The Annals of Statistics
1985, Vol. 13, No. 1, 150–162

A HEAVY CENSORING LIMIT THEOREM FOR THE PRODUCT LIMIT ESTIMATOR

BY JON A. WELLNER

University of Washington

A key identity for the product-limit estimator due to Aalen and Johansen (1978) and Gill (1980) is shown to be a consequence of the exponential formula of Doleans-Dade (1970). The basic counting processes in the censored data problem are shown to converge jointly to Poisson processes under "heavy-censoring": $G_n \to_d \delta_0$, but $n(1 - G_n) \to \alpha$ where G_n is the censoring distribution. The Poisson limit theorem for counting processes implies Poisson type limit theorems under heavy censoring for the cumulative hazard function estimator and product limit estimator. The latter, in combination with the key identity of Aalen-Johansen and Gill and martingale properties of the limit processes, yields a new approximate variance formula for the product limit estimator which is compared numerically with recent finite sample calculations for the case of proportional hazard censoring due to Chen, Hollander, and Langberg (1982).

1. Introduction. Let (X_i, Y_i), $i = 1, \cdots, n$ be independent identically distributed pairs of positive random variables with distribution functions (df's) $F(t) = P(X_i \le t)$ and $G(t) = P(Y_i \le t)$. Suppose that X_i and Y_i are independent for each i. The X's represent "survival times" and the Y's represent "censoring times", but only (Z_i, δ_i), $i = 1, \cdots, n$ with

$$Z_i \equiv \min\{X_i, Y_i\}, \quad \delta_i = \begin{cases} 1 & X_i \le Y_i \\ 0 & X_i > Y_i \end{cases} \tag{1.1}$$

are observed. This is the "random-censorship" model.

For $0 \le t < \infty$ define the counting processes

$$\begin{aligned} &\mathbb{N}_n^u(t) \equiv \textstyle\sum_{i=1}^n 1_{[Z_i \le t]}\delta_i, \quad \mathbb{N}_n^c(t) \equiv \textstyle\sum_{i=1}^n 1_{[Z_i \le t]}(1 - \delta_i), \\ &\mathbb{N}_n(t) \equiv \mathbb{N}_n^u(t) + \mathbb{N}_n^c(t), \end{aligned} \tag{1.2}$$

and, with $f_-(t) \equiv f(t-) \equiv \lim_{s \uparrow t} f(s)$, set

$$\mathbb{R}_n(t) \equiv n - \mathbb{N}_n(t-) = \textstyle\sum_{i=1}^n 1_{[Z_i \ge t]}, \tag{1.3}$$

the number "at risk" at time t. In terms of these counting processes, the cumulative hazard function estimator $\hat{\mathbf{\Lambda}}_n$ of $\Lambda = \int_{(0,\cdot]} dF/(1 - F_-)$ and Kaplan-Meier product-limit estimator $\hat{\mathbb{S}}_n$ of $S \equiv 1 - F$ are defined by

$$\hat{\mathbf{\Lambda}}_n(t) = \int_{(0,t]} \frac{1}{\mathbb{R}_n}\, d\,\mathbb{N}_n^u, \quad 0 \le t < \infty, \tag{1.4}$$

Received June 1983; revised September 1984.

AMS 1980 *subject classifications*. Primary 62G05, 60F05; secondary 62G30, 60G44.

Key words and phrases. Cumulative hazard function estimator, exponential formula, Kaplan–Meier estimator, martingales, Poisson process, small sample moments, variance formulas.

(with $0/0 \equiv 0$) and

$$(1.5) \qquad \hat{\mathbb{S}}_n(t) = \prod_{s \le t} (1 - \Delta \hat{\mathbf{\Lambda}}_n(s)) = \prod_{i: Z_{(i)} \le t} (1 - \delta_{(i)}/(n - i + 1))$$

where $0 \le Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)} \equiv T_n$ denote the ordered Z's and $\delta_{(1)}, \cdots, \delta_{(n)}$ are the corresponding δ's; see e.g. Gill (1980), (1983). Note that this version of the product-limit estimator is strictly positive on $[Z_{(n)}, \infty)$ if the largest observation is censored, $\delta_{(n)} = 0$, and hence differs slightly from the version $\bar{\mathbb{F}}_n(t) \equiv \hat{\mathbb{S}}_n(t) 1_{[0, Z_{(n)})}(t)$ which always equals 0 on $[Z_{(n)}, \infty)$ whether $\delta_{(n)} = 0$ or 1; see e.g. Chen, Hollander, and Langberg (1982) and other references given there. We present small sample numerical evidence in favor of $\hat{\mathbb{S}}_n(t)$ in Section 4.

Our purposes here are three-fold: First, in Section 2 we use the "exponential formula" of Doleans-Dade (1970) to give another proof of the fundamental identity (for $0 < t < \tau_F \equiv \inf\{t: F(t) = 1\}$)

$$(1.6) \qquad \frac{\hat{\mathbb{S}}_n(t)}{S(t)} = 1 - \int_{(0,t]} \frac{\mathbb{S}_{n-}}{S} \, d(\hat{\mathbf{\Lambda}}_n - \Lambda)$$

given in Theorem 3.1 of Aalen and Johansen (1978), (3.2.12) in Gill (1980), and Lemma 2.4 in Gill (1983). While the proof of (1.6) given in Section 2 uses only deterministic special cases of the exponential formula, all of which can be proved via integration by parts (as in Chapter 18 of Liptser and Shiryayev, 1978; Gill, 1980; and Appendix 4 of Bremaud, 1981), we have included the present approach to these results in hopes that the general exponential formula will become more widely known among statisticians.

Second, in Section 3 we present "heavy-censoring" limit theorems for $(\mathbb{N}_n^u, \mathbb{R}_n)$, $\hat{\mathbf{\Lambda}}_n$, and $\hat{\mathbb{S}}_n$ under the assumption that $G = G_n \to_d \delta_0 \equiv$ point mass at 0, but $n(1 - G_n(t)) \to a(t)$ for $0 < t < \infty$. Under these assumptions on G, the processes $(\mathbb{N}_n^u, \mathbb{R}_n)$ converge jointly to a pair of dependent Poisson processes $(\mathbb{N}^u, \mathbb{R})$ where $\mathbb{N}^u$ is a Poisson process with (increasing) mean function $A = \int_0^{\cdot} a \, dF$ and $\mathbb{R}$ is a Poisson process with (decreasing) mean function $B = a(1 - F)$. The basic limit theorem for $(\mathbb{N}_n^u, \mathbb{R}_n)$ combined with (1.4) and (1.5) then yields simple limit processes under heavy censoring for $\hat{\mathbf{\Lambda}}_n$ and $\hat{\mathbb{S}}_n$ represented in terms of the joint Poisson limit process $(\mathbb{N}^u, \mathbb{R})$ of $(\mathbb{N}_n^u, \mathbb{R}_n)$.

Third, martingale arguments and the key identity (1.6) give approximate asymptotic variance formulas under heavy censoring for both $\hat{\mathbf{\Lambda}}_n(t)$ and $\hat{\mathbb{S}}_n(t)$ which we present in Section 4. The statistical implications of these formulas are discussed briefly, and comparisons made with the finite sample calculations of Chen, Hollander, and Langberg (1982). The tables in Section 4, calculated under the assumption that $(1 - G) = (1 - F)^\beta$, $0 < \beta < \infty$, suggest that $\hat{\mathbb{S}}_n$ has smaller bias and variance than $\bar{\mathbb{F}}_n \equiv \hat{\mathbb{S}}_n \, 1_{[0, Z_{(n)})}$ for a substantial range of t's.

2. An exponential formula. In this section we derive both (1.6) and proposition A.4.1 in Appendix 4 of Gill (1980) as special cases of the theory of the "exponential of a semimartingale" due to Doleans-Dade (1970). None of the identities in this section are new, and, as already noted above, the cases of particular interest here can all be derived via more elementary integration by

parts methods thereby avoiding stochastic integration theory. The present approach to these results simply serves to illustrate the power of the general exponential formula, and are included here in hopes that these connections will become more widely known. For further related results see Jacod (1979) pages 190 ff, Meyer (1976) pages 304 ff, Yoeurp (1976), and Yor (1976). For a recent exposition of martingale theory, see Shiryayev (1981).

A process $\mathbb{X}$ is a *semimartingale* if it can be written as $\mathbb{X} = \mathbb{M} + V$ where $\mathbb{M}$ is a local martingale and V is a process of locally bounded variation: $\int_0^t |\, d\, V(t)| < \infty$ with probability one for any $0 \le t < \infty$.

THEOREM 1. (Doleans-Dade, 1970). *Let $\mathbb{X}$ be a semimartingale with $\mathbb{X}(0) = 0$. Then there exists a unique semimartingale, $\mathbb{Z}$, called the exponential of $\mathbb{X}$, satisfying*

$$\mathbb{Z}(t) = 1 + \int_{(0,t]} \mathbb{Z}_- \, d\mathbb{X}, \quad \text{for} \quad 0 \le t < \infty. \tag{2.1}$$

It is given by the "exponential formula"

$$\mathbb{Z}(t) = \exp(\mathbb{X}(t) - \tfrac{1}{2}\langle \mathbb{X}^c \rangle(t)) \prod_{s \le t} (1 + \Delta\mathbb{X}(s))\exp(-\Delta\mathbb{X}(s)) \tag{2.2}$$

where $\mathbb{X}^c$ denotes the continuous martingale part of $\mathbb{X}$ and $\Delta\mathbb{X}(s) = \mathbb{X}(s) - \mathbb{X}(s-)$.

A very special case of Theorem 1 is the following correspondence between a df and its cumulative hazard function which is apparently due to Jacod (1975); see pages 255–256 of Lipster and Shiryayev (1978) for a somewhat more general result in this same vein.

COROLLARY 1. (Jacod, 1975). *If F is a df on R^+ with $F(0) = 0$ and the cumulative hazard function Λ is defined by*

$$\Lambda(t) = \int_{(0,t]} \frac{dF}{1 - F_-}, \tag{2.3}$$

then

$$1 - F(t) = \exp(-\Lambda_c(t)) \prod_{s \le t} (1 - \Delta\Lambda(s)) \tag{2.4}$$

with $\Lambda_c(t) \equiv \Lambda(t) - \sum_{s \le t} \Delta\Lambda(s)$.

PROOF. This follows immediately from Theorem 1 applied to the (deterministic) semimartingale $\mathbb{X} = -\Lambda$ upon noting that (2.3) implies $F(t) = \int_{(0,t]}(1 - F_-)\, d\Lambda$ or $1 - F(t) = 1 - \int_{(0,t]}(1 - F_-)\, d\Lambda$. (In this case $\mathbb{X}^c \equiv 0$.) □

Now we use Theorem 1 and Corollary 1 to establish an identity of Gill (1980). Let A and B be right-continuous nondecreasing functions on $[0, \infty)$ with $A(0) = B(0) = 0$ and $\Delta A \le 1$, $\Delta B < 1$ on $[0, \infty)$. Thus B is a cumulative hazard

function and so is A if $A(s) = A(t)$ for all $s \geq t$ when $\Delta A(t) = 1$; and the right side of (2.4) defines a df when Λ is replaced by either Λ or B in any case. Let $\tau_B \equiv \inf\{t: B(t) = \infty\}$.

COROLLARY 2. (Gill, 1980). *The unique locally bounded solution* $\mathbb{Z}$ *of*

$$\mathbb{Z}(t) = 1 - \int_{(0,t]} \frac{\mathbb{Z}(s-)}{1 - \Delta B(s)}\, d(A(s) - B(s)) \tag{2.5}$$

on $[0, \tau_B)$ *is given by*

$$\mathbb{Z}(t) = \frac{\exp(-A_c(t)) \prod_{s \leq t} (1 - \Delta A(s))}{\exp(-B_c(t)) \prod_{s \leq t} (1 - \Delta B(s))} \tag{2.6}$$

$$= \frac{1 - F_A(t)}{1 - F_B(t)} \tag{2.7}$$

where F_A *and* F_B *are the df's corresponding to* A *and* B.

PROOF. Let

$$\mathbb{X}(t) \equiv -\int_{(0,t]} \frac{1}{1 - \Delta B}\, d(A - B), \quad 0 \leq t < \tau_B. \tag{a}$$

Then $\mathbb{X}$ is a (deterministic) semimartingale since it has locally bounded variation, and (2.5) can be written in the form (2.1) with this choice of $\mathbb{X}$. Thus (2.2) of Theorem 1 gives the solution of (2.5): We calculate

$$1 + \Delta\mathbb{X} = 1 - \frac{1}{1 - \Delta B}(\Delta A - \Delta B) = \frac{1 - \Delta A}{1 - \Delta B} \tag{b}$$

and

$$\begin{aligned} \mathbb{X}_c(t) &\equiv \mathbb{X}(t) - \textstyle\sum_{s \leq t} \Delta\mathbb{X}(s) = -\int_{(0,t]} \frac{1}{1 - \Delta B}\, d(A_c - B_c) \\ &= -(A_c(t) - B_c(t)) \end{aligned} \tag{c}$$

by continuity of A_c and B_c. Note that $\mathbb{X}^c$ in (2.2) is identically 0 here. Hence by (2.2)

$$\mathbb{Z}(t) = \exp(\mathbb{X}_c(t)) \prod_{s \leq t} (1 + \Delta\mathbb{X}(s)) \tag{d}$$

which equals the right side of (2.6) by (b) and (c). Now (2.7) holds by (two applications of) Corollary 1. □

COROLLARY 3. (Aalen and Johansen; Gill). *The identity* (1.6) *holds for* $0 \leq t \leq \tau_\Lambda = \inf\{t: F(t) = 1\}$

$$\frac{\hat{\mathbb{S}}_n(t)}{S(t)} = 1 - \int_{(0,t]} \frac{\hat{\mathbb{S}}_{n-}}{S}\, d(\hat{\mathbf{\Lambda}}_n - \Lambda). \tag{2.8}$$

PROOF. This follows immediately from Corollary 2 with $A \equiv \hat{\Lambda}_n$ (for each fixed ω), and $B = \Lambda$, and using (1.5) and (2.4). Note that $1 - \Delta\Lambda = (1 - F)/(1 - F_-)$. Alternatively, (2.8) follows from Theorem 1 applied directly to the martingale

$$\text{(a)} \qquad \mathbb{X}(t) = \int_{(0,t\wedge T_n]} \frac{1}{1-\Delta\Lambda}\, d(\hat{\Lambda}_n - \Lambda) = \int_{(0,t\wedge T_n]} \frac{1}{1-\Delta\Lambda}\,\frac{1}{\mathbb{R}_n}\, d\mathbb{M}_n^u$$

where $\mathbb{M}_n^u$ is the counting process martingale defined by

$$\text{(b)} \qquad \mathbb{M}_n^u = \mathbb{N}_n^u - \int_0^{\cdot} \mathbb{R}_n\, d\Lambda;$$

see Gill (1980) or (1983). □

3. Heavy censoring limit theorems. We now turn to limit theorems for the counting processes $(\mathbb{N}_n^u, \mathbb{R}_n)$ and $\hat{\Lambda}_n$ and $\hat{\mathbb{S}}_n$ under "heavy censoring", $G_n \to_d \delta_0 = 1_{[0,\infty)}$. Throughout this section we assume that both F and $G = G_n$ are continuous, and we write $Y_{n1}, \cdots, Y_{nn}$ for the independent Y's with df G_n to emphasize the dependence of their distribution on n. The following assumption specifies the rate of increasing censorship:

ASSUMPTION C. For $0 < t < \infty$

$$n(1 - G_n(t)) \to a(t) \tag{3.1}$$

and

$$\int_{(0,t]} n(1 - G_n)\, dF \to \int_{(0,t]} a\, dF < \infty \tag{3.2}$$

as $n \to \infty$ where a is continuous on $(0, \infty)$.

Extreme value theory provides a variety of G_n's satisfying assumption C; see e.g. Lemma 2.2.2 page 62 or Theorem 2.3.1 page 69 of De Haan (1975).

Let

$$\mathcal{F}_n(t) \equiv \sigma\{\mathbb{N}_n^u(s), \mathbb{N}_n^c(s): 0 < s \le t\} \quad \text{for} \quad 0 < t < \infty. \tag{3.3}$$

In the following, $\to_d$ will mean weak convergence in the Skorokhod J_1 topology (or product topology in the case of (3.4)). For the processes of interest here, this may be interpreted as pointwise convergence at all continuity points of the limit process; see Kallenberg (1976) or Brown (1981) for further details.

THEOREM 2. *If F is continuous and assumption C holds, then*

$$(\mathbb{N}_n^u, \mathbb{R}_n) \to_d (\mathbb{N}^u, \mathbb{R}) \quad \textit{as} \quad n \to \infty \tag{3.4}$$

where $\mathbb{N}^u$ is a nonhomogeneous Poisson process with mean function $A = \int_0^{\cdot} a\, dF$; and $\mathbb{R} =_d \mathbb{N} \circ B$ with $\mathbb{N}$ a standard Poisson process and $B(t) = a(t)(1 - F(t))$.

Moreover,

$$\mathbb{M}_n^u = \mathbb{N}_n^u - \int_0^\cdot \mathbb{R}_n \, d\Lambda \to_d \mathbb{N}^u - \int_0^\cdot \mathbb{R} \, d\Lambda \equiv \mathbb{M}^u \tag{3.5}$$

where $\mathbb{M}^u$ is a mean zero square integrable $\mathcal{F}(t)$-martingale with predictable variation process

$$\langle \mathbb{M}^u \rangle = \int_{(0,\cdot]} \mathbb{R} \, d\Lambda. \tag{3.6}$$

Here $\mathcal{F}(t)$ is the limiting filtration corresponding to $\mathcal{F}_n(t)$:

$$\mathcal{F}(t) \equiv \sigma\{\mathbb{N}^u(s);\ \mathbb{R}(s+):\ 0 < s \le t\}. \tag{3.7}$$

PROOF. For $\varepsilon > 0$ define the point process ξ_n on $[0, \infty) \times [\varepsilon, \infty)$ by

(a) $\xi_n(D) \equiv \sum_{i=1}^n 1_D(X_i, Y_{ni})$ for Borel sets $D \subset [0, \infty) \times [\varepsilon, \infty)$.

Then $\xi_n(D) \cong \text{Binomial}(n, (F \times G_n)(D))$ with

(b)
$$\begin{aligned} n(F \times G_n)(D) &= -n \int_D dF \, d(1 - G_n) \\ &\to -\int_D dF \, da \equiv \mu(D) \quad \text{as} \quad n \to \infty \end{aligned}$$

by (3.1) and Helly-Bray for any Borel set $D \subset [0, \infty) \times [\varepsilon, \infty)$. Hence, by Theorem 4.7 of Kallenberg (1976), it follows that $\xi_n \to_d \xi$ where ξ is a Poisson point process on $[0, \infty) \times [\varepsilon, \infty)$ with intensity measure μ.

Let $L \equiv \{(x, y): x \le y\}$. Since

(c) $$\mathbb{N}_n^u(t) - \mathbb{N}_n^u(\varepsilon) = \xi_n((\varepsilon, t] \times [\varepsilon, \infty) \cap L)$$

and

(d)
$$\begin{aligned} \mathbb{R}_n(\varepsilon) - \mathbb{R}_n(t) &= \xi_n([\varepsilon, \infty) \times [\varepsilon, t) \cap L^c) + \mathbb{N}_n^u(t-) - \mathbb{N}_n^u(\varepsilon-) \\ &= \xi_n([\varepsilon, \infty) \times [\varepsilon, t) \cap L^c) + \xi_n([\varepsilon, t) \times [\varepsilon, \infty) \cap L). \end{aligned}$$

the mapping from ξ_n to $(\mathbb{N}_n^u, \mathbb{R}_n)$ as point processes on $[\varepsilon, \infty)$ is continuous, and hence the joint convergence of $(\mathbb{N}_n^u, \mathbb{R}_n)$ on $[\varepsilon, \infty)$ follows from that of ξ_n. That the limit processes have the given mean functions follows from assumption C since

(e) $$E\mathbb{N}_n^u(t) = \int_{(0,t]} n(1 - G_n) \, dF \to \int_{(0,t]} a \, dF \equiv A(t) \quad \text{as} \quad n \to \infty$$

and

(f) $$E\mathbb{R}_n(t) = (1 - F(t))n(1 - G_n(t)) \to (1 - F(t))a(t) \equiv B(t) \quad \text{as} \quad n \to \infty.$$

Note that $A \uparrow$ while $B \downarrow$. Letting $\varepsilon \to 0$ yields $(\mathbb{N}_n^u, \mathbb{R}_n) \to_d (\mathbb{N}^u, \mathbb{R})$ on $(0, \infty)$.

Since the limiting Poisson process ξ on L is independent of ξ on L^c, it follows from (d) that increments of $\mathbb{R}$ may be written as the sum of the same increment of (the left continuous version of) $\mathbb{N}^u$ and a term independent of $\mathbb{N}^u$.

Now $\mathbb{M}_n^u$ on the left side in (3.5) is a continuous function of $(\mathbb{N}_n^u, \mathbb{R}_n)$, and hence the convergence in (3.5) holds. (Note that

$$E\mathbb{N}_n^u(\varepsilon) = E\int_{(0,\varepsilon]} \mathbb{R}_n \, d\Lambda = \int_{(0,\varepsilon]} n(1 - G_n)\, dF \to \int_{(0,\varepsilon]} a\, dF$$

which can be made arbitrarily small by choice of ε since $\int_{(0,t]} a\, dF < \infty$ by assumption C.) That $\mathbb{M}^u$ on the right side of (3.5) is a martingale follows from the fact that $\mathbb{M}_n^u$ is an $\mathcal{F}_n(t)$-martingale and uniform integrability arguments along the lines of Brown (1981). □

The present proof of Theorem 2 was suggested by Tim Brown. Our original proof established only marginal convergence of the processes $\mathbb{N}_n^u$ and $\mathbb{R}_n$ by verifying convergence of the compensators of these counting processes, and then applying Theorem 1 of Kabanov, Lipster, and Shiryayev (1980).

The following alternative approach to the Poisson behavior of $\mathbb{N}_n^u$, $\mathbb{N}_n^c$ was suggested to me by Bernard Silverman. Fix $\varepsilon > 0$ and let $\lambda \equiv nP(Z \geq \varepsilon) = n(1 - F(\varepsilon))(1 - G_n(\varepsilon))$. Let (Z_i^*, δ_i^*), $i = 1, 2, \cdots$ be iid as (Z, δ) *conditional on* the event $[Z \geq \varepsilon]$, and let

$$\nu_n \cong \text{Binomial}(n, \lambda/n), \quad \nu \cong \text{Poisson}(\lambda)$$

be independent of the (Z_i^*, δ_i^*)'s. In view of Hodges and LeCam (1960) and LeCam (1963), ν and ν_n can be constructed in such a way that

$$P(\nu_n \neq \nu) \leq \frac{\lambda^2}{n} \wedge \frac{3\lambda}{n}; \tag{3.8}$$

related results are given by Vervaat (1969) and Brown (1984). Then for $t \geq \varepsilon$ set

$$^*\mathbb{R}_n(t) \equiv \textstyle\sum_{i=1}^{\nu_n} 1_{[Z_i^* \geq t]}, \qquad ^*\mathbb{R}(t) \equiv \textstyle\sum_{i=1}^{\nu} 1_{[Z_i^* \geq t]}$$

$$^*\overline{\mathbb{N}_n^u}(t) \equiv \textstyle\sum_{i=1}^{\nu_n} 1_{[Z_i^* \geq t]}\delta_i^*, \qquad ^*\overline{\mathbb{N}^u}(t) \equiv \textstyle\sum_{i=1}^{\nu} 1_{[Z_i^* \geq t]}\delta_i^*$$

$$^*\overline{\mathbb{N}_n^c}(t) \equiv \textstyle\sum_{i=1}^{\nu_n} 1_{[Z_i^* \geq t]}(1 - \delta_i^*), \quad ^*\overline{\mathbb{N}^c}(t) \equiv \textstyle\sum_{i=1}^{\nu} 1_{[Z_i^* \geq t]}(1 - \delta_i^*).$$

It is easily checked that $^*\overline{\mathbb{N}^u}$, $^*\overline{\mathbb{N}^c}$ are independent Poisson processes on $[\varepsilon, \infty)$ with decreasing mean functions $n\int_{[t,\infty)}(1 - G_n)\, dF$ and $n\int_{[t,\infty)}(1 - F)\, dG_n$ respectively, and that $^*\overline{\mathbb{N}_n^u}$ and $^*\overline{\mathbb{N}_n^c}$ are equal in distribution for $t \geq \varepsilon$ to their unstarred counterparts $\overline{\mathbb{N}_n^u} \equiv \sum_{i=1}^n 1_{[Z_i \geq \cdot]}\delta_i$ and $\overline{\mathbb{N}_n^c} \equiv \sum_{i=1}^n 1_{[Z_i \geq \cdot]}(1 - \delta_i)$. Furthermore in view of (3.8) it follows immediately from the definitions of the starred processes that

$$P(^*\overline{\mathbb{N}_n^u}(t) = {}^*\overline{\mathbb{N}^u}(t) \text{ and } ^*\overline{\mathbb{N}^c}(t) = {}^*\overline{\mathbb{N}^c} \text{ for all } t \geq \varepsilon) \geq 1 - (\lambda^2/n \wedge 3\lambda/n).$$

This gives a very explicit construction of Poisson processes close to the original counting processes on the interval $[\varepsilon, \infty)$.

Now we use Theorem 2 to obtain corresponding limit theorems for the cumulative hazard function estimator $\hat{\mathbf{\Lambda}}_n$ and the product limit estimator $\hat{\mathbb{S}}_n$.

THEOREM 3. *Under the assumptions of Theorem* 2 *we have*

$$\hat{\mathbf{\Lambda}}_n \to_d \int_{(0,\cdot]} \frac{1}{\mathbb{R}}\, d\mathbb{N}^u \equiv \hat{\mathbf{\Lambda}}. \tag{3.9}$$

Moreover, (recall that $T_n \equiv Z_{(n)}$, *the largest* Z*),*

$$(\hat{\mathbf{\Lambda}}_n - \Lambda)^{T_n} = \int_{(0,\cdot\wedge T_n]} \frac{1}{\mathbb{R}_n}\, d\mathbb{M}_n^u \to_d \int_{(0,\cdot]} \frac{1_{[IR>0]}}{\mathbb{R}}\, d\mathbb{M}^u \equiv \mathbb{H} \tag{3.10}$$

where $\mathbb{H}$ *is a mean-zero* $\mathscr{F}(t)$ *local-martingale with predictable variation process*

$$\langle \mathbb{H} \rangle = \int_{(0,\cdot]} \frac{1_{[\mathbb{R}>0]}}{\mathbb{R}}\, d\Lambda. \tag{3.11}$$

THEOREM 4. *Under the assumptions of Theorem* 2

$$\hat{\mathbb{S}}_n = \prod_{s\le\cdot} (1 - \Delta\hat{\mathbf{\Lambda}}_n(s)) \to_d \prod_{s\le\cdot} (1 - \Delta\hat{\mathbf{\Lambda}}(s)) \equiv \mathbb{S} \tag{3.12}$$

where $\hat{\mathbf{\Lambda}}$ *is given by* (3.9). *Moreover, the limit process* $\mathbb{S}$ *satisfies*

$$\frac{\mathbb{S}}{S} = 1 - \int_{(0,\cdot]} \frac{\mathbb{S}_-}{S}\, d\,(\hat{\mathbf{\Lambda}} - \Lambda) \tag{3.13}$$

and hence, with $T \equiv \sup\{t\colon \mathbb{R}(t) > 0\}$,

$$\mathbb{Z} \equiv \left(1 - \frac{\mathbb{S}}{S}\right)^T \tag{3.14}$$

is a mean-zero $\mathscr{F}(t)$ *local-martingale with predictable variation process*

$$\langle \mathbb{Z} \rangle = \int_{(0,\cdot\wedge T]} \left(\frac{\mathbb{S}_-}{S}\right)^2 \frac{1}{\mathbb{R}}\, d\Lambda. \tag{3.15}$$

PROOFS. Since $(\mathbb{N}_n^u, \mathbb{R}_n) \to_d (\mathbb{N}^u, \mathbb{R})$ as $n \to \infty$, it follows from Skorokhod (1956) that there exist equivalent processes $(\mathbb{N}_n^{u*}, \mathbb{R}_n^*) =_d (\mathbb{N}_n^u, \mathbb{R}_n)$ and $(\mathbb{N}^{u*}, \mathbb{R}^*) =_d (\mathbb{N}^u, \mathbb{R})$ defined on a common probability space such that $(\mathbb{N}_n^{u*}, \mathbb{R}_n^*) \to_{\text{a.s.}} (\mathbb{N}^{u*}, \mathbb{R}^*)$ as $n \to \infty$. Now $A(t) \to A(\infty) = \int_0^\infty a\, dF < \infty$ as $t \to \infty$, and A is continuous by assumption C and continuity of F. Hence $\kappa \equiv \mathbb{N}^{u*}(\infty) < \infty$ a.s. and the jump times of $\mathbb{N}^{u*}$ are a.s. distinct: $0 < T_1 < \cdots < T_\kappa < \infty$. Since $\mathbb{N}_n^{u*} \to_{\text{a.s.}} \mathbb{N}^{u*}$ (in the sense that $\mathbb{N}_n^{u*}(t) \to_{\text{a.s.}} \mathbb{N}^{u*}(t)$ for all t such that $\Delta\mathbb{N}^{u*}(t) = 0$), if $0 < T_{n1} < \cdots < T_{n\kappa(n)}$ denote the ordered jump times of $\mathbb{N}_n^{u*}$, it follows that $T_{ni} \to T_i$ for $i = 1, \cdots, \kappa$ a.s. (and $\kappa(n) = \kappa$ for all $n \ge$ some N_ω). Therefore, noting that the jump points of $\mathbb{N}_n^{u*}$ and $\mathbb{N}^{u*}$ are a subset of the jump points of

$\mathbb{R}_n^*$ and $\mathbb{R}^*$ respectively and $\mathbb{R}_n^* \to_{\text{a.s.}} \mathbb{R}^*$, for any t such that $\Delta\Lambda^*(t) = (\mathbb{R}^*(t))^{-1}\Delta\mathbb{N}^{u*}(t) = 0$ we have

$$\Lambda_n^*(t) \equiv \int_{(0,t]} \frac{1}{\mathbb{R}_n^*}\, d\mathbb{N}_n^{u*} = \sum_{i=1}^{\kappa(n)} (\mathbb{R}_n^*(T_{ni}))^{-1} 1_{[T_{ni} \le t]}$$

$$\to_{\text{a.s.}} \sum_{i=1}^{\kappa} (\mathbb{R}^*(T_i))^{-1} 1_{[T_i \le t]} = \int_{(0,t]} \frac{1}{\mathbb{R}^*}\, d\mathbb{N}^{u*} = \Lambda^*(t),$$

and hence (3.9) holds. Convergence in (3.10) follows similarly, and the martingale property holds by uniform integrability arguments.

By Proposition (4.10) of Jacod and Memin (1980), the function $\Lambda \to 1 - F$ given by (2.4) is continuous. Hence Theorem 4 follows from Theorem 3, (1.5), and the identity (1.6). □

4. Approximate variance formulas and small sample behavior of $\hat{\mathbb{S}}_n$. Now we want to exploit (3.11) of Theorem 3 and (3.15) of Theorem 4 to derive "approximate" or "heuristic" formulas, by replacing $\mathbb{R}$ by its expected value $a(1 - F)$. Doing this in (3.11) yields the approximate variance formula

$$\text{Var}[\Lambda(t)] \cong \int_{(0,t]} \frac{1}{(1-F)a}\, d\Lambda = \int_{(0,t]} \frac{1}{(1-F)^2 a}\, dF \equiv C(t) \tag{4.1}$$

which should be compared with the formula resulting from the Gaussian limit theory as given in (7.11) of Breslow and Crowley (1974):

$$\text{Asympt. Var}[\hat{\Lambda}_n(t)] = \int_{(0,t]} \frac{1}{(1-F)^2 n(1-G)}\, dF. \tag{4.2}$$

In the heavy censoring Poisson limit theory, $n(1 - G)$ is replaced by its limit function α, and hence these two approximate variance formulas for $\hat{\Lambda}_n(t)$ are essentially the *same* in both the Gaussian and Poisson limit theories.

This is *not* the case for $\hat{\mathbb{S}}_n(t)$ however. Replacing $\mathbb{R}$ by its expected value $a(1 - F)$ in (3.15) and then taking expectations suggests that, at least approximately, with $M_2(t) \equiv E(\mathbb{S}(t)/S(t))^2$,

$$M_2(t) \cong 1 + \int_{(0,t]} M_2\, dC \tag{4.3}$$

with C as defined in (4.1). If "=" held in (4.3), the solution would be $M_2(t) = \exp(C(t))$, and hence we have

$$\begin{aligned} \text{Var}[\mathbb{S}(t)] &\cong S(t)^2\{\exp(C(t)) - 1\} \\ &= S(t)^2 \left\{ \exp\left(\int_{(0,t]} \frac{1}{(1-F)^2 a}\, dF \right) - 1 \right\}. \end{aligned} \tag{4.4}$$

When the limit function a is (4.4) is replaced by $n(1 - G)$, (4.4) yields the following approximation of $\mathrm{Var}[\hat{\mathbb{S}}_n(t)]$:

$$\mathrm{Var}[\hat{\mathbb{S}}(t)] \cong S(t)^2 \left\{ \exp\left(\int_{(0,t]} \frac{1}{(1 - F)^2 n(1 - G)}\, dF \right) - 1 \right\} \equiv VP_n(t). \tag{4.5}$$

Since $e^x - 1 \geq x$, the right side of (4.5) is always larger than the approximate variance formula resulting from the Gaussian limit theory (the familiar "Green-

TABLE 1

Exact and approximate values of the mean and variance of the product-limit estimator $\bar{\mathbb{F}}_n$

t	$\bar{F}(t)$	β	n	$E\{\bar{\mathbb{F}}_n(t)\}$	$b_n^0(t)$	$\mathrm{Var}\{\bar{\mathbb{F}}_n(t)\}$	$VG_n(t)$	$VP_n(t)$
.5	.60653	.5	10	.60641	.00012	.02814	.02739	.02844
			15	.60653	.00000	.01853	.01826	.01872
			20	.60653	.00000	.01384	.01370	.01396
			25	.60653	.00000	.01105	.01096	.01112
			30	.60653	.00000	.00919	.00913	.00925
.5	.60653	1.0	10	.60485	.00168	.03475	.03161	.03300
			15	.60639	.00014	.02193	.02107	.02169
			20	.60652	.00001	.01621	.01580	.01615
			25	.60653	.00000	.01289	.01264	.01286
			30	.60653	.00000	.01070	.01054	.01069
.5	.60653	2.0	10	.58154	.02499	.06394	.04269	.04527
			15	.60019	.00634	.03586	.02846	.02959
			20	.60488	.00165	.02422	.02135	.02198
			25	.60609	.00044	.01842	.01708	.01748
			30	.60641	.00012	.01499	.01423	.01451
1.0	.36788	.5	10	.36246	.00542	.03613	.03141	.03536
			15	.36665	.00123	.02261	.02094	.02265
			20	.36759	.00029	.01643	.01571	.01665
			25	.36781	.00007	.01295	.01257	.01317
			30	.36786	.00002	.01071	.01047	.01089
1.0	.36788	1.0	10	.33400	.03388	.05889	.04323	.05094
			15	.35379	.01409	.03778	.02882	.03212
			20	.36180	.00608	.02647	.02162	.02344
			25	.36520	.00268	.01999	.01729	.01845
			30	.36668	.00120	.01599	.01441	.01521
1.0	.36788	2.0	10	.21300	.15488	.08516	.08610	.12035
			15	.25646	.11142	.07554	.05740	.07149
			20	.28646	.08142	.06480	.04305	.05068
			25	.30779	.06009	.05496	.03444	.03922
			30	.32323	.04465	.04653	.02870	.03197
2.0	.13534	.5	10	.10688	.02846	.02379	.02330	.04706
			15	.11654	.01879	.01785	.01554	.02446
			20	.12247	.01287	.01401	.01165	.01629
			25	.12633	.00901	.01135	.00932	.01215
			30	.12893	.00640	.00942	.00777	.00968

wood's formula"):

$$(4.6)\quad \text{Asympt. Var}[\hat{\mathbb{S}}_{n(t)}] \cong S(t)^2 \int_{(0,t]} \frac{1}{(1-F)^2 n(1-G)}\, dF \equiv VG_n(t);$$

see e.g. (7.13) in Breslow and Crowley (1974). Also note that (4.5) and (4.6) agree as $n \to \infty$ (with $1 - G$ regarded as fixed).

Table 1 gives a numerical comparison of the exact variance of $\overline{\mathbb{F}}_n(t) = \overline{\mathbb{S}}_n(t)\ 1_{[0,Z_{(n)})}(t)$ (in a case where it can be calculated exactly) with the two approximations $VP_n(t)$ and $VG_n(t)$ defined above. The first 8 columns of Table 1 duplicate (with one added digit, and without the bias bound column) Table 1 of Chen, Hollander, and Langberg (1982) where exact formulas for $E[\overline{\mathbb{F}}_n(t)^\alpha]$ are

TABLE 2

Exact and approximate values of the mean and variance of the product-limit estimator $\hat{\mathbb{S}}_n$

t	$S(t)$	β	n	$E\{\hat{\mathbb{S}}_n(t)\}$	$b_n^p(t)$	$\text{Var}\{\hat{\mathbb{S}}_n(t)\}$	$VG_n(t)$	$VP_n(t)$
.5	.60653	.5	10	.60654	−.00001	.02802	.02739	.02844
			15	.60653	.00000	.01852	.01826	.01872
			20	.60653	.00000	.01384	.01370	.01396
			25	.60653	.00000	.01105	.01096	.01112
			30	.60653	.00000	.00919	.00913	.00925
.5	.60653	1.0	10	.60665	−.00012	.03335	.03161	.03300
			15	.60654	−.00001	.02181	.02107	.02169
			20	.60653	.00000	.01620	.01580	.01615
			25	.60653	.00000	.01289	.01264	.01286
			30	.60653	.00000	.01070	.01054	.01069
.5	.60653	2.0	10	.60868	−.00215	.04776	.04269	.04527
			15	.60692	−.00039	.03126	.02846	.02959
			20	.60661	−.00008	.02294	.02135	.02198
			25	.60655	−.00002	.01807	.01708	.01748
			30	.60653	.00000	.01489	.01423	.01451
1.0	.36788	.5	10	.36883	−.00095	.03335	.03141	.03536
			15	.36803	−.00015	.02191	.02094	.02265
			20	.36791	−.00003	.01625	.01571	.01665
			25	.36789	−.00001	.01291	.01257	.01317
			30	.36788	.00000	.01070	.01047	.01089
1.0	.36788	1.0	10	.37516	−.00728	.04745	.04323	.05094
			15	.37010	−.00222	.03179	.02882	.03212
			20	.36864	−.00076	.02360	.02162	.02344
			25	.36816	−.00028	.01864	.01729	.01845
			30	.36799	−.00011	.01536	.01441	.01521
1.0	.36788	2.0	10	.41643	−.04855	.07794	.08610	.12035
			15	.39464	−.02676	.05805	.05740	.07149
			20	.38389	−.01601	.04617	.04305	.05068
			25	.37793	−.01005	.03816	.03444	.03922
			30	.37440	−.00652	.03238	.02870	.03197
2.0	.13534	.5	10	.15460	−.01926	.02540	.02330	.04706
			15	.14486	−.00953	.01689	.01554	.02446
			20	.14061	−.00527	.01266	.01165	.01629
			25	.13845	−.00311	.01012	.00932	.01215
			30	.13726	−.00192	.00841	.00777	.00968

derived under the proportional hazard censoring assumption $1 - G = (1 - F)^\beta$, $0 < \beta < \infty$. Both their Table 1 and our Table 1 have been calculated with $S(t) \equiv 1 - F(t) = e^{-t}$, $1 - G(t) = e^{-\beta t}$, $\gamma = P[X \leq Y] = (1 + \beta)^{-1}$. Let $b_n^0(t) \equiv S(t) - E[\overline{\mathbb{F}}_n(t)]$ denote the bias of $\overline{\mathbb{F}}_n(t)$.

While the Gaussian limit theory approximation $VG_n(t)$ to $\mathrm{Var}[\overline{\mathbb{F}}_n(t)]$ is often too small, the Poisson limit theory approximation of $\mathrm{Var}[\overline{\mathbb{F}}_n(t)]$ is sometimes too large, less often a little too small, but in general somewhat more conservative. Note that with

$$1 - H(t) = (1 - F(t))(1 - G(t)) = \exp(-(1 + \beta)t),$$

for $\beta = .5, 1, 2$, $1 - H(2) \cong .05, .018, .0025$ respectively, so it is probably unreasonable to expect the Poisson approximation to be accurate for $t = 2$ and $\beta = 1$ or 2 when n is 30 or less. The Poisson approximation does seem to be reasonably good when $n(1 - H(t)) > 1$.

Now the methods of Chen, Hollander, and Langberg (1982) apply equally well to $\hat{\mathbb{S}}_n(t)$ under $1 - G = (1 - F)^\beta$, and easy calculations show that their (3.1) holds with $\overline{\mathbb{F}}_n(t)$ replaced by $\hat{\mathbb{S}}_n(t)$ if the upper terminal of summation $n - 1$ on the right side is replaced by n. Hence

$$(4.7) \qquad E[\hat{\mathbb{S}}(t)^\alpha] = \sum_{q=0}^{n} \binom{n}{q} H(t)^q (1 - H(t))^{n-q} \prod_{i=1}^{q} [\gamma c_{in}^\alpha + (1 - \gamma)]$$

where $c_{in} = (n - i)/(n - i + 1)$ and $1 - H = (1 - F)(1 - G) = (1 - G)^{1+\beta}$ under $1 - G = (1 - F)^\beta$. Table 2 gives the resulting table of means and variances of $\hat{\mathbb{S}}_n(t)$ under the same set of assumptions used to calculate Table 1. Note that both the biases $b_n^p(t) \equiv S(t) - E[\hat{\mathbb{S}}_n(t)]$ and variances of $\hat{\mathbb{S}}_n(t)$ are almost everywhere smaller than those for $\overline{\mathbb{F}}_n(t)$, sometimes substantially so, with the exception of the values for $t = 2$ with heavy censorship. These two tables lead us to prefer $\hat{\mathbb{S}}_n$.

Acknowledgements. We owe thanks to Michael Akritas, to Tim Brown for suggesting the present proof of Theorem 2, and to Bernard Silverman for the Poisson approximation argument.

REFERENCES

AALEN, O. O., and JOHANSEN, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand. J. Statist.* **5** 141–150.

BILLINGSLEY, P. (1971). Weak convergence of measures: applications in probability. *Regional Conf. Series in Applied Mathematics*, No. **5** SIAM, Philadelphia.

BREMAUD, P. (1981). *Point Processes and Queues; Martingale Dynamics.* Springer-Verlag, New York.

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

BROWN, T. C. (1981). Compensators and Cox convergence. *Math. Proc. Camb. Phil. Soc.* **90** 305–319.

BROWN, T. C. (1984). Poisson approximations and the definition of the Poisson process. *Amer. Math. Monthly* **91** 116–123.

CHEN, Y. Y., HOLLANDER, M., and LANGBERG, N. A. (1982). Small sample results for the Kaplan-Meier estimator. *J. Amer. Statist. Assoc.* **77** 141–144.

DE HAAN, L. (1975). On regular variation and its application to the weak convergence of sample extremes. *Mathematical Centre Tract* **32**. Mathematisch Centrum, Amsterdam.

DOLEANS-DADE, C. (1970). Quelques applications de la formule de changement de variables pour les semimartingales. *Z. Wahrsch. verw. Gebiete* **16** 181–194.

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

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

HODGES, J. L. and LE CAM, L. (1960). The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Statist.* **31** 737–740.

JACOD, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. verw. Gebiete* **31** 235–253.

JACOD, J. (1979). Calcul stochastique et problemes des martingales. *Lecture Notes in Math.* **714**. Springer-Verlag, Berlin.

JACOD, J. and MEMIN, J. (1980). Sur la convergence des semimartingalles vers un processus a accroissements independants. Sem. de Probabilites XIV. *Lecture Notes in Math.* **784** 227–248.

KABANOV, YU., LIPTSER, R., and SHIRYAYEV, A. (1980). Some limit theorems for simple point processes (a martingale approach). *Stochastics* **3** 203–216.

KALLENBERG, O. (1976). *Random Measures.* Academic Press, London.

LE CAM, L. (1963). On the distribution of sums of independent random variables. *Bernoulli, Bayes, Laplace; Proceedings of an International Research Seminar* 179–202. Springer-Verlag, Berlin.

LIPTSER, R. S., and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes II: Applications.* Springer-Verlag, Berlin.

MEYER, P. A. (1976). Un cours sur les Integrales Stochastique. Seminaire de Probabilites X. *Lecture Notes in Math.* **511** 246–400.

SHIRYAYEV, A. N. (1981). Martingales: recent developments, results, and applications. *Int. Statist. Rev.* **49** 199–233.

SKOROKHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probab. Applic.* **1** 261–290.

VERVAAT, W. (1969). Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution. *Statistica Neerlandica* **23** 79–86.

YOEURP, C. (1976). Decomposition des martingales locales et formules exponentielles. Seminaire de Probabilites X. *Lecture Notes in Math.* **511** 432–480.

YOR, M. (1976). Sur les integrales stochastiques optionelles et une suite remarquables de formules exponentielles. Seminaire de Probabilites X. *Lecture Notes in Math.* **511** 481–500. Springer-Verlag, Berlin.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195