A product-limit estimator for use with length-biased data

B.B. WINTER* and A. FÖLDES†

University of Ottawa and Hungarian Academy of Sciences

Key words and phrases: Product-limit estimator, length-biased data, nonparametric estimation, strong consistency, estimation for renewal processes, life testing, hazard-function estimation.

AMS 1980 subject classifications: Primary 62G05, 62N05; secondary 62M09, 60K05, 60F15.

ABSTRACT

The following life-testing situation is considered. At some time in the distant past, \( n \) objects, from a population with life distribution \( F \), were put in use; whenever an object failed, it was promptly replaced. At some time \( \tau \), long after the start of the process, a statistician starts observing the \( n \) objects in use at that time; he knows the age of each of those \( n \) objects, and observes each of them for a fixed length of time \( T \leq \infty \), or until failure, whichever occurs first. In the case where \( T \) is finite, some of the observations may be censored; in the case where \( T = \infty \), there is no censoring. The total life of an object in use at time \( \tau \) is a length-biased observation from \( F \). A nonparametric estimator of the (cumulative) hazard function is proposed, and is used to construct an estimator of \( F \) which is of the product-limit type. Strong uniform consistency results (for \( n \rightarrow \infty \)) are obtained. An "Aalen-Johansen" identity, satisfied by any pair of life distributions and their (cumulative) hazard functions, is used in obtaining rate-of-convergence results.

RÉSUMÉ

On considère la situation suivante dans le contexte des tests de durée de vie. Dans le passé lointain on a mis en fonction \( n \) objets d'une population. Leur durée de vie ont \( F \) pour fonction de répartition. Chaque fois qu'un objet tombait en panne on le remplacait immédiatement. Au moment \( \tau \), longtemps après le début du processus, un statisticien commence à observer les \( n \) objets qui sont en fonction à ce moment-là; il connaît l'âge de chacun de ces \( n \) objets et observe chacun d'entre eux durant une période de temps fixée, \( T \leq \infty \), ou jusqu'à ce qu'ils tombent en panne. Si \( T \) est fini certaines observations peuvent être censurées. On propose un estimateur nonparamétrique de la fonction de hasard (cumulative). Cet estimateur est utilisé pour construire un estimateur de \( F \) du type "product-limit". On obtient une convergence uniforme forte. Une identité de "Aalen-Johansen", qui est satisfaite par toute paire formée d'une fonction de répartition et de sa fonction de hasard (cumulative), est utilisée pour obtenir les résultats sur la vitesse de convergence.

1. INTRODUCTION AND SUMMARY

Consider a situation where, at some time in the distant past, \( n \) objects were put in use, one in each of \( n \) positions; e.g., bulbs in \( n \) street lamps. They came from a population of

*This author's research was partly supported by the Natural Sciences and Engineering Research Council of Canada.
†This author's research was partly supported by Hungarian National Foundation for Scientific Research grant no. 1808.
objects such that the *age at failure* of a randomly selected object can be represented by a random variable with distribution function $F$ (with finite mean $\mu$), where "age" is the length of time since the object was put in use. Whenever an object failed, it was promptly replaced by another member of the same population. At some time $\tau$ (0 being the time at which the initial batch of $n$ objects were put in use) a statistician arrives on the scene, intending to estimate the life distribution $F$. It is assumed that observation is restricted to the $n$ objects in use at time $\tau$, that the age of each of the $n$ objects in use at time $\tau$ is known, and that each of the $n$ objects is observed for a fixed length of time $T \leq \infty$, or until failure, whichever occurs first. (The assumption that ages at time $\tau$ are known is not unreasonable: this occurs if a record of the time of the most recent replacement is kept for each of the $n$ positions.) In the case where $T$ is finite, the observation period starts at time $\tau$ and stops at time $\tau + T$; that may cause some of the observations to be censored. In the case where $T = \infty$, there is no censoring.

The situation described above can be represented by $n$ independent renewal processes, one for each position, with underlying distribution function $F$. The age of an object at the start of observation is then represented by the backward recurrence time $q_T$, and the time from the start of observation until failure is represented by the forward recurrence time $\tau$. If $F$ is nonarithmetic (or, as some authors put it, nonlattice), then, for nonnegative $y$ and $z$,

$$\lim_{T \to \infty} \text{prob}[q_T \leq y, \tau \leq z] = \frac{1}{\mu} \int_{[0, y]} \{F(s + z) - F(s)\} \, ds,$$

(1.1)

as can be seen from the limit for $\text{prob}[\eta_r > y, \tau_r > z]$ given on p. 386 of Feller (1971) and p. 308 of Çinlar (1975). It is therefore reasonable to model the situation described in the first paragraph by saying that the data the statistician obtains from the $r$th object (among the $n$ in use at time $\tau$) consist of observed values of random variables $Y_r$, $\hat{Z}_r$, and $\Gamma_r$, where $Y_r$ and $\hat{Z}_r$ are random variables whose joint distribution function is given by the right side of (1.1), $\hat{Z}_r = \min\{Z_r, \tau\}$, and $\Gamma_r = I[Z_r \leq \tau]$; here $I$ is the indicator function, $Y_r$ represents the *age* of the $r$th object at the start of observation, $Z_r$ represents the *remaining life* of that object, from the start of observation and until failure, $\hat{Z}_r$ represents the *length of time that the object is under observation*, and $\Gamma_r$ indicates whether observation is terminated by censoring or failure.

Motivated by the situation described in the first paragraph, and by the modelling considerations in the second paragraph, we address a problem which can be formalized as follows.

**Problem 1.1.** Let $F$ be a distribution function with

$$F(0^-) = 0, \quad F(0) < 1, \quad \text{and} \quad \mu = \int_{\mathbb{R}} x \, dF(x) < \infty.$$

Let $\Lambda$ be the (cumulative) hazard function of $F$, defined by

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

Let $((Y_r, Z_r): r = 1, 2, \ldots)$ be a sequence of pairs of random variables on a probability space $(\Omega, \mathcal{F}, P)$, taking values in $[0, \infty)$, where the pairs are independent of each other and the joint distribution function of each pair is given by
\[
F_{Y, Z}(y, z) = \frac{1}{\mu} \int_{[0, y]} \{F(s + z) - F(s)\} \, ds
\]
for nonnegative \(y\) and \(z\), with \(F_{Y, Z}(y, z) = 0\) elsewhere. Furthermore, \(T\) is fixed positive real or \(+\infty\), \(\bar{Z}_r = \min\{Z_r, T\}\), and \(\Gamma_r = I[Z_r \leq T]\). The problem is to estimate \(F\) and \(\Lambda\) nonparametrically, using observed values of the random triples \((Y_1, \bar{Z}_1, \Gamma_1), \ldots, (Y_n, \bar{Z}_n, \Gamma_n)\), and to study the asymptotic behaviour of the estimator when \(n \to \infty\).

In the case \(T = \infty\), corresponding to no censoring, \(\bar{Z}_r = Z_r\) and \(\Gamma_r = 1\) for \(r = 1, 2, \ldots\); hence estimation is based, in that case, on the random pairs \((Y_1, Z_1), \ldots, (Y_n, Z_n)\). (Note that, although the assumption that \(F\) is nonarithmetic is used in obtaining the limit in (1.1), that assumption is not mentioned in Problem 1.1 and is not used for the results obtained herein.)

When the random variables described in Problem 1.1 are used to model the life-testing setup described in the first paragraph, the random variable \(W_r = Y_r + Z_r\) represents the total (as opposed to observed) life of the object in position \(r\). In the case \(T = \infty\), corresponding to no censoring, the statistician has access to the observed values of the random variables \(W_1, \ldots, W_n\). As will be noted in Section 3 (and as can be surmised from the form of \(\lim_{T \to \infty} \text{prob}[\eta_r + \xi_r \leq w]\) for nonarithmetic \(F\)), the common distribution function of the random variables \(W_1, \ldots, W_n\) is given by

\[
F_W(w) = \frac{1}{\mu} \int_{[0, w]} s \, dF(s).
\]

As \(F_W\) is the length-biased form of \(F\) [see, e.g., Patil and Rao (1977) and Stein and Dattero (1985)], it is clear that the problem under consideration is a variant of the problem of estimating \(F\) on the basis of "length-biased observations from \(F\)."

It is interesting to note that the random variables described in Problem 1.1 can also be used to model data obtained in some situations other than the life-testing setup described in the first paragraph. They occur, for instance, in the situation studied by Simon (1980).

In Section 2, we define estimators of \(\Lambda\) and \(F\) and state our principal results, which concern strong uniform consistency (as \(n \to \infty\)). The estimator of \(F\) is of the product-limit type and is derived from the estimator of \(\Lambda\). Proofs are given in Section 3. We use a result due to Gill (1981) to infer strong uniform consistency of our estimator of \(F\) from the strong uniform consistency of the estimator of \(\Lambda\). Rate-of-convergence results for the estimator of \(\Lambda\) propagate to the estimator of \(F\) by means of an identity, related to a result of Aalen and Johansen (1978), which is satisfied by any pair of life distributions and their hazard functions. In Section 4 we summarize some simulation results regarding the quality of our estimators, indicate some other estimators which could be used in the life-testing setup described above, and briefly mention other research related to the work reported herein.

In the case \(T = \infty\), corresponding to no censoring, one has access to the observed values of the r.v.'s \(W_1, \ldots, W_n\). Instead of the estimator of \(F\) described in Section 2, one could, in this case, use an estimator introduced by Cox (1969)—which is a particular case of a more general estimator subsequently introduced by Vardi (1985). Our product-limit estimator and the Cox-Vardi estimator are briefly compared in Section 4. (Whereas both the product-limit estimator and the Cox-Vardi estimator are available in the case where there is no censoring, the Cox-Vardi estimator cannot be adapted to the case \(T < \infty\) where censoring is present.)
2. DEFINITIONS AND RESULTS

2.1. Notation and Terminology.

The jump $G(t^+)-G(t^-)$ of a function $G$ at $t \in \mathbb{R}$ is denoted $\Delta G(t)$, the indicator of a set $A$ is $I_A$, or $[\cdots]$ when the set is an event $[\cdots]$, and $\bar{\mathbb{R}}$ denotes the extended real line. When $G$ is a random function with domain $\mathbb{R} \times \Omega$, $G(\cdot, \omega)$ denotes the realization corresponding to $\omega \in \Omega$ and $G(t)$ denotes the random variable corresponding to a fixed $t$; if $G$ has nondecreasing, right-continuous sample paths, $\int \cdot dG$ denotes the random variable whose value at $\omega \in \Omega$ is obtained by integrating with respect to the measure determined by $G(\cdot, \omega)$.

As frequently happens in survival analysis, we shall be encountering functions which are just like distribution functions of nonnegative random variables, except that they may carry a total mass which is less than 1. It is convenient to give a name to such functions.

**Terminology 2.1.** We say that a function $Q : \mathbb{R}^+ \to \mathbb{R}$ is a life distribution if it is the distribution function of a random variable taking values in $[0, m]$, i.e., if it is a probability (sub)distribution function and $Q(0^-) = 0$. We say that it is a proper life distribution if $Q(cO) = 1$.

Following Gill (1980), we define the hazard function of a life distribution, proper or not, as follows.

**Definition 2.2.** If $Q$ is a life distribution, its hazard function (h.f.) is the function $\mathcal{Q}: \mathbb{R} \to \mathbb{R}$ given by

$$\mathcal{Q}(t) = \int_{[0, t]} \frac{1}{1 - Q(x^-)} \, dQ(x).$$

What we call the hazard function of a life distribution is usually referred to in the literature as its cumulative hazard function. The qualifier "cumulative" can be eliminated without risk of confusion—just as one nowadays speaks of "distribution functions" where one used to refer to "cumulative distribution functions". [If $Q$ has a density $q$, then the function $x \mapsto q(x)/(1 - Q(x))$ is the hazard rate or failure rate function for $Q$.]

The mapping which sends a life distribution to its h.f. is one-to-one. For a characterization of hazard functions, and a derivation of an inversion formula which yields the life distribution corresponding to a h.f., see, e.g., Winter (1989a). Without recourse to those general considerations, one can establish the following result by straightforward computation.

**Proposition 2.3.** If $\mathcal{Q}$ is an increasing right-continuous step function with jumps at $0 \leq t_1 < \cdots < t_k$, such that $\mathcal{Q}(t) = 0$ for $t < t_1$, $\Delta \mathcal{Q}(t_i) < 1$ for $1 \leq i < k$, and $\Delta \mathcal{Q}(t_k) \leq 1$, then $\mathcal{Q}$ is the h.f. of the life distribution $Q$ given by $Q(t) = 1 - \prod_{t_i \leq t} (1 - \Delta \mathcal{Q}(t_i))$. Moreover, $Q$ is proper if and only if $\Delta \mathcal{Q}(t_k) = 1$.

2.2. The Estimators.

We shall motivate the definition of an estimator of the hazard function $\mathcal{Q}$ by considering the r.v.'s in Problem 1.1 in relation to the life-testing setup described at the beginning of Section 1.

Let $((Y_r, Z_r) : r = 1, 2, \ldots), \tilde{Z}_r,$ and $\Gamma_r$ be as in Problem 1.1, and consider a fixed posi-
In relation to the life-testing setup described in Section 1, $Y_r$ can be thought of as the age (at the start of observation) of the $r$th object (among the $n$ objects in use at the start of observation), and $Z_r$ as its remaining life (i.e., time from the start of observation until failure of that object). Clearly, the number of observed failures (“deaths”) at ages up to and including $t$ is represented by the r.v.

$$D_n(t) = \sum_{r=1}^{n} I[\Gamma_r = 1, Y_r + Z_r \leq t].$$

(2.1)

An object is at risk at an age $t$ if it has not failed prior to that age and it is actually under observation at that age. In the life-testing setup considered herein, the $r$th object comes under observation at age $Y_r$ and remains under observation until age $Y_r + Z_r$; it is thus under observation at ages $t$ such that $Y_r \leq t \leq Y_r + Z_r$. Furthermore, since the $r$th object fails at age $Y_r + Z_r$, it does not fail prior to any age $t$ which satisfies the inequality $Y_r \leq t \leq Y_r + Z_r$. Thus $[Y_r \leq t \leq Y_r + Z_r]$ is the event “the $r$th object is at risk at age $t$”; hence the number of objects at risk at age $t$ is represented by the random variable

$$R_n(t) = \sum_{r=1}^{n} I[Y_r \leq t \leq Y_r + Z_r].$$

(2.2)

Now, in light of a reasoning frequently used in survival analysis, it appears plausible that one should estimate $A$ by a step function $\hat{A}_n$ with jump $\Delta D_n(s)/R_n(s)$ at every age $s$ at which failures are observed and no jumps except at such ages; i.e.,

$$\hat{A}_n(t) = \int_{[0,t]} \frac{1}{R_n(s)} dD_n(s) = \int_{[0,t]} \frac{1}{(1/n)R_n(s)} d\left(\frac{1}{n} D_n\right)(s).$$

(2.3)

Then, if this step function is a h.f., it is quite natural to estimate $F$ by the life distribution for which $\hat{A}_n$ is the h.f.

In life-testing setups such as the classical competing-risks situation, the estimator arrived at by the reasoning which led to $\hat{A}_n$ is such that any realization of the estimator is indeed the h.f. of a life distribution. That may however fail to be the case in the situation considered herein. One could, for instance, encounter situations such as in the following illustration.

Suppose that three devices are in use at time $\tau$, that devices 1, 2, and 3 were put in use, respectively, at times $\tau - 1$, $\tau - 6$, and $\tau - 2$, and that the devices fail, respectively, at times $\tau + 2$, $\tau + 1$, and $\tau + 3$. Thus devices 1, 2, and 3 are at risk during the age intervals $[1, 3]$, [6, 7], and [2, 5], respectively. Therefore the graphs of $D_n, R_n$, and $\hat{A}_n$ corresponding to this outcome $\omega$ are as shown in Figure 1.

For the case illustrated above, the step function $\hat{A}_n(\cdot, \omega)$ has a jump of size 1 at $t_2 = 5$ as well as at $t_3 = 7$, which implies that it is not the h.f. of a life distribution. This problem occurs whenever there are two or more ages such that an object fails at that age, and it is the only object at risk at that age. The problem arises from the fact that, in the life-testing setup considered herein, it is possible to have an age interval such that there are objects at risk before and after that age interval, but no objects are at risk at any age within the interval. Such problems tend to disappear as $n$ becomes sufficiently large.

In order to avoid such problems altogether, we shall rely on an estimator which is a slight perturbation of the estimator defined in (2.3). By adding a small quantity $\theta_n$ to the denominator in the second expression in (2.3), we avoid some technical difficulties in consistency proofs as well as the potential small-sample problem illustrated above.
DEFINITION 2.4. Let \((\theta_n)_{n=1}^{\infty}\) be a sequence of positive reals such that \(\theta_n \to 0\), and put

\[
\Lambda_n(t) = \int_{[0,t]} \frac{1}{\theta_n + (1/n)R_n(s)} \, d\left(\frac{1}{n} D_n\right)(s) = \int_{[0,t]} \frac{1}{n\theta_n + R_n(s)} \, dD_n(s),
\]

where \(D_n\) and \(R_n\) are as in (2.1) and (2.2).

As will be shown, the perturbations \(\theta_n\) can be arbitrarily small; our consistency results are valid if, e.g., \(\theta_n = \theta_0 n^{-t}\), with \(\theta_0\) an arbitrarily small positive constant. Thus these perturbations are of no practical consequence; the estimator \(\Lambda_n\) specified in Definition 2.4 can be constructed so as to be—in practice—indistinguishable from the estimator \(\hat{\Lambda}_n\) described in (2.3).

It is easily seen that (as suggested by the “number of failures” and “number at risk” interpretations) \(\Delta D_n(t)/R_n(t) \leq 1\). The addition of a positive perturbation to the denominator then ensures that every jump of the step function \(\Lambda_n(\cdot, \omega)\) will in fact be strictly smaller than 1. Thus every realization \(\Lambda_n(\cdot, \omega)\) is a hazard function of the type considered in Proposition 2.3. Therefore \(\Lambda_n\) determines a life distribution.

DEFINITION 2.5. The estimator \(F_n\) is the life distribution whose h.f. is \(\Lambda_n\); more explicitly, for \(\omega \in \Omega\), \(F_n(\cdot, \omega)\) is the life distribution whose h.f. is the function \(\Lambda_n(\cdot, \omega)\), with \(\Lambda_n\) as in Definition 2.4.

Our results, stated below and proved in Section 3, pertain to the estimators \(\Lambda_n\) and \(F_n\) defined in Definitions 2.4 and 2.5. The estimator \(\Lambda_n\) is a step function, with jumps at ages at which failures are observed. Thus, in light of Proposition 2.3, \(F_n\) can be described as follows.
Proposition 2.6. The estimator $F_n$ is given by

$$F_n(t, \omega) = 1 - \prod_{t_i \leq t} \left( 1 - \frac{d_i}{n \theta_n + r_i} \right),$$

where $0 \leq t_1 < \cdots < t_k$ are the ages at which failures are observed, $d_i = \Delta D_n(t_i, \omega)$ is the number of failures seen to occur at age $t_i$, and $r_i = R_n(t_i, \omega)$ is the number of objects at risk at age $t_i$.

As will be noted in Section 3, the random variable $Y_r + Z_r$ is absolutely continuous. It follows that the occurrence of multiple failures at the same age has probability zero; hence every $d_i$ in Proposition 2.6 is, a.s., 1.

2.3. Results.

The statements of our principal results refer to the following two conditions, where $(\rho_n)^n_{n=1}$ is a sequence of positive reals:

Condition (a). $F(0) = 0, \sigma \in (0, \infty)$ is such that $F(\sigma^-) < 1$, and

$$\lim_{n \to \infty} \theta_n^{-2} \sqrt{n^{-1} \log n} = 0.$$

Condition (b). $F(0) = 0, \sigma \in (0, \infty)$ is such that

$$F(\sigma) < 1 \quad \text{and} \quad \int_{[0, \sigma]} x^{-1} dF(x) < \infty,$$

$$\lim_{n \to \infty} \theta_n \rho_n = 0, \quad \text{and} \quad \lim_{n \to \infty} \rho_n \theta_n^{-2} \sqrt{n^{-1} \log n} = 0.$$

The sequence $(\theta_n)^n_{n=1}$ satisfies the requirement in Condition (a) if, for instance, $\theta_n = \theta_0 n^{-1/(4 + \delta)}$ with $\delta > 0$. The sequences $(\theta_n)^n_{n=1}$ and $(\rho_n)^n_{n=1}$ satisfy the requirement in Condition (b) if, for instance, $\theta_n = \theta_0 (n^{-1} \log n)^{\delta}$ and $\rho_n = \rho_0 (n/\log n)^{\delta} (\log n)^{-\delta}$ with $\delta > 0$.

Our principal results assert the strong uniform consistency (with a rate factor, if suitable conditions are satisfied) of our estimators of $\Lambda$ and $F$.

Theorem A. If Condition (a) is satisfied,

$$\lim_{n \to \infty} \sup_{0 \leq t \leq \sigma} |\Lambda_n(t) - \Lambda(t)| = 0 \quad \text{a.s.}$$

If Condition (b) is satisfied,

$$\lim_{n \to \infty} \left( \rho_n \sup_{0 \leq t \leq \sigma} |\Lambda_n(t) - \Lambda(t)| \right) = 0 \quad \text{a.s.}$$

Theorem B. If Condition (a) is satisfied,

$$\lim_{n \to \infty} \sup_{0 \leq t \leq \sigma} |F_n(t) - F(t)| = 0 \quad \text{a.s.}$$

If Condition (b) is satisfied,

$$\lim_{n \to \infty} \left( \rho_n \sup_{0 \leq t \leq \sigma} |F_n(t) - F(t)| \right) = 0 \quad \text{a.s.}$$
We close this section by mentioning an identity—used in the proofs in Section 3—which is satisfied by pairs of life distributions and their hazard functions. This result, stated in Theorem C, asserts the general validity of an identity previously presented in some restricted contexts. Identities corresponding to the one in Theorem C were stated by Aalen and Johansen (1978) for an estimator applicable to data from continuous-time Markov chains, by Gill (1980) for the product-limit estimator in a random censorship situation (which, although very general, does not include the setup considered herein), and by Gill (1983) and Wellner (1985) for the product-limit estimator in the independent-competing-risks situation. Such an identity is in fact true in the general form given in Theorem C, as can be seen by arguing essentially as in Appendix 4 of Gill (1980); for details, see Winter (1989a).

**THEOREM C.** Let $Q$ and $R$ be life distributions, and let $\mathcal{Q}$ and $\mathcal{R}$ be their hazard functions. If $t \in [0, \infty)$ is such that $R(t) < 1$ and $\mathcal{Q}(t) < \infty$, then

$$
\frac{1 - Q(s)}{1 - R(s)} = 1 - \int_{[0, s]} \frac{1 - Q(x^-)}{1 - R(x)} d(\mathcal{Q} - \mathcal{R})(x) \quad \text{for} \quad 0 \leq s \leq t.
$$

3. PROOFS

3.1. Preliminaries.

The hypotheses and notation introduced in Sections 1 and 2 apply in this section. In what follows, we put $\bar{F}(t) = 1 - F(t^-)$ and $\bar{t} = \min(t, T)$, $t \in \mathbb{R}$, and observe the usual conventions that $0 \cdot (\pm \infty) = 0$, $/0 = 0$, $[a, b] = \emptyset$ when $b < a$, $[a, b] = \emptyset$ when $b \leq a$, and $F.$ denotes the distribution function of the random variable in the subscript. Whenever convenient, we write $Y, Z,$ and $W$ for $Y_1, Z_1,$ and $W_1$.

Throughout this section, we rely on the fact that $D_n$ can also be written as

$$
D_n(t) = \sum_{r=1}^{n} I[Z_r \leq T, Y_r + Z_r \leq t]
$$

(3.1)

and, as is easily seen, $R_n$ can be written as

$$
R_n(t) = \sum_{r=1}^{n} I[t - T \leq Y_r \leq t \leq Y_r + Z_r].
$$

(3.2)

To simplify the notation, we put

$$
C_n(s) = n\theta_n + R_n(s),
$$

(3.3)

so that

$$
\Lambda_n(t) = \int_{[0, t]} \frac{1}{C_n(s)} dD_n(s) = \int_{[0, s]} \frac{1}{(1/n)C_n(s)} d\left(\frac{1}{n} D_n\right)(s).
$$

(3.4)

The proofs rely on this representation of $\Lambda_n$—and on the approximation of $\Lambda_n$ which results when sample averages are replaced by the corresponding expectations, namely the function $\bar{\Lambda}_n$ given by

$$
\bar{\Lambda}_n(t) = \int_{[0, t]} \frac{1}{\mathcal{E}(1/n)\bar{C}_n(s)} d\left(\mathcal{E} \frac{1}{n} D_n\right)(s).
$$

(3.5)

Our proofs make use of the essentially renewal-theoretic results stated in Propositions 3.1 and 3.2; for a proof of Proposition 3.1, see Winter (1989b).
Proposition 3.1.

(a) $F_Y(t) = F_Z(t) = (1/\mu) \int_{0,\alpha} \{1 - F(s)\} \, ds$.

(b) $P[Y > y, Z > z] = (1/\mu) \int_{(y+z,\infty)} \{1 - F(s)\} \, ds$ for $y \geq 0, z \geq 0$.

(c) $F_W(w) = (1/\mu) \int_{[0,w]} s \, dF(s) = (1/\mu) \int_{[0,w]} \{F(w) - F(s)\} \, ds$.

(d) The joint distribution of $Y$ and $Z$ is the same as the joint distribution of $UV$ and $(1 - U)V$, where $U$ and $V$ are independent positive random variables, with $U \sim \text{uniform}(0,1)$ and $F_U(v) = (1/\mu) \int_{[0,v]} s \, dF(s)$.

Proposition 3.2. The joint distribution of $Y$ and $W$ is such that

$$P[Y \leq \alpha, W \leq \beta] = \frac{1}{\mu} \int_{[0,\alpha]} \{F(\beta) - F(s)\} \, ds$$

when $0 \leq \alpha \leq \beta < \infty$.

Proof. Consider $0 \leq \alpha \leq \beta < \infty$, and let $U$ and $V$ be random variables as in Proposition 3.1(d), defined on $(\Omega', \mathcal{F}', P')$. Then

$$P[Y \leq \alpha, W \leq \beta] = P'[UV \leq \alpha, UV + (1 - U)V \leq \beta]$$

$$= \int_{[0,\alpha]} \int_{\mathbb{R}} P'[0 \leq U \leq \frac{\alpha}{V}] \, dF_U(u) \, dF_V(v)$$

$$= \int_{[0,\alpha]} \frac{\alpha}{\mu} \, dF(v) + \int_{(\alpha,\beta]} \frac{\alpha}{\mu} \, dF(V)$$

$$= \frac{1}{\mu} \int_{[0,\alpha]} \{F(\alpha) - F(s)\} \, ds + \frac{1}{\mu} \{F(\beta) - F(\alpha)\} \tag{using Proposition 3.1(c)}$$

$$= \frac{1}{\mu} \int_{[0,\alpha]} \{F(\alpha) - F(s) + F(\beta) - F(\alpha)\} \, ds$$

$$= \frac{1}{\mu} \int_{[0,\alpha]} \{F(\beta) - F(s)\} \, ds.$$

Q.E.D.

3.2. Lemmas.

Lemma 3.3.

$$\mathcal{C} \frac{1}{n} C_n(t) = \theta_n + \frac{t}{\mu} \bar{F}(t) \quad \text{when} \quad 0 \leq t < \infty.$$

Proof. Since $(1/n)C_n = \theta_n + (1/n)R_n$, it suffices to compute $\mathcal{C}(1/n)R_n(t)$. From (3.2),

$$\mathcal{C} \frac{1}{n} R_n(t) = P[t - T \leq Y \leq t \leq W]$$

$$= P[t - T \leq Y \leq t] - P[t - T \leq Y \leq t, W < t]$$

$$= P[t - T \leq Y \leq t] - (P[Y \leq t, W < t] - P[Y < t - T, W < t]). \tag{i}$$
By Proposition 3.1(a) and the monotone convergence theorem,

\[ P(t - T \leq Y \leq t) = \frac{1}{\mu} \int_{[0,t]} \{1 - F(s)\} \, ds - \frac{1}{\mu} \int_{[0,t-T]} \{1 - F(s)\} \, ds. \tag{ii} \]

By Proposition 3.2 and monotone convergence,

\[ P(Y \leq t, W < t) = \frac{1}{\mu} \int_{[0,t]} \{F(t^-) - F(s)\} \, ds \tag{iii} \]

and

\[ P(Y < t - T, W < t) = \frac{1}{\mu} \int_{[0,t-T]} \{F(t^-) - F(s)\} \, ds. \tag{iv} \]

Therefore, by (i)-(iv),

\[ \mathbb{E} \left( \frac{1}{n} R_n(t) \right) = \frac{1}{\mu} \int_{[0,t]} \{1 - F(t^-)\} \, ds - \frac{1}{\mu} \int_{[0,t-T]} \{1 - F(t^-)\} \, ds. \]

Thus if \( t - T \leq 0 \) then

\[ \mathbb{E} \left( \frac{1}{n} R_n(t) \right) = \frac{1}{\mu} \int_{[0,t]} \{1 - F(t^-)\} \, ds = \frac{t}{\mu} \{1 - F(t^-)\} = \frac{t}{\mu} \{1 - F(t^-)\}, \]

whereas if \( t - T > 0 \) then

\[ \mathbb{E} \left( \frac{1}{n} R_n(t) \right) = \frac{1}{\mu} \int_{[t-T,t]} \{1 - F(t^-)\} \, ds = \frac{T}{\mu} \{1 - F(t^-)\} = \frac{T}{\mu} \{1 - F(t^-)\}. \]

Q.E.D.

In the next proof, as well as in the proof of Lemma 3.10, we use integration by parts, namely the fact that if \(-\infty < a < \alpha < \beta < b < \infty \) and \( f : \mathbb{R} \to \mathbb{R} \) and \( g : \mathbb{R} \to \mathbb{R} \) are of bounded variation on \( [a, b] \), then

\[ \int_{[a,\beta]} f(s^+) \, dg(s) + \int_{[\alpha,\beta]} g(s^-) \, df(s) = f(\beta^+)g(\beta^-) - f(\alpha^-)g(\alpha^+); \]

see, e.g., (21.67) in Hewitt and Stromberg (1965).

**Lemma 3.4.**

\[ \mathbb{E} \left( \frac{1}{n} D_n(t) \right) = \frac{1}{\mu} \int_{[0,t]} \hat{s} \, dF(s) \quad \text{when} \quad 0 \leq t < \infty. \]

**Proof.** It is easily seen that \( P[Z \leq T, Y + Z \leq t] = P[Z \leq \hat{t}, Y + Z \leq t] \). Therefore, by (3.1),

\[ \mathbb{E} \left( \frac{1}{n} D_n(t) \right) = P[Z \leq \hat{t}, Y + Z \leq t]. \]

By Proposition 3.1(b), the distribution of \( (Z, Y) \) is the same as that of \( (Y, Z) \); consequently, by Proposition 3.2 and integration by parts,

\[ \mathbb{E} \left( \frac{1}{n} D_n(t) \right) = P[Y \leq \hat{t}, Y + Z \leq t] \]

\[ = \frac{1}{\mu} \int_{[0,\hat{t}]} \{F(t) - F(s)\} \, ds = \frac{1}{\mu} \left( \int_{[0,\hat{t}]} s \, dF(s) + \hat{t}[F(t) - F(\hat{t})] \right). \]
For no so large that $\theta^0$.

\[ \left( \int \left( t \mu^0 \right) \right)_{\text{d}} \geq \left( \int \left( t \mu^0 \right) \right)_{\text{d}} \]

1. **Proof.** Consider a fixed $s$ such that $0 \leq x \leq 0 \leq x$. As in the proof of Lemma 3.6.

\[ 0 = \left( \left( \int t^\mu \right) \right)_{\text{d}} \]

From which it follows that

\[ 0 = t^\mu_{\text{d}} = \left( \frac{\mu}{\text{d}} \right)_{\text{d}} \]

Lemma 3.6. Suppose that $\mu^0$ is such that $1 > (\omega)^\mu_{\text{d}} > 0$ and $0 = (0)^\mu_{\text{d}}$.

The last integrand is dominated by a function whose value at $s$ is $I$.

\[ \left( \int \left( t \mu^0 \right) \right)_{\text{d}} \]

1. **Proof.** Recall that $\theta^0$ and $0 \leq \theta^0$.

\[ \left( \int \left( t \mu^0 \right) \right)_{\text{d}} \]

The last integrand is replaced by $\mu^0_{\text{d}}$. Now, by (5), (7) and Lemma 3.9 and 3.3 and 3.4, and Lemma 3.5.

\[ \left( \int \left( t \mu^0 \right) \right)_{\text{d}} \]

1. **Proof.** Recall that $\theta^0$ and $0 \leq \theta^0$.

\[ \left( \int \left( t \mu^0 \right) \right)_{\text{d}} \]

Thus, if $\mu^0_{\text{d}}$.

\[ \left( \int \left( t \mu^0 \right) \right)_{\text{d}} \]

1. **Proof.** Recall that $\theta^0$ and $0 \leq \theta^0$.
\[ I_{(0,\sigma)}(s) \frac{\rho_n \theta_n}{\bar{F}(s)\{\theta_n + (\delta/\mu)\bar{F}(s)\}} \leq I_{(0,\sigma)}(s) \frac{1}{\bar{F}(s)\{\theta_n + (\delta/\mu)\bar{F}(s)\}} \]

\[ \leq I_{(0,\sigma)}(s) \frac{\mu}{\delta} \frac{1}{\bar{F}(\sigma)}^{-2}. \]  

(ii)

Since

\[ \int_{(0,\sigma)} \frac{1}{\delta} dF(s) \leq \int_{(0,\sigma)} \frac{1}{\delta} dF(s) + \frac{1}{T} \{F(\sigma) - F(T)\}^+ < \infty, \]

\( \bar{F}(\sigma) > 0, \) and \( \int_{[0,\sigma]} x^{-1} dF(x) \) is finite, we see that the last expression in (ii) is \( dF \)-integrable; therefore, by the DCT, the right side of (i) goes to 0 when \( n \to \infty \). Q.E.D.

The proof of the next lemma makes use of the following fact.

**Proposition 3.7.** If \( f : \mathbb{R} \to \mathbb{R} \) and \( g : \mathbb{R} \to \mathbb{R} \) are nondecreasing on a bounded interval \( [a, b] \) and \( \gamma = \inf\{ |f(x) - g(x)| : a \leq x \leq b \} \) is greater than 0, then the variation of \( (f - g)^{-1} \) over \( [a, b] \) is at most \( \gamma^{-2}\{f(b) - f(a) + g(b) - g(a)\} \).

This follows from the observation that if \( a = x_0 < \cdots < x_n = b \) is a partition of \( [a, b] \) then

\[ \sum_{i=1}^{m} \left| (f - g)^{-1}(x_i) - (f - g)^{-1}(x_{i-1}) \right| \leq \left( \inf_{a \leq x \leq b} |f(x) - g(x)| \right)^{-2} \times \left( \sum_{i=1}^{m} \{ f(x_i) - f(x_{i-1}) \} + \sum_{i=1}^{m} \{ g(x_i) - g(x_{i-1}) \} \right) = \gamma^{-2}\{f(b) - f(a) + g(b) - g(a)\}. \]

**Lemma 3.8.** If \( U(s) = \{\xi(1/n)C_n(s)\}^{-1} \) and \( -\infty < a \leq 0 \leq b < \infty, \) the variation of \( U \) over \( [a, b] \) is at most \( \theta_n^{-1}(2\delta/\mu) \).

**Proof.** Since \( \xi C_n \) is constant on \((-\infty, 0]\), it suffices to deal with the variation on \([0, b] \). Put

\[ f(s) = \theta_n + \frac{1}{\mu} \int_{(0,\sigma)} \{1 - F(x)\} dx \quad \text{and} \quad g(s) = \frac{1}{\mu} \int_{(0,\sigma)} \{F(s) - F(x)\} dx, \]

so that

\[ f(s) - g(s) = \theta_n + \frac{1}{\mu} \int_{(0,\sigma)} \{1 - F(s)\} dx \]

\[ = \theta_n + \frac{\delta}{\mu} \bar{F}(s) = \xi \frac{1}{n} C_n(s) \Rightarrow \{U(s)\}^{-1}. \]  

(i)

The functions \( f \) and \( g \) are nondecreasing and, from (i),

\[ \inf_{0 \leq s \leq t} |f(s) - g(s)| \geq \theta_n > 0 \quad \text{when} \quad 0 \leq t < \infty. \]
Therefore, by Proposition 3.7 and (i), the variation of \( U = (f - g)^{-1} \) over \([0, b]\) is at most

\[
\theta_n^{-2} \left( \frac{1}{\mu} \int_{(0, b)} (1 - F(x)) \, dx + \frac{1}{\mu} \int_{(0, b)} (F(b) - F(x)) \, dx \right).
\]

(ii)

Since each of the integrals in (ii) is bounded by \( \hat{b} \), the desired conclusion follows. Q.E.D.

**Remark 3.9.** For \( d \in \mathbb{N}^+ \), let \((X_i^{(1)}, \ldots, X_i^{(d)}): i = 1, 2, \ldots\) be a sequence of independent identically distributed \( \mathbb{R}^d \)-valued random vectors, defined on a probability space \((\Omega', \mathcal{F}', P')\). It is known [see Kiefer (1961)] that if \( 1 < \alpha < 2 \), then there exists a constant \( K \), dependent on \( \alpha \) and the dimension \( d \) but not on the distribution of the random vectors, such that, for any \( n \in \mathbb{N}^+ \) and any \( \delta > 0 \),

\[
P^r \left[ \sup_{a \in \mathbb{R}^d} |Q_n^*(a) - Q(a)| > \delta \right] < Ke^{-an\delta^2},
\]

where, for \( a = (a_1, \ldots, a_d) \),

\[
Q(a) = P'[X_1^{(1)} \leq a_1, \ldots, X_1^{(d)} \leq a_d] \quad \text{and} \quad Q_n^*(a) = \frac{1}{n} \sum_{i=1}^{n} I[X_i^{(1)} \leq a_1, \ldots, X_i^{(d)} \leq a_d].
\]

It follows by the Borel-Cantelli lemma that there exists \( \Omega_0 \in \mathcal{F}' \) with \( P'(\Omega_0) = 1 \) and such that if \( \omega \in \Omega_0 \) then one can find \( n_0(\omega) \) such that \( n \geq n_0(\omega) \) implies

\[
\sup_{a \in \mathbb{R}^d} |Q_n^*(a, \omega) - Q(a)| \leq \sqrt{n^{-1} \log n}.
\]

**Lemma 3.10.** Let \((\rho_n)_{n=1}^\infty\) be a sequence of positive reals. If

\[
0 < \sigma < \infty \quad \text{and} \quad \lim_{n \to \infty} \rho_n \theta_n^{-2} \sqrt{n^{-1} \log n} = 0,
\]

then

\[
\lim_{n \to \infty} \rho_n \sup_{0 \leq t \leq \sigma} |\Lambda_n(t) - \bar{\Lambda}_n(t)| = 0 \quad \text{a.s.}
\]

**Proof.** Observe that

\[
D_n(t) = \sum_{r=1}^{n} I[Z_r \leq T, W_r \leq t]
\]

and

\[
C_n(t) - n\theta_n = R_n(t) = \sum_{r=1}^{n} I[t - T \leq Y_r, t \leq W_r]
\]

\[
= \sum_{r=1}^{n} I[Y_r \leq t, t \leq W_r] - \sum_{r=1}^{n} I[Y_r < t - T, t \leq W_r]
\]

\[
= \left( \sum_{r=1}^{n} I[Y_r \leq t] - \sum_{r=1}^{n} I[Y_r \leq t, W_r < t] \right) - \left( \sum_{r=1}^{n} I[Y_r < t - T] - \sum_{r=1}^{n} I[Y_r < t - T, W_r < t] \right).
\]

In the light of the above, and recalling that \( F_W(0) = 0 \), we conclude from Remark 3.9 that there exists \( \Omega_0 \in \mathcal{F} \) with \( P(\Omega_0) = 1 \) and such that, for any \( \omega \in \Omega_0 \):
(1) \( (1/n)D_n(0, \omega) = 0 \) for \( n \in \mathbb{N}^+ \);
(2) one can find \( n_0(\omega) \) such that \( n \geq n_0(\omega) \) implies

\[
\sup_{t \in \mathbb{R}} \left| \frac{1}{n} D_n(t, \omega) - \frac{1}{n} D_n(t) \right| \leq \sqrt{n^{-1} \log n}
\]

and

\[
\sup_{t \in \mathbb{R}} \left| \frac{1}{n} C_n(t, \omega) - \frac{1}{n} C_n(t) \right| \leq 4\sqrt{n^{-1} \log n}.
\]

Consider an arbitrary but fixed \( t \in [0, \sigma] \) and an arbitrary but fixed \( \omega \in \Omega_0 \), and let \( n_0(\omega) \) be as above. To simplify the notation, we omit \( \omega \) from some expressions which in fact depend on \( \omega \); for instance, an expression such as

\[
\int \cdots d\left( \frac{1}{n} D_n \right)(s)
\]

denotes a Lebesgue-Stieltjes integral in which the integrator is the function \( s \mapsto (1/n)D_n(s, \omega) \).

Now by (3.4) and (3.5), and since \((1/n)C_n = \theta_n > 0\),

\[
|A_n(t, \omega) - \bar{A}_n(t)| = \left| \int_{[0,t]} \frac{1}{(1/n)C_n(s, \omega)} d\left( \frac{1}{n} D_n \right)(s)
\right|
\]

\[
- \int_{[0,t]} \frac{1}{\mathcal{E}(1/n)C_n(s)} d\left( \frac{1}{n} D_n \right)(s)
\]

\[
= \left| \int_{[0,t]} \left\{ \frac{1}{(1/n)C_n(s, \omega)} - \frac{1}{\mathcal{E}(1/n)C_n(s)} \right\} d\left( \frac{1}{n} D_n \right)(s)
\right|
\]

\[
+ \int_{[0,t]} \frac{1}{\mathcal{E}(1/n)C_n(s)} d\left( \frac{1}{n} D_n \right)(s)
\]

\[
- \int_{[0,t]} \frac{1}{\mathcal{E}(1/n)C_n(s)} d\left( \frac{1}{n} D_n \right)(s)
\]

\[
\leq A_n + B_n,
\]

where

\[
A_n = \int_{[0,t]} \left| \frac{\mathcal{E} - \frac{1}{n} C_n(s)}{\frac{1}{n} C_n(s) - \frac{1}{n} C_n(s)} \right| d\left( \frac{1}{n} D_n \right)(s)
\]

and

\[
B_n = \int_{[0,t]} \frac{1}{\mathcal{E}(1/n)C_n(s)} d\left( \frac{1}{n} D_n - \frac{1}{n} D_n \right)(s)
\]

For \( n \geq n_0(\omega) \), the integrand in \( A_n \) is bounded by \( \theta_n^{-2}4\sqrt{n^{-1} \log n} \); since \((1/n)D_n(\cdot, \omega)\) is a probability subdistribution function, it follows that

\[
A_n \leq \theta_n^{-2}4\sqrt{n^{-1} \log n}.
\]

To obtain a bound for \( B_n \), put \( U(s) = \{\mathcal{E}(1/n)C_n(s)\}^{-1} \) and \( V(s) = (1/n)D_n(s, \omega) - \mathcal{E}(1/n)D_n(s) \) and observe that \( V(s^+) = V(s) \) and, as is clear from Lemma 3.3, \( U(s^-) = \).
$U(s)$. By Lemma 3.8, for an arbitrary $a < 0$ and $b > \sigma$, $U$ is of bounded variation over $[a,b]$; and $V$ is obviously of bounded variation over any compact interval. Therefore one can use integration by parts to find that

$$B_n = \left| \int_{[0,1]} U(s^-) \, dV(s) \right|$$

$$= \left| U(t^+)V(t^+) - U(0^-)V(0^-) - \int_{[0,1]} V(s^+) \, dU(s) \right|$$

$$= \left| \frac{1}{\mathcal{E}(1/n)C_n(t^+)} \left( \frac{1}{n} D_n(t, \omega) - \mathcal{E} \frac{1}{n} D_n(t) \right) \right.$$

$$\left. - \frac{1}{\mathcal{E}(1/n)C_n(0)} \left( \frac{1}{n} D_n(0^-, \omega) - \mathcal{E} \frac{1}{n} D_n(0^-) \right) \right.$$ 

$$\left. - \int_{[0,1]} \left( \frac{1}{n} D_n(s, \omega) - \mathcal{E} \frac{1}{n} D_n(s) \right) \frac{d}{\mathcal{E}(1/n)C_n}(s) \right| .$$

The second term in the last expression equals $0/\theta_n = 0$. When $n \geq n_0(\omega)$, the first term in that expression is bounded by $\theta_n^{-1} \sqrt{n^{-1} \log n}$ and the integrand in the third term is bounded by $\sqrt{n^{-1} \log n}$. Thus, in view of Lemma 3.8,

$$B_n \leq \theta_n^{-1} \sqrt{n^{-1} \log n} + \frac{2b}{\mu} \theta_n^{-2} \sqrt{n^{-1} \log n}. \quad (iii)$$

Since $b > \sigma$ is arbitrary, we see from (i)–(iii) that

$$|L_n(t, \omega) - \Lambda_n(t)| \leq \left( \frac{2\sigma}{\mu} + 5 \right) \theta_n^{-2} \sqrt{n^{-1} \log n} \quad (iv)$$

for all $n$ so large that $\theta_n \leq 1$. And, clearly, (iv) implies the statement of the lemma. Q.E.D.

### 3.3. Proofs of Theorems A and B.

**Proof of Theorem A.** The first assertion in Theorem A follows from Lemma 3.5 and Lemma 3.10 with $\rho_n = 1$; the second assertion follows from Lemma 3.6 and Lemma 3.10. Q.E.D.

Theorem A establishes some strong uniform consistency properties of our estimator of the hazard function $\Lambda$. Corresponding properties of our estimator of $F$ are stated in Theorem B. The transition from Theorem A to Theorem B is based on the next two lemmas.

As suggested to us by Richard Gill in relation to a competing-risks problem, one can use Theorem C (stated at the end of Section 2) and integration by parts to obtain the following inequality—which gives some precision to the general notion that if the hazard functions of two life distributions are "close", then so are those life distributions.

**Lemma 3.11.** Let $Q$ and $R$ be life distributions, and let $Q$ and $R$ be their hazard functions. If $t \in [0, \infty)$ is such that $R(t) < 1$, then

$$\sup_{0 \leq s \leq t} |Q(s) - R(s)| \leq 5(1 - R(t))^{-1} \sup_{0 \leq s \leq t} |Q(s) - R(s)|.$$
For details of the argument that Theorem C implies Lemma 3.11, see Winter (1989a).

The next result shows that the mapping which sends a hazard function to the corresponding life distribution is, in a certain sense, continuous.

**Lemma 3.12 (Gill).** Let \( Q \) and \( Q_1, Q_2, \ldots \) be life distributions, let \( \mathcal{L} \) and \( \mathcal{L}_1, \mathcal{L}_2, \ldots \) be their hazard functions, and let \( \sigma \in [0, \infty) \) be such that \( Q(\sigma^-) < 1 \). Now if

\[
\lim_{n \to \infty} \sup_{0 \leq t \leq \sigma} |\mathcal{L}_n(t) - \mathcal{L}(t)| = 0 \quad \text{then} \quad \lim_{n \to \infty} \sup_{0 \leq t \leq \sigma} |Q_n(t) - Q(t)| = 0.
\]

If the condition \( Q(\sigma^-) < 1 \) is replaced by \( Q(\sigma^-) < 1 \), the above result follows from Lemma 3.11. The fact that the result is true under the less restrictive condition \( Q(\sigma^-) < 1 \) is implicit in the proof of Lemma 2 in Gill (1981).

**Proof of Theorem B.** The first assertion in Theorem B follows from the first part of Theorem A "by continuity", namely by Lemma 3.12. The second assertion in Theorem B follows from the second part of Theorem A and Lemma 3.11.

4. DISCUSSION

4.1. Other Estimators.

When observations are generated by the life-testing setup described in Section 1, and there is no censoring (i.e., in the case \( T = \infty \)), one can also use the estimator \( F_n^c \) defined by

\[
F_n^c(t) = \frac{\sum_{r=1}^{n} (1/W_r)I[W_r \leq t]}{\sum_{r=1}^{n} 1/W_r}.
\]

This estimator, introduced by Cox (1969), is a particular case of a more general estimator subsequently introduced by Vardi (1985): the Cox estimator corresponds to the \( s = 1 \), \( w(x) = x \) case of Vardi’s general model. Gill, Vardi, and Wellner (1988) have established a number of asymptotic properties of the Vardi estimator; in the particular case which yields Cox’s estimator, they show, inter alia, that \( \sup_{0 < t < m} |F_n^c(t) - F(t)| \to 0 \) a.s. Horvath (1985) gives a simple proof of the strong uniform consistency of \( F_n^c \) and proves the weak convergence and strong approximation of

\[
(n^{1/2}|F_n^c(t) - F(t)|: 0 \leq t < \infty),
\]

a process also studied by Sen (1984). Unlike our Theorem B, these authors’ results do not give any information regarding the rate at which the sup-norm error converges to 0.

It should be noted that the Cox-Vardi estimator cannot be used when \( T < \infty \) and some of the observations are censored: in that case, if one only knows (for \( r = 1, 2, \ldots, n \)) the initial age \( y_r \) and the time \( z_r = \min\{T, z_r\} \) that the \( r \)-th object is under observation, the Cox-Vardi estimator cannot even be computed.

Both the Cox-Vardi estimator and the product-limit estimator can be used when \( T = \infty \), and, as will be noted below, they appear to perform similarly. That may appear surprising if one believes that the product-limit estimator uses more information—in that it explicitly uses information about backward and forward recurrence times, instead of relying on this information as it is amalgamated in the total life \( W_r \). However, the pair \((Y_r, Z_r)\) does not carry more information than the random variable \( W_r \): the joint distribution of \( Y_r \) and \( Z_r \) is the same as the joint distribution of \( UW_r \) and \( (1 - U)W_r \), where \( U \) is a uniform \((0,1)\) random variable, independent of \( W_r \); see Proposition 3.1(d) and (c). The product-limit
estimator does not gain on the Cox-Vardi estimator by using more information; rather, its advantage lies in that it can be adapted to the case with censoring.

It is interesting to note that, when the ages of the objects in use at the start of observation are known, one can estimate $F$ without further observation. As is clear from Proposition 3.1(a), the random variables $Y_1, Y_2, \ldots$ have $f_Y(s) = \mu^{-1}[1 - F(s)] f_{(0, \infty)}(s)$ as their common density. As $1 - F(s) = \mu f_Y(s) = f_Y(s)/f_Y(0^+)$ when $s > 0$, it is clear that any nonparametric estimator of the monotone density $f_Y$ yields a nonparametric estimator of $F$. The classical monotone-density estimator introduced by Grenander (1956) [see also Prakasa Rao (1969)] is not suitable for such an application, because it severely overestimates $f_Y(0^+)$; see Winter (1987) for a discussion. But a “monotonized histogram” estimator of $f_Y$ introduced by Barlow and van Zwet (1970) is suitable: Watelet (1984) [see also Winter and Watelet (1989)] showed that if $F$ is continuous, then under reasonable conditions regarding the manner in which the partitions used for computing the histogram “shrink” with increasing sample size, this method yields a strongly uniformly consistent estimator of $F$.

Since $F_Y = F_Z$, it is clear that the histogram-based approach described above is also applicable if ages are not known and the data consist of observed values of $Z_1, Z_2, \ldots$, with or without censoring. If both backward and forward recurrence times are known, one can use them separately to estimate $F$ (via monotonized histograms, as above), and it is then reasonable to take the average of those two estimates as an estimate of $F$. In the discussion below, we use $F_n^w$ to designate such an estimator.

4.2. Simulation.

Simulation studies were performed by Luc Watelet (Department of Biostatistics, University of Washington) to compare $F_n$, $F_n^w$, and $F_n^w$. Only the case without censoring ($T = \infty$) was considered, and simulations were performed with $F$ a uniform distribution, with $F$ an exponential distribution, and in several cases where $F$ was lognormal. It was found that, quite generally, the Cox-Vardi estimator $F_n^v$ and the product-limit estimator $F_n$ produced very similar results and were of comparable quality. With one exception—discussed below—those two estimators performed much better than the histogram-based estimator $F_n^w$.

In the cases examined—apart from the exception discussed below—the Cox-Vardi and product-limit estimators appeared to be essentially unbiased estimators of $F(t)$ for $t \geq \gamma$ the 10th percentile of $F$, when based on samples of size $n = 100$.

For the empirical distribution function (e.d.f.), the distribution of the Kolmogorov distance, namely the maximum absolute difference between $F$ and the estimator, is extensively tabulated; see Birnbaum (1952). For the three estimators which were examined, that distribution was determined empirically, in the cases considered, from 1000 replications of samples of size 100. In terms of this error distribution, both the Cox-Vardi and the product-limit estimators were, for three choices of $F$, comparable to the e.d.f. based on a sample of size 80, 40, and 25, respectively; but in the exceptional case discussed below, this “equivalent sample size” appeared to be as low as 5. Although the cases examined by Watelet were not sufficient to establish a clear pattern, the available cases suggest that the Cox-Vardi and product-limit estimators perform nearly as well as the e.d.f. when $F_w$ is not very different from $F$, but they perform much less well than the e.d.f. when $F_w$ and $F$ differ appreciably; i.e., even though those estimators can be used with length-biased data, they are not very efficient when the length bias is appreciable.

The exceptional case mentioned above comes about when $F$ is a lognormal distribution,
i.e. the distribution function of $e^{\tau N}$, where $N$ is a normal $(0, 1)$ random variable and $\tau$ is a positive real constant. Cox (1969) examined the behaviour of $F_n^c$ for lognormal $F$, for various values of $\tau$. Having established that the estimator is asymptotically unbiased and normal, Cox examined the variance and determined that the efficiency of his estimator decreased sharply with increasing $\tau$. The simulations performed by Watelet showed that the Cox-Vardi and product-limit estimators behave very well for $\tau = 0.5$ and moderately well for $\tau = 1$, but become very inadequate when $\tau = 2$. Cox had anticipated poor performance for such $\tau$, but expected the problem to be one of large variance, hence limited efficiency. Simulation showed that, in fact, the problem is one of bias: when $F$ is the lognormal distribution function with $\tau = 2$, and $F(t)$ is being estimated with $t = \text{median}$, $F(t) = 0.5$, the simulation averages (over 1000 replications) of the Cox-Vardi and product-limit estimates were 0.4 when the sample size was $n = 100$, and it took sample sizes of $n = 500$ to bring those averages to 0.48. On the other hand, $F_n^w$ appeared to be nearly unbiased in this case, even with $n$ as small as 25.

4.3. Other Related Work.

Gill (1981) and Winter and Foldes (1984) have obtained strong uniform consistency results for a life-testing setup related to the $T < \infty$ case of the setup studied herein. The principal difference is that in the previously examined setup, observation starts when the initial batch of $n$ objects are put in use; a further difference is that in that setup each of the $n$ "positions" is observed for a fixed length of time $T$, so that the replacement objects—which are used to replace any objects that fail prior to time $T$—are also observed. Winter and Foldes were concerned with estimation of the failure-rate function when $F$ is absolutely continuous. Gill showed that $\sup_{0 \leq t \leq T} |F_n(t) - F(t)| \to 0$ a.s., where $F_n$ is the product-limit estimator.

Since a life-length greater than $T$ cannot be observed in the setup studied by Gill, the supremum in his consistency result cannot be over an interval extending beyond $T$. In the setup studied herein, arbitrarily long lives can be observed: although it is impossible in our setup (in the case $T < \infty$) to observe a forward recurrence time longer than $T$, arbitrarily long backward recurrence times are possible; hence our results involve a supremum over $[0, \sigma]$, with $\sigma > T$ permitted.

Our results depend on the assumption that $F(0) = 0$, whereas Gill’s result is valid for $F(0^-) = 0$. That difference is due to the fact that the distribution function of the random variable $W$ is continuous at 0, which corresponds to the fact that, in the setup studied herein, it is essentially impossible to observe a life length which equals 0: if there is an instantaneous failure at the time when observation starts, then "the object in use at the start of observation" is the object (with nonzero life length) which replaces the one that suffered an instantaneous failure.

Keiding and Gill (1987) obtain a weak convergence result for our estimator of $F$ in the case where $F$ is absolutely continuous.

REFERENCES


