The Annals of Statistics
1974, Vol. 2, No. 3, 437-453

A LARGE SAMPLE STUDY OF THE LIFE TABLE AND PRODUCT LIMIT ESTIMATES UNDER RANDOM CENSORSHIP[1]

BY N. BRESLOW AND J. CROWLEY

University of Washington and
International Agency for Research on Cancer, Lyon

Using the model of random censorship, a necessary and sufficient condition for the consistency of the standard (actuarial) life table estimate of a survival distribution is derived. We establish the asymptotic normality of this estimate, showing that Greenwood's variance formula is nearly correct. In the case of a continuous survival distribution we establish limiting normality for the product limit estimate and for the closely related cumulative hazard process. Some applications of these results are outlined.

1. Introduction. Although the life table is one of the statistical tools most commonly used by applied statisticians, rigorous derivations of many of its formal properties seem strangely to be lacking from the literature. This is true even of properties which are widely quoted and used. For example, Greenwood's (1926) formula for the variance of the cumulative survival probability (cf. (5.9) below and discussion) depends for its validity on the asymptotic independence of the estimates of the conditional probabilities of survival over the intervals used for grouping of the data. Chiang (1968, page 228) is often cited as a source for this result, although his proof applies only to the case of no live withdrawals. Derivations of the same result for life table estimates based on a specific parametric model are given only under the assumed model (Elveback (1958)).

The purpose of the present paper is to outline a general theory for the life table in which its familiar large sample properties can be rigorously established. In large part the material presented consists of a review and extension, in the light of both classical and modern large sample methods, of the fundamental papers on the subject by Kaplan and Meier (1958) and Chiang (1960a, b, 1961). Our Theorem 5 was first stated without proof by Efron (1967), so that the paper also consists of a review and formalization of his work.

In order to keep the mathematics as simple as possible, the theoretical development uses the device of random censorship introduced by Gilbert (1962) and later exploited by Efron (1967), Breslow (1969, 1970), Thomas (1972) and others. This is a very convenient tool for studying the large sample effects of censorship and the results obtained can, in many cases, easily be extended to the case of fixed or conditional censorship.

Received October 1972; revised June 1973.

[1] This research was supported in part by USPHS grant GM01269 to the University of Washington.

AMS 1970 *subject classifications.* Primary 62E20; Secondary 62G05.

Key words and phrases. Life table, product limit, censored data, consistency, weak convergence.

The type of life table which is considered in this paper is the cohort table used for estimation of a survival distribution from right censored data. Hence the material presented here will be of greatest interest to statisticians concerned with medical follow-up studies and life testing, and of some interest to actuaries and demographers. It is anticipated that the methodology employed may prove useful in the study of extensions of the life table method, such as that proposed recently by D. R. Cox (1972).

2. The statistical model: random censorship. Let $X_1^\circ, \cdots, X_N^\circ$ denote the true survival times for the N individuals included in the life table. These are assumed to be independent random variables having a common distribution $F^\circ(x) = P[X_n^\circ \leqq x]$ such that $F^\circ(0) = 0$. (This notation is borrowed in part from Efron (1967), although he works with left continuous survival functions in place of distribution functions.) The period of observation, or follow-up, for the nth individual will typically be limited by an amount Y_n. Formally speaking, the X_n° are censored on the right by the Y_n since one observes only

$$X_n = \min(X_n^\circ, Y_n) \quad \text{and} \quad \delta_n = I_{[X_n^\circ \leqq Y_n]}, \tag{2.1}$$

[2] where δ_n indicates whether X_n is censored ($\delta_n = 0$) or not ($\delta_n = 1$).

Under the random censorship model the censoring variables Y_n ($n = 1, \cdots, N$) are also assumed to be a random sample, drawn independently of the X_n°, from a distribution $H(y) = P[Y_n \leqq y]$. Hence the observed X's constitute a random sample from the distribution function F given by

$$(1 - F) = (1 - F^\circ)(1 - H), \tag{2.2}$$

while the sub-distribution function $\tilde{F}$ of an uncensored observation may be written

$$\tilde{F}(x) = P[X_n \leqq x, \delta_n = 1] = \textstyle\int_0^x (1 - H)\, dF^\circ. \tag{2.3}$$

3. The standard life table estimate (grouped data). Classical life table estimates are calculated from grouped data arising from a partition of the range $[0, T]$ of observation into, let us say, K intervals $I_k = (\xi_{k-1}, \xi_k]$ with endpoints $0 = \xi_0 < \xi_1 < \cdots < \xi_K < T$. The conditional probabilities of death in each interval

$$q_k = (F^\circ(\xi_k) - F^\circ(\xi_{k-1}))/(1 - F^\circ(\xi_{k-1})), \tag{3.1}$$

are the parameters of interest. They are combined by multiplication in order to obtain the probability of survival past ξ_k, written $P_k = 1 - F^\circ(\xi_k) = p_1 \cdots p_k$, where $p_k = 1 - q_k$.

Before giving an explicit definition of the most commonly used estimate of the q_k, we introduce the following statistics ($k = 1, \cdots, K$):

$$\begin{aligned} D_{1k} &= \textstyle\sum_{n=1}^N I_{[X_n \in I_k, Y_n > \xi_k, \delta_n = 1]}, \qquad & D_{2k} &= \textstyle\sum_{n=1}^N I_{[X_n \in I_k, Y_n \in I_k, \delta_n = 1]} \\ W_k &= \textstyle\sum_{n=1}^N I_{[X_n \in I_k, \delta_n = 0]} & & \\ N_{1k} &= \textstyle\sum_{n=1}^N I_{[X_n > \xi_{k-1}, Y_n > \xi_k]}, & N_{2k} &= \textstyle\sum_{n=1}^N I_{[X_n > \xi_{k-1}, Y_n \in I_k]} \end{aligned} \tag{3.2}$$

[2] Here as elsewhere I_A denotes the characteristic function of the event A.

and further $D_k = D_{1k} + D_{2k}$, $N_k = N_{1k} + N_{2k}$. Thus N_k, D_k, and W_k represent, respectively, the number of individuals alive at the beginning of I_k (and hence "at risk" of death in I_k), the number known to have died in I_k, and the number "withdrawn alive" in I_k. N_k and D_k may be further subdivided according to whether the individual is "due for withdrawal" (N_{2k} and D_{2k}) or not due for withdrawal (N_{1k} and D_{1k}) in I_k. However, this subdivision is possible only if one knows the censoring variables Y_n for all N individuals and this information is not always available. In general one has only the data specified by (2.1).

The standard (SD) life table estimate used most often in practice (Berkson and Gage (1950), Cutler and Ederer (1958), Gehan (1969)) is given by

$$\hat{q}_k = D_k/(N_k - \tfrac{1}{2}W_k)\,. \tag{3.3}$$

The fact that it is undefined if $N_k = 0$ is of no consequence since we are concerned only with large sample properties. However, the ambiguity may (and will) be resolved by taking $\hat{q}_k = 1$ in such cases.

Maximum likelihood estimates based on models which specify a parametric form for the survival distribution, and in which all withdrawals occur at the midpoint of each interval, have been proposed by Elveback (1958) and Chiang (1968) among others. While large sample properties for such estimates may be derived from the corresponding likelihoods, these can only be expected to hold under the assumed model. Kaplan and Meier (1958) studied the reduced sample (RS) estimate

$$\hat{q}_k = D_{1k}/N_{1k} \tag{3.4}$$

which is calculated solely from individuals not due for withdrawal in I_k. This estimate is consequently not commonly used and we introduce it here mainly for reference during later (theoretical) developments.

Only the RS estimate will generally be consistent for q_k. The SD estimate, in common with the parametric estimates, utilizes information from the N_{2k} individuals who are at risk of death for less than the entire interval and is thus (see Section 4) consistent only under special conditions relating the survival and censoring distributions. Nevertheless, it has been used by actuaries on an *ad hoc* basis for centuries: the $\frac{1}{2}W_k$ term in the denominator is supposed to adjust for the fact that the N_k individuals are not all at risk for the entire interval. Littel (1952) has suggested modifying the constant $\frac{1}{2}$ in order to improve the approximation in certain circumstances. Since the SD estimate is the one most widely used and quoted, we confine our attention to it. However, it will be readily apparent that the techniques developed can also be used to establish similar large sample properties for the other estimates.

4. Requirements for consistency of the SD estimate. Restricting attention for the moment to I_1, note that the SD estimate may be written

$$\hat{q}_1 = \left(\frac{N_{11}}{N_{11} + N_{21} - \frac{1}{2}W_1}\right)\left(\frac{D_{11}}{N_{11}}\right) + \left(\frac{N_{21} - \frac{1}{2}W_1}{N_{11} + N_{21} - \frac{1}{2}W_1}\right)\left(\frac{D_{21}}{N_{21} - \frac{1}{2}W_1}\right) \tag{4.1}$$

as a weighted average of the RS estimate and the estimate $D_{21}/(N_{21} - \frac{1}{2}W_1)$. This latter is just the SD estimate calculated exclusively from the N_{21} individuals due for withdrawal in I_1. Under the random censorship model, each of the frequencies N_{11}, N_{21}, D_{11}, etc. appearing in (4.1) is a sum of N i.i.d. Bernoulli variables whose respective probabilities can be calculated from (2.2) and (2.3). For example, $E(D_{11}) = NF^\circ(\xi_1)(1 - H(\xi_1))$ and $E(N_{11}) = N(1 - H(\xi_1))$ while $E(D_{21}) = N \int_0^{\xi_1} F^\circ \, dH$ and $E(W_1) = N \int_0^{\xi_1} (1 - F^\circ) \, dH$. When divided by N, these frequencies will converge a.s. to the corresponding probabilities. This provides a formal proof of consistency for the RS estimate and shows that the SD estimate is consistent if and only if it is consistent when all individuals are due for withdrawal in I_1, i.e., when $N_1 = N_{21}$. In this case the requirement for consistency becomes

$$F^\circ(\xi_1) = \frac{H^{-1}(\xi_1) \int_0^{\xi_1} F^\circ \, dH}{1 - \frac{1}{2}H^{-1}(\xi_1) \int_0^{\xi_1} (1 - F^\circ) \, dH} . \tag{4.2}$$

Insisting that this equation must hold for all choices of $0 < \xi_1 < T$ leads to severe restrictions on F° as shown by

THEOREM 1. *Let the censoring distribution H be absolutely continuous with density h on* $[0, T]$. *A necessary and sufficient condition that the* SD *estimate yield a consistent estimate of* F° *at the endpoints of each of the K intervals, for any choice of interval endpoints, is that* F° *satisfy*

$$F^\circ(x) = 1 - [1/(1 + cH(x))]^{\frac{1}{2}} \tag{4.3}$$

for some constant $c > 0$.

PROOF. For necessity, suppose the SD estimate is consistent at ξ_1 for any $0 < \xi_1 < T$. Then (4.2) must hold for all $\xi = \xi_1$ between 0 and T. Define $G(\xi) = H^{-1}(\xi) \int_0^{\xi} F^\circ(x)h(x) \, dx$. G satisfies the differential equation

$$\frac{G'(1 + G)}{G(1 - G)} = \frac{h}{H}$$

whose solution $\ln (G/(1 - G^2)) = \ln (cH)$ in terms of $F^\circ = 2G/(1 + G)$ is precisely (4.3).

Conversely, any F° satisfying (4.3) will yield a consistent estimate at ξ_1. Furthermore, (4.3) ensures that the conditional survival distributions over each of the $K - 1$ subsequent intervals satisfy an analogous condition with respect to the distribution of censoring variables in that interval, and this proves the theorem.

A good approximation to the censoring distribution in many cases will be the uniform $H(x) = x/T$, $0 < x < T$, in which case the distributions yielding a consistent estimate satisfy $F^\circ(x) = 1 - (1/(1 + cx))^{\frac{1}{2}}$. Similarly, if H places mass only at the interval midpoints as assumed by some of the cited authors, the requirement for consistency (4.2) becomes $F^\circ(2\xi_1) = 2F^\circ(\xi_1)/(1 + F^\circ(\xi_1))$. This has as a family of solutions

$$F^\circ(x) = 1 - 1/(1 + cx) , \qquad c > 0 . \tag{4.4}$$

The fact that the distribution (4.4) may be consistently estimated by the SD estimate in this case was previously noted by Elveback (1958, page 438).

Theorem 1 implies that the SD estimate is generally inconsistent, with the estimates of the conditional probabilities converging to values $q_k^* \neq q_k$ (cf. (5.3) below). The magnitude of the asymptotic bias, $b = \prod_{k=1}^{K} p_k^* - \prod_{k=1}^{K} p_k$, has been investigated by Crowley (1970) for the case of censoring variables uniformly distributed on $[0, T]$ and both exponential and uniform survival distributions. Following Elveback (1958), he chooses $\xi_K = \frac{5}{6}T$, adjusts the parameters of the survival distributions so that $1 - F^\circ(\xi_K) = \exp(-2) = .135335$, and selects equidistant interval endpoints. A summary of these results is presented in Table 1. The positive bias may be explained on the basis that the uniform and exponential distributions show, respectively, positive and zero aging, i.e., increasing and constant failure rates. However, the distribution (4.3) with $H(x) = x$, under which the SD estimate is consistent, has negative aging. Given that all three distributions have the same value at ξ_K, an individual withdrawn before ξ_K will have a greater probability of dying before withdrawal under the distribution (4.3) than for the exponential, and still less for the uniform. Since the SD estimate correctly estimates the probabilities q_k under (4.3), it may be expected to underestimate them for the other distributions and hence to overestimate the survival probability. However, the bias appears not to be serious, from a practical viewpoint, unless the number of intervals used is fewer than ten. Related results for the SD estimate are given by Littel (1952) and for some of the other estimates by Elveback (1958) and Crowley (1970).

TABLE 1

Asymptotic bias of the SD *estimate for estimating the survival probability* .13533 *for uniform withdrawals and uniform and exponential survival distributions*

Survival distribution	Number of intervals = K					
	1	5	10	20	40	80
Uniform	.20401	.01945	.00543	.00141	.00036	.00009
Exponential	.11189	.00626	.00160	.00040	.00010	.00002

5. Asymptotic normality of the SD estimate: large sample covariance structure. Denote by $N_{K+1} = N_K - D_K - W_k$ the number of individuals at risk of death beyond the last interval. Under the model of random censorship the random vector of frequencies

$$\mathbf{M} = (D_1, W_1, D_2, W_2, \cdots, D_K, W_K, N_{K+1})' \tag{5.1}$$

will have a multinomial distribution whose cell occupancy probabilities may be calculated from (2.2) and (2.3) as illustrated in Section 4. Let us denote this vector of probabilities (using an obvious notation) by

$$\mathbf{\Pi} = (\Pi_1^D, \Pi_1^W, \Pi_2^D, \Pi_2^W, \cdots, \Pi_K^D, \Pi_K^W, \Pi_{K+1}^K)' . \tag{5.2}$$

It follows that the distribution of $N^{-\frac{1}{2}}(\mathbf{M} - N\mathbf{\Pi})$ converges to a multivariate normal with mean $\mathbf{0}$ and covariance matrix $\Sigma = D_\Pi - \mathbf{\Pi\Pi}'$, where D_Π is the diagonal matrix with $\mathbf{\Pi}$ on the diagonal. Since the estimates $\hat{q}_k = D_k/(N_k - \frac{1}{2}W_k)$ are smooth functions of these frequencies, it follows further that the random vector $N^{\frac{1}{2}}(\hat{\mathbf{q}} - \mathbf{q}^*)$ will likewise converge in distribution to a multivariate normal having mean $\mathbf{0}$ and a covariance matrix which may be determined by the "δ-method" (Rao (1965), page 322). Here the asymptotic mean vector $\mathbf{q}^*$ has components

$$q_k^* = \Pi_k^D/(\Pi_k^N - \tfrac{1}{2}\Pi_k^W) \tag{5.3}$$

which of course will generally not equal the true underlying q_k. Finally, since the vector of estimated survival probabilities $\hat{\mathbf{P}} = (\hat{P}_1, \cdots, \hat{P}_K)'$ given by

$$\hat{P}_k = \prod_{j=1}^k (1 - \hat{q}_j), \tag{5.4}$$

is again a smooth function of $\hat{\mathbf{q}}$, a further application of the δ-method suffices to establish its asymptotic normality. Hence it remains only to carry out the covariance calculations.

Expanding in a Taylor's series

$$\hat{q}_k - q_k^* = \left(\frac{-\Pi_k^D}{(\Pi_k^N - \frac{1}{2}\Pi_k^W)^2}, \frac{1}{\Pi_k^N - \frac{1}{2}\Pi_k^W}, \frac{\frac{1}{2}\Pi_k^D}{(\Pi_k^N - \frac{1}{2}\Pi_k^W)^2}\right)\frac{1}{N} \times \begin{pmatrix} N_k - N\Pi_k^W \\ D_k - N\Pi_k^D \\ W_k - N\Pi_k^W \end{pmatrix} + o_p\left(\frac{1}{N^{\frac{1}{2}}}\right), \tag{5.5}$$

and similarly for $\hat{q}_l - q_l^*$, it follows that the asymptotic covariance of $N^{\frac{1}{2}}(\hat{q}_k - q_k^*)$ with $N^{\frac{1}{2}}(\hat{q}_l - q_l^*)$ is, for $k < l$, $(\Pi_k^N - \frac{1}{2}\Pi_k^W)^{-2}(\Pi_l^N - \frac{1}{2}\Pi_l^W)^{-2}$ times

$$(-\Pi_k^D, \Pi_k^N - \tfrac{1}{2}\Pi_k^W, \tfrac{1}{2}\Pi_k^D)\begin{pmatrix} 1 - \Pi_k^N \\ -\Pi_k^D \\ -\Pi_k^W \end{pmatrix}(\Pi_l^N, \Pi_l^D, \Pi_l^W)\begin{pmatrix} -\Pi_l^D \\ \Pi_l^N - \frac{1}{2}\Pi_l^W \\ \frac{1}{2}\Pi_l^D \end{pmatrix}. \tag{5.6}$$

The matrix formed by the product of the two middle terms in (5.6) is N^{-1} times the covariance matrix of (N_k, D_k, W_k) with (N_l, D_l, W_l). Since the product of the last two terms in (5.6) is zero, this proves that the individual estimates $\hat{q}_k$ are indeed asymptotically uncorrelated.

Essentially the same argument leads to the conclusion that the asymptotic variance of $N^{\frac{1}{2}}(\hat{q}_k - q_k^*)$ is given by

$$(\Pi_k^N - \tfrac{1}{2}\Pi_k^W)^{-4}(-\Pi_k^D, \Pi_k^N - \tfrac{1}{2}\Pi_k^W, \tfrac{1}{2}\Pi_k^D)\begin{pmatrix} 0 & 0 & 0 \\ \Pi_k^D & \Pi_k^D & 0 \\ \Pi_k^W & 0 & \Pi_k^W \end{pmatrix}\begin{pmatrix} -\Pi_k^D \\ \Pi_k^N - \frac{1}{2}\Pi_k^W \\ \frac{1}{2}\Pi_k^D \end{pmatrix}$$

$$= (\Pi_k^N - \tfrac{1}{2}\Pi_k^W)^{-4}[\Pi_k^D(\Pi_k^N - \tfrac{1}{2}\Pi_k^W)^2 + (\tfrac{1}{2}\Pi_k^D)^2\Pi_k^W - (\Pi_k^D)^2\Pi_k^N]$$

$$\begin{aligned} &\mathrm{Var}_A\,[N^{\frac{1}{2}}(\hat{q}_k - q_k^*)] \\ &\quad = (\Pi_k^N - \tfrac{1}{2}\Pi_k^W)^{-1}[q_k^* - (q_k^*)^2((\Pi_k^N - \tfrac{1}{4}\Pi_k^W)/(\Pi_k^N - \tfrac{1}{2}\Pi_k^W))]. \end{aligned} \tag{5.7}$$

These results may be summarized in

THEOREM 2. *The normalized vector* $N^{\frac{1}{2}}(\hat{\mathbf{q}} - \mathbf{q}^*)$ *of the* SD *estimates of the conditional probabilities of death has a limiting multivariate normal distribution with mean* $\mathbf{0}$ *and a diagonal covariance matrix, the elements of which are given by* (5.7).

It is of interest to compare (5.7) with the classical formula used routinely to estimate the variance of $\hat{q}_k$. Derived by analogy with binomial sampling theory, this takes the form

$$\widehat{\mathrm{Var}}\,(\hat{q}_k) = \hat{q}_k \hat{p}_k / N_k' = (N_k')^{-1}(\hat{q}_k - \hat{q}_k^2)\,, \tag{5.8}$$

where $N_k' = N_k - \frac{1}{2}W_k$ is the term appearing in the denominator of the SD estimate. N_k' is often thought of as the effective number exposed to the risk of death in I_k, which explains the origin of (5.8). Since N_k'/N is a consistent estimate of $\Pi_k^N - \frac{1}{2}\Pi_k^W$, (5.8) is seen to overestimate the true limiting variance of $\hat{q}_k$. Chiang (1968, page 275) observes a similar phenomenon for the estimate he proposes. If the kth interval is sufficiently short that $\hat{q}_k^2$ is small compared with $\hat{q}_k$, then the leading terms of (5.8) and (5.7) will predominate. This is important since the former is a consistent estimate of the latter, indicating that the error in (5.8) may not be too serious if, once again, the number of intervals is reasonably large.

One further routine application of the δ-method leads us to

COROLLARY 1. *The limiting distribution of* $N^{\frac{1}{2}}(\hat{\mathbf{P}} - \mathbf{P}^*)$ *for the* SD *estimate is multivariate normal with mean* $\mathbf{0}$ *and a covariance matrix whose* (k, l) *term for* $k \leq l$ *is*

$$P_k^* P_l^* \textstyle\sum_{j=1}^{k} \mathrm{Var}_A\,[N^{\frac{1}{2}}(\hat{q}_j - q_j^*)]/(1 - q_j^*)^2\,, \tag{5.9}$$

the asymptotic variance being given by (5.7).

Formula (5.9), with (5.8) used in place of (5.7) to estimate $\mathrm{Var}\,(\hat{q}_k)$ and with $\hat{P}_k$ substituted for P_k^*, is the classical approximation established by Greenwood (1926). In view of the preceding discussion, it too, from the viewpoint of asymptotic theory, will yield a slight overestimate of the variation in the estimated survival probability.

6. The product limit (PL) estimate: heuristic approach. Let us denote explicitly the dependence of the life table estimate (5.4) on the partition $0 < \xi_{1,K} < \cdots < \xi_{K,K}$ and sample size N by writing

$$1 - \hat{F}^{\circ}_{K,N}(t) = \hat{P}_k \qquad \text{for} \quad t \in [\xi_{k,K}, \xi_{k+1,K}) \qquad k = 1, \cdots, K-1\,. \tag{6.1}$$

Kaplan and Meier (1958) studied the product limit (PL) estimate $\hat{F}_N{}^{\circ} = \lim_{K\to\infty} \hat{F}^{\circ}_{K,N}$, where the right continuous limit is taken under any nested sequence of partitions such that $\sup_{1\leq k\leq K} |\xi_{k,K} - \xi_{k-1,K}| \to 0$. In calculating the limit they adopted the convention, also used here, of adjusting each of the censoring variables an infinitesimal amount to the right. Thus withdrawals occurring at the endpoint ξ_k do not contribute to W_k. This is reasonable since these individuals are not at risk for even a portion of the interval I_k and hence should not be used in determining the "effective number at risk." Likewise, uncensored observations are ranked ahead of censored observations with which they are tied. With these adjustments to the data and with the $\hat{q}_k$ calculated according to any of the methods previously mentioned, the limit becomes

$$(6.2)\qquad \begin{aligned} 1 - F_N{}^\circ(t) &= \prod_{n:\, X_n \leqq t} [(N - R_n)/(N - R_n + 1)]^{\delta_n}\,, & t < X_{(N)} \\ &= 0\,, & t \geqq X_{(N)} \end{aligned}$$

where $X_{(N)} = \max(X_1, \cdots, X_N)$ and R_n is the rank of $(X_n, 1 - \delta_n)$ in the lexicographic ordering of the sequence $(X_1, 1 - \delta_1), \cdots, (X_N, 1 - \delta_N)$.

A heuristic derivation of the large sample properties of $\hat{F}_N{}^\circ$ can be given from the corresponding properties of the life table estimate studied in Sections 4 and 5, merely by interchanging the two limiting operations $K \to \infty$ and $N \to \infty$. For example, since the PL estimate is the limit as $K \to \infty$ of the RS estimate, and since this latter is consistent as $N \to \infty$, it "follows" that the PL estimate will be consistent as well. To obtain the asymptotic distribution define the stochastic process $Z_N{}^*(t) = N^{\frac{1}{2}}(\hat{F}_N{}^\circ(t) - F^\circ(t))$ for $0 < t < T$. Choose a sequence of partitions and let k, l and K approach infinity in such a way that $s \in I_k$ and $t \in I_l$. In view of Corollary 1 it follows upon interchanging the two limits that the finite dimensional distributions of $Z_N{}^*$ are asymptotically normal. It would follow further from this that $Z_N{}^*$ converges weakly to a mean zero Gaussian process Z^*, provided the property of uniform tightness could be established (Billingsley (1968)). The covariance structure of the limiting process Z^* can be formally obtained by taking the limit in (5.7) and (5.9). Since the bias of the life table estimate tends to 0 in the limit, we have $P_k{}^* \to 1 - F^\circ(t)$ and likewise $\Pi_k{}^N \to 1 - F(t - 0)$ and $q_k{}^* \to dF^\circ(t)/(1 - F^\circ(t - 0))$ so that for $s \leqq t$

$$(6.3)\qquad \begin{aligned} &\operatorname{Cov}(Z^*(s), Z^*(t)) \\ &\quad = (1 - F^\circ(s))(1 - F^\circ(t)) \int_0^s (1 - F(x - 0))^{-1} \\ &\qquad \times (1 - F^\circ(x))^{-1}\, dF^\circ(x)\,. \end{aligned}$$

It is easy to formalize this argument for the case of a discrete survival distribution taking values $t_1 < \cdots < t_K$, say. For, *redefining* $N_k = \sum_1^N I_{[X_n \geqq t_k]}$ and $D_k = \sum_1^N I_{[X_n = t_k, \delta_n = 1]}$ to be the number of individuals at risk and dying at t_k, respectively, it follows from the same multinomial arguments used in Section 5 that, almost surely as $N \to \infty$, $N_k/N \to \Pi_k{}^N = 1 - F(t_k - 0) = (1 - F^\circ(t_k - 0)) \times (1 - H(t_k - 0))$ and $D_k/N \to \Pi_k{}^D = P[X_n = t_k, Y_n \geqq t_k] = (F^\circ(t_k) - F^\circ(t_k - 0)) \times (1 - H(t_k - 0))$ so that $\hat{q}_k = D_k/N_k \to (F^\circ(t_k) - F^\circ(t_k - 0))/(1 - F^\circ(t_k - 0)) = q_k$. Working through equations (5.3)—(5.7) in this simpler situation shows that $N^{\frac{1}{2}}(\hat{\mathbf{q}} - \mathbf{q})'$ is again asymptotically normal, with mean $\mathbf{0}$ and a diagonal covariance matrix with diagonal elements $q_k p_k/\Pi_k{}^N$. Consequently the normalized vector of estimated survival probabilities, as in Corollary 1, will be multivariate normal with a covariance matrix given by

$$P_k P_l \sum_{j=1}^{k} q_j/(\Pi_j{}^N p_j)\,, \qquad k \leqq l\,. \tag{6.4}$$

But (6.4) is just (6.3) for the case of a discrete distribution.

7. Weak convergence of the cumulative hazard and PL processes: continuous case. For a number of technical reasons it is convenient to study the behavior of the PL estimate and its relative, the empirical cumulative hazard process, under the assumption that F° and H are continuous. It follows that F and $\tilde{F}$ are also continuous. This allows us to perform numerous changes of variable in integrations of the form $\int h(F^\circ)\,dF^\circ$ (cf. (7.5) and the Appendix). It also means that, almost surely, there will be no ties among the observations, so that the ranks R_n especially of the uncensored observations among $X_1, \cdots, X_N$ are uniquely defined, as required below by Lemma 1 and the following definition.

The empirical cumulative hazard process $\Lambda_N{}^e$ is defined by

$$\Lambda_N{}^e(t) = \sum_{X_n \leqq t} \delta_n/(N + 1 - R_n)\,, \tag{7.1}$$

where R_n is the rank of X_n among all N observations. While the PL estimate may be derived as the unrestricted maximum likelihood estimate of the cumulative distribution function, $\Lambda_N{}^e$ may be derived as the maximum likelihood estimate of the cumulative hazard function in the class of distributions which have constant hazard functions between each pair of uncensored observations (Breslow (1972)). The two estimates are related through the following inequality.

LEMMA 1. *Let $N(t) = \sum_{n=1}^{N} I_{\{X_n \geqq t\}}$ be the number of individuals still "at risk" at time t. Then with probability* 1 *for all* $0 < t < X_{(N)}$,

$$0 < -\ln(1 - \hat{F}_N{}^\circ(t)) - \Lambda_N{}^e(t) < (N - N(t))/(N \cdot N(t))\,. \tag{7.2}$$

PROOF. Using the elementary inequality[3] $0 < -\ln(1-(x+1)^{-1}) - (x+1)^{-1} < (x(x+1))^{-1}$, valid for $x > 0$, it follows upon substituting $x = N - R_n$ that

$$\begin{aligned}
0 &< -\ln \textstyle\prod_{X_n \leqq t} (1 - 1/(N + 1 - R_n))^{\delta_n} - \sum_{X_n \leqq t} \delta_n/(N + 1 - R_n) \\
&< \textstyle\sum_{X_n \leqq t} \delta_n/((N - R_n)(N + 1 - R_n)) \\
&\leqq \textstyle\sum_{n=1}^{N-N(t)} 1/((N - n)(N - n + 1)) \\
&< \textstyle\sum_{n=N(t)}^{N-1} n^{-2} < \int_{N(t)}^{N} y^{-2}\,dy = (N - N(t))/(N \cdot N(t))\,.
\end{aligned}$$

Denote by

$$F_N{}^e(t) = N^{-1} \sum_{n=1}^{N} I_{[X_n < t]} \tag{7.3}$$

and

$$\tilde{F}_N{}^e(t) = N^{-1} \sum_{n=1}^{N} I_{[X_n < t,\, \delta_n = 1]}\,,$$

respectively, the left continuous versions of the EDF of the observations and the sub-EDF of the uncensored observations. Then $\Lambda_N{}^e$ may be written

$$\Lambda_N{}^e(t) = \int_0^t (1 - F_N{}^e(y))^{-1}\,d\tilde{F}_N{}^e(y)\,. \tag{7.4}$$

[3] Due to Thomas (1972).

Since F_N^e and $\tilde{F}_N^e$ converge almost surely to F and $\tilde{F}$ as defined by (2.2) and (2.3), we expect Λ_N^e to converge to

$$\Lambda = \Lambda(t) = \int_0^t (1 - F)^{-1}\, d\tilde{F} = \int_0^t (1 - F^\circ)^{-1}\, dF^\circ = -\ln(1 - F^\circ(t))\,, \tag{7.5}$$

i.e., we expect Λ_N^e to consistently estimate the cumulative hazard function of the underlying distribution F°.

The relevance of (7.4) is that it allows us to study the limiting distribution of Λ_N^e in terms of the joint limiting distribution of F_N^e and $\tilde{F}_N^e$. These latter are treated as random functions in $D[0, T]$, the well-known space of functions on $[0, T]$ having jump discontinuities, using the Skorohod topology (Billingsley (1968)). Here T is any finite value such that $1 - F(T) > 0$. The pertinent facts concerning the joint distribution of F_N^e and $\tilde{F}_N^e$ are summarized in

THEOREM 3. *Define* $(X_N, Y_N) \in D[0, T] \times D[0, T]$ *by* $X_N = N^{\frac{1}{2}}(F_N^e - F)$ *and* $Y_N = N^{\frac{1}{2}}(\tilde{F}_N^e - \tilde{F})$. *Then* (X_N, Y_N) *converges weakly to a bivariate Gaussian process* (X, Y) *which has mean* $\mathbf{0}$ *and a covariance structure given for* $s \leqq t$ *by*

$$\begin{aligned} \operatorname{Cov}(X(s), X(t)) &= F(s)(1 - F(t)) \\ \operatorname{Cov}(Y(s), Y(t)) &= \tilde{F}(s)(1 - \tilde{F}(t)) \\ \operatorname{Cov}(Y(s), X(t)) &= \tilde{F}(s)(1 - F(t)) \\ \operatorname{Cov}(X(s), Y(t)) &= \tilde{F}(s) - F(s)\tilde{F}(t)\,. \end{aligned} \tag{7.6}$$

PROOF. The fact that the finite dimensional distributions of (X_N, Y_N) are multivariate normal with covariance structure (7.6) follows, as with the EDF, from the representation

$$(X_N(t), Y_N(t)) = N^{\frac{1}{2}} \sum_{n=1}^N (I_{[X_n < t]} - F(t), I_{[X_n < t, \delta_n = 1]} - \tilde{F}(t)) \tag{7.7}$$

of (X_N, Y_N) as the normalized sum of i.i.d. processes. Hence it remains only to prove tightness. However it is well known (Billingsley (1968, Theorem 16.4)) that the sequence of distributions induced by X_N is tight, and the same convergence criteria apply equally well to the Y_N. Consequently (X_N, Y_N) induces a tight sequence of distributions on the product space $D[0, T] \times D[0, T]$.

For technical reasons it will also prove helpful to introduce the special Skorohod (1956) constructions, as elucidated by Pyke and Shorack (1968), and replace (X_N, Y_N) and (X, Y) with a sequence of random functions having the same distributions for each N, but which satisfy also

$$\rho((X_N, Y_N), (X, Y)) \to_{\text{a.s.}} 0\,, \tag{7.8}$$

where ρ is the Skorohod metric on $D[0, T] \times D[0, T]$. As explained by Pyke and Shorack, conclusions regarding limiting distributions of functions of the specially constructed processes will apply equally well to the same functions of the original processes. Thus, in what follows, think of $\Lambda_N^e = \Lambda_N^e(X_N, Y_N)$ as a mapping from $D[0, T] \times D[0, T]$ to $D[0, T]$. Assume that the arguments (X_N, Y_N) satisfy (7.8) and that (X, Y) is the Gaussian process defined in (7.6). We then have the expansion

$$N^{\frac{1}{2}}(\Lambda_N^e - \Lambda) = A_N + B_N + R_{1N} + R_{2N}$$

where

$$\begin{aligned} A_N(t) &= \textstyle\int_0^t X_N(1 - F)^{-2}\, d\tilde{F} \\ B_N(t) &= Y_N(t)(1 - F(t))^{-1} - \textstyle\int_0^t Y_N(1 - F)^{-2}\, dF \\ R_{1N}(t) &= N^{-\frac{1}{2}} \textstyle\int_0^t X_N^2(1 - F)^{-2}(1 - F_N^e)^{-1}\, d\tilde{F} \\ R_{2N}(t) &= \textstyle\int_0^t X_N(1 - F)^{-1}(1 - F_N^e)^{-1}\, d(\tilde{F}_N^e - \tilde{F})\,. \end{aligned} \tag{7.9}$$

and

We will show that A_N and B_N converge a.s. in ρ_T, the supremum metric on $[0, T]$, to random functions A and B defined by $A(t) = \int_0^t X(1 - F)^{-2}\, d\tilde{F}$ and $B(t) = Y(t)(1 - F(t))^{-1} - \int_0^t Y(1 - F)^{-2}\, dF$, respectively. Likewise we will show that R_{1N} and R_{2N} converge in probability to 0. Since convergence to a continuous limit (such as A, B, X, Y and 0 are a.s.) in ρ_T is equivalent to convergence in ρ, this will establish

THEOREM 4. *Let $T < \infty$ satisfy $F(T) < 1$ and suppose F° and H are continuous. Then the random function $N^{\frac{1}{2}}(\Lambda_N^e - \Lambda)$ converges weakly to the Gaussian process Z defined by*

$$Z(t) = \textstyle\int_0^t X(1 - F)^{-2}\, d\tilde{F} + Y(t)(1 - F(t))^{-1} - \int_0^t Y(1 - F)^{-2}\, dF\,, \tag{7.10}$$

where (X, Y) is the bivariate mean **0** *Gaussian process satisfying* (7.6). *Furthermore, the covariance structure of the limiting process Z is given for $s \leqq t$ by*

$$\operatorname{Cov}(Z(s), Z(t)) = \textstyle\int_0^s (1 - F)^{-2}\, d\tilde{F} = \int_0^s (1 - F)^{-1}(1 - F^\circ)^{-1}\, dF^\circ\,. \tag{7.11}$$

PROOF. To prove limiting normality it suffices to examine the convergences mentioned above. Since $\rho_T(A_N, A) \leqq \rho_T(X_N, X)\cdot \int_0^T (1-F)^{-2} d\tilde{F}$ and $\rho_T(B_N, B) \leqq \rho_T(Y_N, Y)[(1 - F(T))^{-1} + \int_0^T (1 - F)^{-2}\, dF]$, convergence of the two leading terms in (7.9) follows straight away from (7.8). Turning to the remainder terms, $\rho_T(R_{1N}, 0) \leqq N^{-\frac{1}{2}}(1 - F(T))^{-2}(1 - F_N^e(T))\rho_T^2(X_N, 0)$. Since the distributions induced by the X_N are tight and since $F_N^e(T) \to F(T)$ a.s., the last two terms in this expression are bounded in probability. Hence $\rho_T(R_{1N}, 0) = o_p(1)$. Next $\rho_T(R_{2N}, 0) \leqq 2\rho_T(X_N, X)(1 - F(T))^{-1}(1 - F_N^e(T))^{-1} + \sup_{0\leqq t\leqq T} |\int_0^t X(1 - F)^{-1} \times [(1 - F_N^e)^{-1} - (1 - F)^{-1}]\, d(\tilde{F}_N^e - \tilde{F})| + \sup_{0\leqq t\leqq T} |\int_0^t X(1 - F)^{-2}\, d(\tilde{F}_N^e - \tilde{F})|$. The first term is $o_p(1)$ in view of (7.8) and the a.s. boundedness of $(1 - F_N^e(T))$. The second term is bounded by $2N^{-\frac{1}{2}}\rho_T(X, 0)\rho_T(X_N, 0)(1 - F(T))^{-2}(1 - F_N^e(T))^{-1}$ and is thus $o_p(1)$ by the same argument. For the third term, consider a subset Ω_0 of the underlying probability space such that (i) $P(\Omega_0) = 1$ and (ii) for $\omega \in \Omega_0$, X is uniformly continuous on $[0, T]$ while $\rho_T(\tilde{F}_N^e, \tilde{F}) = N^{-\frac{1}{2}}\rho_T(Y_N, 0) \to 0$. Choose a partition (depending on ω) of $[0, T]$ into K intervals $I_k = (\xi_{k-1}, \xi_k]$ such that $\sup_{t\in I_k} |X(t)(1 - F(t))^{-2} - X(\xi_k)(1 - F(\xi_k))^{-2}| < \varepsilon$ for $k = 1, \cdots, K$. Then we have the third term bounded by $2\varepsilon + [(k - 1)\varepsilon + \rho_T(X(1 - F)^{-2}, 0)]\rho_T(\tilde{F}_N^e, \tilde{F})$, which tends to 2ε as $N \to \infty$. Since ε is arbitrary, this shows that the third term also converges to 0 a.s. and completes the proof of normality.

The evaluation of the covariance structure of the limiting process consists of a lengthy and tedious but straightforward calculation of the covariances for the additive terms (7.10) which make up $Z(t)$ and $Z(s)$, using (7.6) and repeated integration by parts. It is given in detail in the Appendix.

The asymptotic normality of the PL estimate follows from Theorem 4 upon one further application of the δ-method. For thinking now also of $\hat{F}_N^\circ$ as a random function in $D[0, T]$, we have from (7.2) and (7.5) on the set $[X_{(N)} > T]$,

$$(7.12)\qquad \begin{aligned} N^{\frac{1}{2}}(\hat{F}_N^\circ - F^\circ) &= -N^{\frac{1}{2}}(e^{-\Lambda_N^e} - e^{-\Lambda}) - N^{\frac{1}{2}}[\exp(\ln(1 - \hat{F}_N^\circ)) - e^{-\Lambda_N^e}] \\ &= -e^{-\Lambda}N^{\frac{1}{2}}(\Lambda_N^e - \Lambda) - e^{-\Lambda_N^*}N^{\frac{1}{2}}(\Lambda_N^e - \Lambda)^2 \\ &\quad + N^{\frac{1}{2}}e^{-\Lambda_N^{**}}(-\ln(1 - \hat{F}_N^\circ) - \Lambda_N^e) \end{aligned}$$

where $\rho_T(\Lambda_N^*, \Lambda) \leqq \rho_T(\Lambda_N^e, \Lambda)$ and, from Lemma 1, $\rho_T(\Lambda_N^{**}, \Lambda_N^e) \leqq \rho_T(-\ln(1 - \hat{F}_N^\circ), \Lambda_N^e) \leqq N^{-1}N^{-1}(T)(N - N(T))$. The two remainder terms converge to 0 in probability while the leading term converges in distribution to the Gaussian process $Z^*(t) = -(1 - F^\circ(t))Z(t)$. Since the set $[X_{(N)} > T]$ has probability one in the limit, this proves

THEOREM 5. *Let $T < \infty$ satisfy $F(T) < 1$ and suppose F° and H are continuous. Then the random function $N^{\frac{1}{2}}(\hat{F}_N^\circ - F^\circ)$, for $0 < t < T$, converges weakly to a mean 0 Gaussian process $Z^*(t)$ with*

$$(7.13)\qquad \operatorname{Cov}(Z^*(s), Z^*(t)) = (1 - F^\circ(s))(1 - F^\circ(t)) \textstyle\int_0^s (1 - F)^{-2}\, d\tilde{F}, \quad s \leqq t.$$

If the distribution H of censoring variables has support on all of $(0, \infty)$, then $F(t) < 1$ for all $0 < t < \infty$. In this case Theorem 5 can probably be extended to yield weak convergence on the entire half line as required in one of the applications suggested below. However, in most realistic applications the censoring variables will be bounded and as T approaches the upper limit of the range of observation, the variance of the limiting process increases to $+\infty$ unless the limit is also the limit of support of the underlying F°. This suggests that care be taken in applying the result in a region where only a few uncensored observations are available.

8. Applications. We outline briefly a few of the possible applications of the preceding. According to Theorem 5, any continuous functional of $\hat{F}_N^\circ$ will have, when appropriately normalized, a limiting distribution which can be determined from the distribution of Z^*. Since many statistics calculated from simple random samples can be expressed as linear functionals of EDF's, they are easily generalized to the censored data case by substituting PL estimates for the EDF's. The asymptotic distribution of such a statistic will be normal with a variance which may be estimated by substituting F_N^e, $\tilde{F}_N^e$ and $\hat{F}_N^\circ$ for F, $\tilde{F}$ and F° in (7.13). Thus, for example, providing that H has support on all of $(0, \infty)$, the moments of the underlying distribution may be estimated by integrating the following expression by parts:

$$(8.1)\qquad \hat{u}_N^r = \textstyle\int_0^\infty x^r\, d\hat{F}_N^\circ(x).$$

The asymptotic variance of the mean, estimated from (8.1) with $r = 1$, is for example

$$\text{Var}_A \hat{u}_N^1 = \frac{1}{N} \int_0^\infty (1 - F(s))^{-2} [\int_s^\infty (1 - F^\circ(t))\, dt]^2\, d\tilde{F}(s) , \tag{8.2}$$

which reduces to the usual σ^2/N in case $\tilde{F} = F = F^\circ$, i.e., there is no censoring.

Kimura (1973) uses the PL estimate to obtain smoothed estimates of the underlying survival distributions by means of orthogonal series expansions. Following Kronmal and Tarter (1968), he writes

$$f^\circ(t) = \frac{d}{dt} F^\circ(t) = \sum_{k=1}^\infty a_k \psi_k(t) \tag{8.3}$$

where the $[\psi_k]$ form an orthogonal series of functions on $[0, T]$. Coefficients in this expansion are estimated by

$$\hat{a}_k = \int_0^T \psi_k\, d\hat{F}_N^\circ . \tag{8.4}$$

Theorem 5 is useful in establishing asymptotic variances for the estimated coefficients, which can then be used in a selection rule to choose the terms in the asymptotic expansion (8.4) which are actually used for estimation of the density.

It is not inconceivable that Theorem 5 could be used to investigate asymptotic properties of a Kolmogorov–Smirnov type test appropriately generalized to censored data.

Finally, in a two sample application, Efron (1967) obtains a generalization of Wilcoxon's statistic by setting

$$W = \int_0^T \hat{F}_N^\circ\, d\hat{G}_M^\circ , \tag{8.5}$$

where $\hat{F}_N^\circ$ and $\hat{G}_M^\circ$ are PL estimates calculated from independent samples of size N and M.

9. The case of fixed censorship. It is in many ways more satisfying to regard the censoring variables Y_n as fixed numbers rather than as random variables: one would like to study the asymptotic properties of the life table and PL estimates conditionally, in terms of the Y_n actually observed, rather than in terms of an unknown distribution. In conclusion we merely note that the previous results may indeed be extended to fixed censoring variables, provided that these behave in the limit as if they were a random sample from some distribution. Liapunov's version of the central limit theorem and a moment inequality established by Koul (1970) are useful in making these extensions. However, due to limitations of space, we leave the detailed arguments to the reader.

Acknowledgments. The authors wish to express their appreciation to Professor Galen Shorack for invaluable assistance in the construction of the proof of Theorem 4. We also appreciate the suggestion of a referee that Theorem 1 be generalized to arbitrary H (as stated) in place of the uniform distribution originally considered.

Note added in proof. While this article was in press the authors' attention was directed to the paper by Odd Aalen (1973) titled "Nonparametric inference in connection with multiple decrement models," Statistical Research Report No. 6, Department of Mathematics, University of Oslo. This paper, based in a competing risks framework, presents results similar to those outlined in Section 6 and derives a Kolmogorov–Smirnov type test as suggested in Section 8.

APPENDIX

Covariance structure of the limiting process. Write $\mathrm{Cov}(Z(s), Z(t)) = \mathrm{Var}\, Z(s) + \mathrm{Cov}(Z(s), Z(t) - Z(s))$ where Z is the process defined in (7.10) and $s \leqq t$. We use repeatedly the relationships (7.6) and integration by parts. $\mathrm{Var}\, Z(s)$ may be expressed as the sum of the terms (A.1) through (A.6) below where, by convention, the variables of integration r and u satisfy $0 \leqq r \leqq u \leqq s$.

$$\mathrm{Var} \int_0^s X(1-F)^{-2}\, d\tilde{F}$$

$$= 2 \int_0^s \int_0^u \frac{1-(1-F^\circ(r))(1-H(r))}{(1-F(r))(1-F^\circ(r))(1-F^\circ(u))}\, dF^\circ(r)\, dF^\circ(u) \tag{A.1}$$

$$= 2 \int_0^s \frac{\ln(1-F^\circ(r)) - \ln(1-F^\circ(s))}{(1-F(r))(1-F^\circ(r))}\, dF^\circ(r) - \ln^2(1-F^\circ(s)) .$$

$$\mathrm{Var}\, Y(s)(1-F(s))^{-1} = \tilde{F}(s)(1-\tilde{F}(s))(1-F(s))^{-2} . \tag{A.2}$$

$$\mathrm{Var} - \int_0^s Y(1-F)^{-2}\, dF$$

$$= 2 \int_0^s \int_0^u \frac{\tilde{F}(r)(1-\tilde{F}(u))}{(1-F(r))^2(1-F(u))^2}\, dF(r)\, dF(u)$$

$$= 2 \int_0^s (1-\tilde{F})(1-F)^{-2}(\tilde{F}(1-F)^{-1} + \ln(1-F^\circ))\, dF \tag{A.3}$$

$$= \tilde{F}(s)(1-\tilde{F}(s))(1-F(s))^{-2} + \int_0^s (1-F)^{-1}(1-F^\circ)^{-1}\, dF^\circ$$

$$+ 2\ln(1-F^\circ(s))(1-\tilde{F}(s))(1-F(s))^{-1} - \ln^2(1-F^\circ(s)) .$$

$$2\,\mathrm{Cov}\left(\int_0^s X(1-F)^{-2}\, d\tilde{F},\ Y(s)(1-F(s))^{-1}\right)$$

$$= 2 \int_0^s \frac{\tilde{F}(r) - F(r)\tilde{F}(s)}{(1-F(r))^2(1-F(s))}\, d\tilde{F}(r) \tag{A.4}$$

$$= 2(1-F(s))^{-1}\left[\int_0^s \tilde{F}(1-F)^{-1}(1-F^\circ)^{-1}\, dF^\circ\right.$$

$$\left. - \tilde{F}(s) \int_0^s (1-F)^{-1}(1-F^\circ)^{-1}\, dF^\circ - \tilde{F}(s)\ln(1-F^\circ(s))\right] .$$

$$2\,\mathrm{Cov}\left(\int_0^s X(1-F)^{-2}\, d\tilde{F},\ -\int_0^s Y(1-F)^{-2}\, dF\right)$$

$$= -2 \int_0^s \int_0^u \frac{\tilde{F}(r) - F(r)\tilde{F}(u)}{(1-F(r))^2(1-F(u))^2}\, d\tilde{F}(r)\, dF(u)$$

$$- 2 \int_0^s \int_0^u \frac{\tilde{F}(r)(1-F(u))}{(1-F(r))^2(1-F(u))^2}\, dF(r)\, d\tilde{F}(u) \tag{A.5}$$

$$= -2(1-F(s))^{-1} \int_0^s \tilde{F}(1-F)^{-1}(1-F^\circ)^{-1}\, dF^\circ$$

$$+ 2[\tilde{F}(s)(1-F(s))^{-1} + \ln(1-F^\circ(s))]$$

$$\times \int_0^s (1-F)^{-1}(1-F^\circ)^{-1}\, dF^\circ .$$

(A.6) $$2\,\mathrm{Cov}\,(Y(s)(1-F(s))^{-1}, -\textstyle\int_0^s Y(1-F)^{-2}\,dF) = -2(1-\tilde F(s))(1-F(s))^{-1}[\tilde F(s)(1-F(s))^{-1} + \ln(1-F^\circ(s))]\,.$$

Addition of (A.1)—(A.6) yields the expression (7.11). It remains to show that $\mathrm{Cov}\,(Z(s), Z(t)-Z(s)) = 0$. For these calculations there are nine terms (B.1)—(B.9) for which, by convention, the range of integration is $0 \leqq r \leqq s \leqq u \leqq t$.

(B.1) $$\begin{aligned}&\mathrm{Cov}\,(\textstyle\int_0^s X(1-F)^{-2}\,d\tilde F, \int_s^t X(1-F)^{-2}\,d\tilde F)\\ &\quad = \int_s^t\int_0^s \frac{F(r)(1-F(u))}{(1-F(r))^2(1-F(u))^2}\,d\tilde F(r)\,d\tilde F(u)\\ &\quad = [\ln(1-F^\circ(s) - \ln(1-F^\circ(t))][\textstyle\int_0^s(1-F)^{-1}(1-F^\circ)^{-1}\,dF^\circ\\ &\qquad + \ln(1-F^\circ(s))]\,.\end{aligned}$$

(B.2) $$\begin{aligned}&\mathrm{Cov}\,(\textstyle\int_0^s X(1-F)^{-2}\,d\tilde F, Y(t)(1-F(t))^{-1} - Y(s)(1-F(s))^{-1})\\ &\quad = ((1-F(t))^{-1} - (1-F(s))^{-1})\textstyle\int_0^s \tilde F(1-F)^{-1}(1-F^\circ)^{-1}\,dF^\circ\\ &\qquad - (\tilde F(t)(1-F(t))^{-1} - \tilde F(s)(1-F(s))^{-1})\\ &\qquad \times[\textstyle\int_0^s(1-F)^{-1}(1-F^\circ)^{-1}\,dF^\circ + \ln(1-F^\circ(s))]\,.\end{aligned}$$

(B.3) $$\begin{aligned}&\mathrm{Cov}\,(\textstyle\int_0^s X(1-F)^{-2}\,d\tilde F, -\int_s^t Y(1-F)^{-2}\,dF)\\ &\quad = -\int_s^t\int_0^s \frac{\tilde F(r) - F(r)\tilde F(u)}{(1-F(r))^2(1-F(u))^2}\,d\tilde F(r)\,dF(u)\\ &\quad = -((1-F(t))^{-1} - (1-F(s))^{-1})\textstyle\int_0^s \tilde F(1-F)^{-1}(1-F^\circ)^{-1}\,dF^\circ\\ &\qquad + [\tilde F(t)(1-F(t))^{-1} + \ln(1-F^\circ(t)) - \tilde F(s)(1-F(s))^{-1}\\ &\qquad - \ln(1-F^\circ(s))][\textstyle\int_0^s(1-F)^{-1}(1-F^\circ)^{-1}\,dF^\circ\\ &\qquad + \ln(1-F^\circ(s))]\,.\end{aligned}$$

(B.4) $$\begin{aligned}&\mathrm{Cov}\,(Y(s)(1-F(s))^{-1}, \textstyle\int_s^t X(1-F)^{-2}\,d\tilde F)\\ &\quad = \tilde F(s)(1-F(s))^{-1}(\ln(1-F^\circ(s)) - \ln(1-F^\circ(t)))\,.\end{aligned}$$

(B.5) $$\begin{aligned}&\mathrm{Cov}\,(Y(s)(1-F(s))^{-1}, Y(t)(1-F(t))^{-1} - Y(s)(1-F(s))^{-1})\\ &\quad = \tilde F(s)(1-F(s))^{-1}((1-\tilde F(t))(1-F(t))^{-1}\\ &\qquad - (1-\tilde F(s))(1-F(s))^{-1})\,.\end{aligned}$$

(B.6) $$\begin{aligned}&\mathrm{Cov}\,(Y(s)(1-F(s))^{-1}, -\textstyle\int_s^t Y(1-F)^{-2}\,dF)\\ &\quad = -\tilde F(s)(1-F(s))^{-1}\textstyle\int_s^t(1-\tilde F)(1-F)^{-2}\,dF\\ &\quad = -\tilde F(s)(1-F(s))^{-1}[(1-\tilde F(t))(1-F(t))^{-1}\\ &\qquad - (1-\tilde F(s))(1-F(s))^{-1} - \ln(1-F^\circ(t))\\ &\qquad + \ln(1-F^\circ(s))]\,.\end{aligned}$$

(B.7) $$\begin{aligned}&\mathrm{Cov}\,(-\textstyle\int_0^s Y(1-F)^{-2}\,dF, \int_s^t X(1-F)^{-2}\,d\tilde F)\\ &\quad = -\int_s^t\int_0^s \frac{\tilde F(r)(1-F(u))}{(1-F(r))^2(1-F(u))^2}\,dF(r)\,d\tilde F(u)\\ &\quad = [\ln(1-F^\circ(t)) - \ln(1-F^\circ(s))]\textstyle\int_0^s \tilde F(1-F)^{-2}\,dF\,.\end{aligned}$$

$$\text{(B.8)}\quad \begin{aligned}&\operatorname{Cov}\left(-\textstyle\int_0^s Y(1-F)^{-2}\,d\tilde{F},\ Y(t)(1-F(t))^{-1}-Y(s)(1-F(s))^{-1}\right)\\ &\quad=-((1-\tilde{F}(t))(1-F(t))^{-1}\\ &\qquad-(1-\tilde{F}(s))(1-F(s))^{-1})\textstyle\int_0^s \tilde{F}(1-F)^{-2}\,dF\,.\end{aligned}$$

$$\text{(B.9)}\quad \begin{aligned}&\operatorname{Cov}\left(-\textstyle\int_0^s Y(1-F)^{-2}\,dF,\ -\int_0^t Y(1-F)^{-2}\,dF\right)\\ &\quad=\int_s^t\int_0^s \frac{\tilde{F}(r)(1-\tilde{F}(u))}{(1-F(r))^2(1-F(u))^2}\,dF(r)\,dF(u)\\ &\quad=((1-\tilde{F}(t))(1-F(t))^{-1}-(1-\tilde{F}(s))(1-F(s))^{-1}\\ &\qquad-\ln(1-F^\circ(t))+\ln(1-F^\circ(s)))\textstyle\int_0^s \tilde{F}(1-F)^{-2}\,dF\,.\end{aligned}$$

Note that the sum of (B.1)—(B.3) is zero as is the sum of (B.4)—(B.6) and (B.7)—(B.9). This completes the proof of Theorem 4.

REFERENCES

[1] BERKSON, J. (1954). Estimation of the interval rate in actuarial calculations (summary). *J. Amer. Statist. Assoc.* **47** 363.

[2] BERKSON, J. and GAGE, R. P. (1950). Calculation of survival rates for cancer. *Proc. Mayo Clinic* **25** 270-286.

[3] BILLINGSLEY, P. (1968). *Weak Convergence of Probability Measures.* Wiley, New York.

[4] BRESLOW, N. (1969). On large sample sequential analysis with applications to survivorship data. *J. Appl. Probability* **6** 261-274.

[5] BRESLOW, N. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* **57** 579-594.

[6] BRESLOW, N. (1972). Contribution to discussion on the paper of D. R. Cox cited below.

[7] CHIANG, C. L. (1960a). A stochastic study of the life table and its applications: I. Probability distributions of the biometric functions. *Biometrics* **16** 618-635.

[8] CHIANG, C. L. (1960b). A stochastic study of the life table and its applications: II. Sample variance of the observed expectation of life and other biometric functions. *Human Biol.* **32** 221-238.

[9] CHIANG, C. L. (1961). A stochastic study of the life table and its applications: III. The follow-up study with consideration of competing risks. *Biometrics* **17** 57-78.

[10] CHIANG, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics.* Wiley, New York.

[11] COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B* **34** 187-220.

[12] CROWLEY, J. (1970). A comparison of several life table methods. M. S. thesis. Univ. of Washington.

[13] CUTLER, S. J. and EDERER, F. (1958). Maximum utilization of the life table method in analyzing survival. *J. Chron. Dis.* **8** 699-712.

[14] EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **4** 831-853. Univ. of California Press.

[15] ELVEBACK, L. (1958). Estimation of survivorship in chronic disease: the "actuarial" method. *J. Amer. Statist. Assoc.* **53** 420-440.

[16] GEHAN, E. (1969). Estimating survival functions from the life table. *J. Chron. Dis.* **21** 629-644.

[17] GILBERT, J. P. (1962). Random censorship. Ph. D. thesis, Univ. of Chicago.

[18] GREENWOOD, M. (1926). The natural duration of cancer. *Reports on Public Health and Medical Subjects* **33.** H. M. Stationery Office, London.

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

[20] KIMURA, D. (1973). Ph. D. dissertation. Univ. of Washington.

[21] KOUL, H. L. (1970). Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.* **41** 1768–1773.

[22] KRONMAL, R. and TARTER, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Statist. Assoc.* **64** 925-952.

[23] LITTLE, A. S. (1952). Estimation of the *T*-year survival rate from follow-up studies over a limited period of time. *Human Biol.* **24** 87-116.

[24] PYKE, R. and SHORACK, G. (1968). Weak convergence of a two-sample empirical process and a new approach to the Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755–771.

[25] RAO, C. R. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.

[26] SKOROHOD, A. V. (1958). Limit theorems for stochastic processes. *Theor. Probability Appl.* (translation in SIAM) **1** 261-290.

[27] THOMAS, D. R. (1972). On a generalized Savage statistic for comparing two arbitrarily censored samples. Technical Report, Oregon State Univ.

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER
WORLD HEALTH ORGANIZATION
150 COURS ALBERT THOMAS
69008 LYON, FRANCE