist positive integers p, q such that $p/q = (1 + r)/(k + 1)$. Let Ψ satisfy

$$(9) \qquad \Psi(t) = \phi(t) - p/q\int_0^t \Psi_-/\overline{B}^q\, d\overline{B}^q, \qquad 0 \le t < \tau_B,$$

where $\phi(t) = (1 + r)/r\int_0^t \alpha\, d\beta$ and B is a c.d.f. given by $\overline{B}^q = \overline{F}$. Then $L_k \le \Psi$. This and (8), (9) give, for $0 \le s < t < \tau_B$,

$$(10) \qquad L_k(t) - L_k(s) \le \Psi(t) - \Psi(s).$$

Integrating by parts [Shorack and Wellner (1986), page 868] and using (9) we have

$$(11) \qquad \begin{aligned} \Psi(t)\overline{B}^p(t) &= \int_0^t d(\Psi\overline{B}^p) = \int_0^t \left(\Psi_-\, d\overline{B}^p + \overline{B}^p\, d\Psi\right) \\ &= \int_0^t \left(\Psi_-\, d\overline{B}^p - \frac{p}{q}\Psi_-\overline{B}^{p-q}\, d\overline{B}^q\right) + \int_0^t \overline{B}^p\, d\phi. \end{aligned}$$

Now using [Shorack and Wellner (1986), page 868]

$$d\overline{B}^p = \left(\sum_{i=0}^{p-1} \overline{B}_-^i \overline{B}^{p-i-1}\right) d\overline{B}, \tag{12}$$

we can rewrite the first integral on the RHS of (11) as

$$\int_0^t p\Psi_- \left(\frac{1}{p}\sum_{i=0}^{p-1} \overline{B}_-^i \overline{B}^{p-i-1} - \frac{1}{q}\sum_{i=0}^{q-1} \overline{B}_-^i \overline{B}^{p-i-1}\right) d\overline{B} = \int_0^t A\, dB,$$

say. Since $\overline{B}_- \geq \overline{B}$, $\overline{B}_-^i \overline{B}^{p-i-1}$ is nondecreasing in i and so their average is nondecreasing as the function of the number of summands. Hence we get $A \leq 0$ if $p \geq q$ and $A \geq 0$ if $p < q$; also apparently $A = 0$ if B is continuous. Thus from (11), we have for $0 \leq s < t < \tau_B$,

$$\begin{aligned} &\Psi(t) - \Psi(s) \\ &\quad = \overline{B}^{-p}(t)\int_s^t A\, dB + \left(\overline{B}^{-p}(t) - \overline{B}^{-p}(s)\right)\int_0^s A\, dB + D(t) - D(s) \\ &\quad \leq D(t) - D(s), \end{aligned} \tag{13}$$

when either $p \geq q$ or B is continuous. Now our stated result follows from (10) and (13). □

REMARK 1. *Applications.* In nonparametric inference, some sort of Taylor expansion, that is, the δ-method, is often used. In such a situation, Theorem 1 can be used in dealing with the second order term. It can also be used in a L^2 space setup, for example, in the Cramér–von Mises type minimum distance estimation, and in dealing with the variation process in a martingale framework. Specifically, we mention two applications of Theorem 1 below. References are given for the details. These are illustrative examples and we do not strive for the most general results here.

(i) *The integrated square error of a kernel density estimator*. Suppose F has a density f with respect to the Lebesgue measure λ. Define a kernel density estimator of f by

$$f_n(x) = \frac{1}{a_n}\int_{-\infty}^{\infty} K\left(\frac{x-y}{a_n}\right) dF_n(y), \tag{14}$$

where $a_n > 0$ is called the smoothing parameter and K is a kernel function. To measure the global performance of f_n, one often uses its integrated square error

$$\mathrm{ISE}(f_n) = \int_{-\infty}^{\infty} (f_n - f)^2 w\, d\lambda,$$

where w is a weight function. The asymptotic property of ISE(f_n) is closely related to the limiting behavior of the smoothing parameter a_n. Let A be a left continuous nonrandom function. The proof of Theorem 1 shows that, for any

$r > 0$, there exists a constant $k_r > 0$ such that for any $t < \tau_H$,

$$E\left(\int_{-\infty}^{t} A\,dZ_{n0}\right)^2 = E\int_{-\infty}^{t} A^2\,d\langle Z_{n0}\rangle \le k_r \int_{-\infty}^{t} A^2\left(\overline{F}^{2+r}\overline{G}\right)^{-1} dF,$$

provided that the last integral is finite. These types of inequalities will be exploited in studying the ISE(f_n) in a separate paper.

(ii) *Hellinger-differentiable functionals.* Let Ψ be a functional of the density. Suppose that Ψ is Hellinger-differentiable, with derivative ψ. That is,

$$\Psi(d_n) - \Psi(d) = \int_{-\infty}^{\infty} \psi\left(d_n^{1/2} - d^{1/2}\right) d\lambda + o\left(\left\| d_n^{1/2} - d^{1/2}\right\|\right),$$

as $\|d_n^{1/2} - d^{1/2}\| \to 0$, where $\|\cdot\|$ denotes the $L^2(\lambda)$ norm. One natural estimator of $\Psi(f)$ is $\Psi(f_n)$, where f_n is defined in (14). From the result in Theorem 1 for $k = 0$ and $r = 1$, the asymptotic normality of $\sqrt{n}\,(\Psi(f_n) - \Psi(f))$ can be established under the conditions $\sqrt{n}\,a_n^2 \to 0$, $\sqrt{n}\,a_n^{1+\varepsilon} \to \infty$ for some $\varepsilon > 0$ and some smoothness and boundedness conditions on ψ and f. This follows from the proofs of Lemma 4.1 and Theorem 4.1 in Yang (1991) (note that in that paper the functional involved is a little more complex, due to the estimation of the common unknown censoring distribution and some restriction on the integration range). Now if we truncate F_n at $X_{(n-1)}$ in the definition of f_n in (14), then, by the result in Theorem 1 for $k = 1$, $r < 1$, the conditions on the smoothing parameter can be weakened to those used in the i.i.d. complete data case: $\sqrt{n}\,a_n^2 \to 0$, $\sqrt{n}\,a_n \to \infty$. For details see the proof of Lemma 4.1 in Yang (1991).

3. The ratio of the survival distribution function to the PL estimator. With the same notation as in Section 2, let

$$R_n(t) = \frac{1 - F(t)}{1 - F_n(t)}, \qquad 0 \le t < X_{(n)}, \tag{15}$$

$$\Lambda_n(t) = \int_0^t \overline{H}_{n-}^{-1}\,dH_n^1, \qquad 0 \le t \le X_{(n)}. \tag{16}$$

Again we will consider the stopped versions $R_{nk}(\cdot) = R_n(\cdot \wedge X_{(n-k)})$, $\Lambda_{nk}(\cdot) = \Lambda_n(\cdot \wedge X_{(n-k)})$ for $k = 0, 1, \dots, n-1$. Define

$$S_{nk}(t) = \sum_{s \le t} \left(\Delta\Lambda_{nk}(s)\right)^2, \qquad 0 \le t < \infty, \tag{17}$$

$$A_{nk}(t) = 1 - \prod_{s \le t} \left(1 - \Delta S_{nk}(s)\right), \qquad 0 \le t < \infty, \tag{18}$$

that is, $\overline{A}_{nk}$ is the product integral of $-S_{nk}$ [Gill and Johansen (1989)]. Note that A_{nk} is a (sub) c.d.f. in t and $A_{nk}(t)$ is measurable with respect to $\sigma\{[X_i \le t], \delta_i[X_i \le t], X_i[X_i \le t], i = 1, \dots, n\}$. Also $dS_{nk} = \overline{A}_{nk-}^{-1}\,dA_{nk}$. The following result represents R_{nk} as the product of a bounded process and a martingale.

THEOREM 2. *Suppose F is continuous and $F(0) = G_i(0) = 0$, $i = 1, \dots, n$. Then for $k = 0, 1, \dots, n-1$ and $0 \le t < X_{(n)}$,*

$$R_{nk}(t) = \bar{A}_{nk}^{-1}(t) M_{nk}(t), \tag{19}$$

where

$$M_{nk}(t) = 1 + \int_0^t \bar{A}_{nk-} R_{nk-} J_{nk}\, d(\Lambda_n - \Lambda). \tag{20}$$

PROOF. Noting that $\Delta\Lambda_{nk} = J_{nk} \bar{H}_{n-}^{-1} \Delta H_n^1 \le (k+1)^{-1}$, with equality only possible at $X_{(n)}$, we can use Proposition A.4.1 of Gill (1980) in order to show that, for $0 \le t < X_{(n)}$,

$$\begin{aligned} R_{nk}(t) &= 1 - \int_0^t J_{nk} R_{nk-} (1 - \Delta\Lambda_n)^{-1} d(\Lambda - \Lambda_n) \\ &= 1 - \int_0^t J_{nk} R_{nk-}\, d(\Lambda - \Lambda_n) \\ &\quad + \int_0^t J_{nk} R_{nk-} \Delta\Lambda_n (1 - \Delta\Lambda_n)^{-1} d(\Lambda_n - \Lambda) \\ &= N_n(t) + V_n(t), \end{aligned} \tag{21}$$

say. Since F is continuous and $\Delta\Lambda_n = \bar{F}_{n-}^{-1} \Delta F_n$ (from the definition of F_n), we have $J_{nk} R_{nk-} (1 - \Delta\Lambda_n)^{-1} = J_{nk} R_{nk}$ and therefore $V_n(t)$ is simplified to be $\sum_{s \le t} R_{nk}(s)(\Delta\Lambda_{nk}(s))^2 = \int_0^t R_{nk}\, dS_{nk}$. Thus from (18) and (21), we have

$$R_{nk}(t) = N_n(t) + \int_0^t R_{nk} \bar{A}_{nk-}^{-1}\, dA_{nk}. \tag{22}$$

Integrating by parts yields

$$R_{nk} \bar{A}_{nk}(t) = \int_0^t d\big(R_{nk} \bar{A}_{nk}\big) + 1 = 1 + \int_0^t R_{nk}\, d\bar{A}_{nk} + \int_0^t \bar{A}_{nk-}\, dR_{nk}.$$

Now using (22) in the last integral we obtain

$$R_{nk} \bar{A}_{nk}(t) = 1 + \int_0^t \bar{A}_{nk-}\, dN_n = M_{nk}(t),$$

hence (19) follows. □

Let

$$c_k = \sum_{i=k+1}^{\infty} \frac{1}{i^2}, \qquad p_k = \exp(2c_k), \qquad k = 0, 1, \dots. \tag{23}$$

Two consequences of the representation (19) are given in the following corollary. Let $\vee$ denote the maximum.

COROLLARY 1. *Suppose the same conditions as in Theorem 2 hold. Then for $k = 0, 1, \dots, n-1$, and $\lambda > 0$,*

$$P[\| R_{n-} \|_k > \lambda] \le \frac{p_{k \vee 1}}{\lambda}, \tag{24}$$

where $\|\cdot\|_k$ *denotes the* sup norm *on* $[0, X_{(n-k)}]$ *and for* $k = 1, \dots, n-1$,

$$ER^2_{nk-}(t) \le p^2_k / \bar{F}^{p^2_k/(k+1)}(t), \qquad 0 \le t < \tau_F. \tag{25}$$

PROOF. Note that $\Delta\Lambda_{nk} = J_{nk}\bar{H}^{-1}_{n-}\Delta H^1_n \le J_{nk}(i+1)^{-1}$ at $X_{(n-1)}$, $i = 0, 1, \dots, n-1$. Hence from the inequality $\ln(1-x) \ge -2x$ for $0 \le x \le 1/2$, we have, for $0 \le t < X_{(n)}$ and $k = 0, 1, \dots, n-1$,

$$\begin{aligned} \bar{A}^{-1}_{nk}(t) &= \prod_{s \le t} (1 - \Delta S_{nk}(s))^{-1} \\ &\le \exp\left\{2 \sum_{s \le t} \Delta S_{nk}(s)\right\} \le e^{2c_{k \vee 1}} = p_{k \vee 1}. \end{aligned} \tag{26}$$

By applying Doob's inequality to the positive martingale M_{nk} on $[0, t]$, we obtain, for $k = 0, \dots, n-1$ and any $0 \le t < \infty$,

$$P\left[\sup_{[0,t]} M_{nk} \ge \lambda\right] \le \frac{1}{\lambda} EM_{nk}(t) = \frac{1}{\lambda}.$$

Hence letting $t \uparrow \infty$, we have $P[\sup M_{nk} \ge \lambda] \le (1/\lambda)$. Further, using (19) and (26), we obtain that $\|R_{n-}\|_k \le p_{k \vee 1} \sup M_{nk}$. Hence (24) follows. As for (25), first note that $X_{(n-1)} < X_{(n)}$ a.s. and therefore $ER^2_{nk-}(t)$ is well defined for $k = 1, \dots, n-1$. The martingale $M_{nk}(t)$ has a nondecreasing predictable variation process $\langle M_{nk}\rangle(t) = 1 + \int_0^t \bar{A}^2_{nk-} R^2_{nk-} J_{nk} (n\bar{H}_{n-})^{-1}(1 - \Delta\Lambda)\, d\Lambda$, hence by Fatou's lemma, $EM^2_{nk-}(t) \le \lim_{s \uparrow t} EM^2_{nk}(s) \le E < M_{nk} > (t)$. Now denote $ER^2_{nk-}(t)$ by $L(t)$. Then from (19), we have

$$L(t) \le p^2_k EM^2_{nk-}(t) \le p^2_k E\langle M_{nk}\rangle(t) \le p^2_k\left\{1 + \int_0^t L/(k+1)\, d\Lambda\right\}. \tag{27}$$

To obtain the last inequality we have used the fact that $\bar{A}_{nk} \le 1$, $nJ_{n-1}\bar{H}_{n-} \ge J_{nk}(k+1)$ and $\Delta\Lambda = 0$. Starting from (27), we can proceed as in the proof of Theorem 1 and finally obtain that

$$ER^2_{nk-}(t) = L(t) \le p^2_k / \bar{F}^{p^2_k/(k+1)}(t). \tag{28}$$

□

REMARK 2. (i) Gill's inequality [Shorack and Wellner (1986), page 317], dealing with R_{n0-} for general F but identical G_i's, has a bound $3/\sqrt{\lambda}$ for the tail probability of $\|R_{n-}\|_0$, while (24) for $k = 0$ gives the bound $p_1/\lambda = e^{\pi^2/3-2}/\lambda < 3.64/\lambda$. Discarding some large observations only helps reducing the constant multiple, not the rate, of our bound. For the mean square of the ratio R_{nk}, note that we could have obtained $ER^2_n((t \wedge X_{(n)})_-) \le p^2_1/\bar{F}^{p^2_1}(t)$ in Corollary 1, with the power $p^2_1 < 13.3$. If we discard $X_{(n)}$, then the power in (25) becomes $p^2_1/2 < 6.7$. In general, $p^2_k/(k+1) \to 0$ as $k \to \infty$.

(ii) By the symmetry of the censoring problem in U_i's and C_i's, results parallel to (24) and (25) for different lifetime c.d.f. F_i's and common censoring c.d.f. G also hold. Those results can be used in studying the censored linear regression problem considered by Koul, Susarla and Van Ryzin (1981). In that problem, one observes

$$T_i \wedge Y_i, \qquad [T_i < Y_i], \qquad i = 1, \ldots, n,$$

where the censoring times Y_i's are i.i.d. and independent of T_i's and for $1 \le i \le n$,

$$T_i = \alpha + \beta x_i + \varepsilon_i,$$

with known x_i's and i.i.d. zero mean ε_i's. Koul, Susarla and Van Ryzin (1981) suggested estimating α, β, respectively, by

$$\hat{\alpha} = \sum a_{ni}\delta_i Z_i \overline{G}_n(Z_i)[Z_i \le M_n], \qquad \hat{\beta} = \sum b_{ni}\delta_i Z_i \overline{G}_n(Z_i)[Z_i \le M_n],$$

where $b_{ni} = (x_i - \bar{x})/\Sigma(x_i - \bar{x})^2$, $a_{ni} = 1/n - \bar{x}b_{ni}$, $\bar{x} = 1/n\Sigma x_i$, G_n is the Kaplan–Meier estimator for the censoring distribution (defined similarly to F_n with $\delta_i = 1$ replaced by $\delta_i = 0$) and the nonrandom quantity $M_n \to \infty$ at a proper rate so that various negligibility conditions are satisfied. Thus the ratio $\overline{G}_-/\overline{G}_{n-}$ is naturally involved. Results parallel to our (24), (25) can be used to relax and simplify some conditions in that paper; see Zhou (1989) for a related argument in detail.

(iii) After the submission of this paper, I learned of an alternative bound for the tail probability of $\|R_{n-}\|_0$, given by Zhou (1990). He considered $\|\Lambda - \Lambda_n\|_0$ first and used a delicate truncation on the integrating interval. Then from the exponential identity he obtained a bound $C(\ln(\lambda/5))^{-2/3}$, where C can be taken to be 358 and λ is restricted to be $> 5e^2$. Our result here seems simpler, with a more straightforward proof.

(iv) When all censoring distributions are the same: $G_i = G$, $i = 1, \ldots, n$, a kernel density estimator alternative to that in (14) is sometimes used. It is defined as [cf. Blum and Susarla (1980)]

$$f_n(x) = \frac{1}{na_n\overline{G}_{n-}(x)} \sum_{i=1}^{n} \delta_i K\left(\frac{x - X_i}{a_n}\right),$$

where G_n is the Kaplan–Meier estimator for G. Thus the ratio $\overline{G}_-/\overline{G}_{n-}$ is also involved. We expect that again (24), (25) can be useful, but will not dwell on it here.

Acknowledgments. I would like to thank Professor Hira L. Koul for some stimulating discussions which led to my consideration of the problem on the ratio of the survival distribution function to the PL estimator. Thanks are also due to Professor Lawrence Schovanec for suggesting the use of inequality (7) and to the referees for valuable comments which led to this improved version of the manuscript.

REFERENCES

BLUM, J. R. and SUSARLA, V. (1980). Maximal derivation theory of density and failure estimates based on censored data. In *Multivariate Analysis V* (P. R. Krishnaiah, ed.) 213–222. North-Holland, Amsterdam.

GILL, R. (1980). *Censoring and Stochastic Integrals. Math. Centre Tract* **124**. Math. Centrum, Amsterdam.

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

GILL, R. and JOHANSEN, S. (1990). A survey of product-integration with a view towards application in survival analysis. *Ann. Statist.* **18** 1501–1555.

GRONWALL, T. (1919). Notes on the derivatives with respect to a parameter of solutions of a system of differential equations. *Ann. Math.* **20** 292–296.

HOEFFDING, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27** 713–721.

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

KOUL, H., SUSARLA, V. and VAN RYZIN, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9** 1276–1288.

KOUL, H. and YANG, S. (1989). Minimum distance estimation in two sample scale problem: Uncensored and censored data. In *Recent Developments in Statistics and Their Applications* (J. P. Klein and J. C. Lee, eds.) 117–134. Freedom Academy, Seoul.

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

SUSARLA, V. and VAN RYZIN, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. *Ann. Statist.* **8** 1002–1016.

VOLTERRA, V. (1987). Sulle equazione differnezeniali lineari. *Rendiconti Accademia dei Lincei* **3** 393–396.

YANG, S. (1991). Minimum Hellinger distance estimation of parameter in the random censorship model. *Ann. Statist.* **19** 579–602

ZHOU, M. (1989). A new proof of CLT for the Koul–Susarla–van Ryzin estimator. Mimeo Series 1770, Dept. Statistics, Univ. North Carolina.

ZHOU, M. (1990). Some properties of the Kaplan–Meier estimator for independent nonidentically distributed random variables. Technical report, Dept. Statistics, Univ. Kentucky.

DEPARTMENT OF MATHEMATICS
TEXAS TECH UNIVERSITY
MAIL STOP 1042
LUBBOCK, TEXAS 79409-1042