Contributions to the Theory of Product-Limit Estimators

Michael L. Moher

Ottawa-Carleton Institute
for
Graduate Studies and Research
in
Mathematics and Statistics

Thesis submitted to
the School of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Mathematics

Université d'Ottawa / University of Ottawa
1989

ISBN 0-315-68041-5

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

The product-limit estimator is shown to be a strongly uniformly consistent estimator of the distribution function of a renewal process which started long before the commencement of observation. This product-limit estimator is based on the censored data obtained from independent realizations of such a process in one of two scenarios: one observation per renewal process, and multiple observations per renewal process. In the former scenario a lower bound on the rate of convergence is obtained.

ACKNOWLEDGEMENTS

I especially thank Dianne, Elizabeth, and Kathleen for their understanding and patience, shown while I was working on this thesis. I also want to thank Ben Winter for the many hours of effort he spent reviewing this material.

Table of Contents

Table of Notation

		Page No.
$\mathfrak{B}$	the Borel σ-algebra on the real line	7
d.f.	distribution function	1
$D(t)$	number of observed renewals	11
$\Delta D(t)$	$D(t^+)-D(t^-)$	6
$\mathbb{E}$	the expectation operator	7
F	the common distribution function	1
$\bar{F}(t)$	$1-F(t^-)$	6
F_n	the product-limit estimate of F	15
$\mathfrak{F}$	σ-algebra on Ω	7
γ	a positive real number such that $F(\gamma^-)>0$	13
I_A	indicator of set A	1
$I[...]$	indicator of event [...]	6
Λ	the hazard function of F	8
Λ^c	the continuous part of Λ	9
Λ_n	the Nelson hazard estimator	14
N	the renewal function	29
P	probability measure	7
$\rho_\sigma(\cdot,\cdot)$	sup-norm metric on $[0,\sigma]$	6
r.v.	random variable	1
$R(t)$	the number at risk	11
$\mathbb{R}$	the real line	6
σ	an element of $\mathbb{R}$ such that $\bar{F}(\sigma)>0$	19
$\tilde{t}$	min $\{t,T\}$	7
T	a deterministic censoring time	7
τ	sup $\{s\in\mathbb{R}: \Lambda(s)<\infty\}$	9
ω	element of the probability space	7
Ω	probability space	7
$V_\gamma^\sigma(f)$	total variation of function f on $[\gamma,\sigma]$	19
X_i	inter-renewal times	13
Y	consumed life	3
Z	remaining life	3
$\#\{...\}$	cardinality of the set $\{...\}$	11
$\wedge$	minimum	2

CHAPTER 1: Introduction

This thesis considers one form of the problem of estimating the underlying distribution function of a renewal process from censored observations. In particular, we investigate the properties of a product-limit estimator for the distribution function when the observations begin between renewals and are deterministically censored on the right.

Nonparametric estimation of distribution functions

In the classical approach to estimating distribution functions one makes n independent observations of a random variable (r.v.) which has the distribution function (d.f.) F. If these observations are denoted by the random variables $X_1, \ldots, X_n$, then a natural nonparametric estimator for the distribution function is the step function

$$F_n(x,\omega) = \frac{1}{n} \sum_{r=1}^{n} I_{(-\infty,x]}(X_r(\omega)) \tag{1}$$

where I_A is the indicator function of the set A. This is the *empirical distribution function* for the independent and identically distributed random variables $X_1, \ldots, X_n$. By the Glivenko-Cantelli theorem [B1] the error in this empirical distribution function has the property that $\sup_x |F_n(x) - F(x)| \to 0$ a.s. That is, there is uniform convergence of the empirical distribution to the distribution function F a.s. This is referred to as *strong uniform consistency* of the estimator. (Strong consistency refers to the a.s. convergence; for convergence in probability the qualifier strong is omitted.) Some authors have attached such importance to this result that they have referred to it as the fundamental theorem of statistics [L1]. There has been considerable research effort spent searching for estimators which have similar convergence properties under less restrictive conditions.

In the area of survival analysis and life testing, where observations may be censored, a frequently useful estimator of the distribution function is the *product-limit* or *Kaplan-Meier estimator* [K1]. For instance, to represent the independent random censoring situation, we let $(X_n: n \geq 1)$ be an independent sequence of positive random variables with common distribution function F, and let $(U_n: n \geq 1)$ be an independent sequence of positive random variables, independent of $(X_n: n \geq 1)$, with a common distribution function G. The actual observations (and their types: censored or not) are represented by

$$\tilde{X}_r = X_r \wedge U_r , \qquad \Gamma_r = I[X_r \leq U_r] \tag{2}$$

where $\wedge$ denotes minimum. The two processes D_n and R_n defined by

$$D_n(t,\omega) = \sum_{r=1}^{n} I[\, \tilde{X}_r(\omega) \leq t,\ \Gamma_r(\omega) = 1] \tag{3}$$

$$R_n(t,\omega) = \sum_{r=1}^{n} I[\, \tilde{X}_r(\omega) \geq t\} \tag{4}$$

are combined to form the product-limit estimate,

$$1 - F_n(t,\omega) = \prod_{s \leq t} \left(1 - \frac{\Delta D_n(s,\omega)}{R_n(s,\omega)}\right), \tag{5}$$

where $\Delta D_n(s,\omega) = D(s^+,\omega) - D(s^-,\omega)$, and the product is over that finite number of s where $\Delta D_n(s,\omega) \neq 0$.

The product-limit estimator is used in lieu of the empirical distribution function in many situations which involve censored data. In a number of cases it has been shown to have strong uniform consistency and other properties analogous to those of the empirical distribution function.

The application — stabilized renewal processes

The prototype physical model of a renewal process involves the successive replacement of light bulbs. A bulb is installed for service at time 0, fails at time X_1, and is then renewed with a new bulb. The second bulb fails at time X_1+X_2 and is renewed with a third bulb, and so on. The lifetimes of the bulbs (inter-renewal times) are assumed to be independently and identically distributed according to a distribution function F. If one observes a single renewal process until the n*th* renewal, or equivalently n independent renewal processes until the first renewal, then the observations which are made on the process are the random variables X_1, X_2, . . . , X_n which represent the inter-renewal times. For this situation the empirical distribution function (1) is a strongly uniformly consistent estimator of the distribution function F.

However, the process we will consider corresponds to the physical situation of commencing observations at some time after the renewal process has started; we will refer to this as a stabilized renewal process.

Definition A *stabilized renewal process* with underlying distribution F is an independent sequence of random variables ((Y,Z), X_1, X_2, ...) defined on a common probability space $(\Omega,\mathcal{F},P)$. The inter-renewal times X_1, X_2, ... are distributed according to F where $F(0^-)=0$ and $\mu=\int s\,dF(s)<\infty$. The pair of r.v.s (Y, Z) are positive and independent of the X_i but not necesssarily of each other, and are distributed according to

$$F_{Y,Z}(y,z) = \frac{1}{\mu}\int_{[0,y]} (F(s+z) - F(s))\,ds \tag{6}$$

for non-negative y and z, with $F_{Y,Z}(y,z) = 0$ elsewhere.

The r.v. Y represents the age of the object in use at the time the observations commenced; Z represents the remaining life of that object; and X_1, X_2, ... represent the times between all renewals after the first one observed. The d.f. (6) is the limiting joint distribution of the backward and forward recurrence times of a renewal process if F is non-arithmetic. A stabilized renewal process intuitively models the situation where observations of a renewal process commence a relatively long time after the process started.

We shall assume that a sequence of independent and identically distributed stabilized renewal processes can be observed. Consequently, even with a finite censoring time, one can make statements about the asymptotic properties of various functions as the number of renewal processes under observation tends to infinity.

We will consider the problem of estimating the d.f. F of a stabilized renewal process for two different cases. Both cases are subject to deterministic censoring at a time T after the observations commence. In the first case the only observations are the pair of random variables (Y,Z) or their censored version. In the second case, one continues observing each stabilized renewal process until censoring occurs.

Previous results

The product-limit estimator was proposed by Kaplan and Meier in 1958 [K1], and since that time it has been the subject of much study regarding its application to the statistics of censored data. Recent developments in this area have seen the application of the theory of counting processes, martingales and stochastic integration to the analysis of the product-limit estimator [A1], [G1], [S1], and [F2]. The application of product-limit estimators to renewal processes is of interest because it is one instance where standard techniques are not readily applicable [G1]. Previous work in the area of applying product-limit estimators to renewal processes includes [G1], [G2], and [W1].

In [G2] Gill shows the strong uniform consistency of the product-limit estimator in the case of an ordinary renewal process with deterministic censoring. By an ordinary renewal process we mean the observations commence at the same time as the process. Gill goes on to show that, in the case of integer-valued observations the error at the integers converges to a normal distribution; and in the case of continuous F, the error in the product-limit estimator converges weakly to a continuous zero-mean Gaussian process.

In [W1] Winter and Földes consider the first form of the problem that we will study. They assume that the pair (Y,Z), or a censored version thereof, is the only observation made of each of n stabilized renewal processes. Under these conditions Winter and Földes show the strong uniform consistency of a modified form of the product-limit estimator as the number of renewal processes approaches infinity. They also obtain lower bounds on the rate of convergence of the estimate.

Outline

In Chapter 2 we summarize the assumptions, notation and conventions used throughout this thesis, and describe the relationship between product-limit estimators and hazard functions.

In Chapter 3 we further investigate the case of a stabilized renewal process studied by Winter and Földes [W1]. We refer to this as a *stabilized renewal process with a single observation.* By modifying and extending the approach used in [W1] we are able to obtain strong uniform consistency results over the whole real line. We also obtain a lower bound on the rate of convergence.

In Chapter 4 we look at *stabilized renewal processes with multiple observations*, that is, observations are not stopped after the first renewal but continue until the censoring time T. We are able to show that a product-limit type estimator also provides a strongly uniformly consistent estimate of the underlying distribution function over the whole real line in this case.

Considerable effort was spent trying to obtain weak convergence results for the estimators of Chapters 3 and 4, with very limited success. The results which were obtained are not included because of their incomplete nature.

Results which were straightforward to prove but could not be found in the literature have been included in the appendix, for the sake of completeness. Also proofs which are only minor variations on those which can be found in the literature have been relegated to the appendix. References to the appendix use the form, for example, (A1).

CHAPTER 2: Stochastic processes, hazard functions, and related miscellanea

Before we delve further into the theory of product-limit estimators for distribution functions, the requisite number of definitions, notations and conventions are in order.

Notation and conventions

All real-valued or extended real-valued functions or processes, $Q:\mathbb{R}\rightarrow\mathbb{R}$ or $Q:\mathbb{R}\rightarrow\bar{\mathbb{R}}$, shall be assumed to be 0 on $(-\infty,0)$. A function shall be said to be *increasing* if it is monotone nondecreasing; *decreasing* functions are defined similarly. If Q has left-hand limits everywhere we define

$$\bar{Q}(t) = 1 - Q(t^-)\,. \tag{1}$$

If Q has both left- and right-hand limits at t we define $\Delta Q(t)=Q(t^+)-Q(t^-)$, provided the right-hand side is defined. For Q, $Q':\mathbb{R}\rightarrow\mathbb{R}$ and $0\leq t<\infty$ the sup-norm metric will be defined by

$$\rho_t(Q,Q') = \sup_{0\leq s\leq t} \left|Q(s)-Q'(s)\right|\,. \tag{2}$$

We use the notation I_A and I[...] to denote the indicator of the set A and the event [...], respectively, and

$$\bar{t} = \min\{t,T\} \tag{3}$$

where T is a fixed positive real (a deterministic censoring time). We follow the convention that $\frac{1}{0}\cdot 0 = \infty\cdot 0 = 0$ and that $\frac{0}{0} = 0$.

Stochastic processes

Let $(\Omega,\mathcal{F},P)$ be the probability space underlying the discussions which follow. Let $\mathcal{B}$ be the Borel σ-algebra on $\mathbb{R}$. A real-valued ***stochastic process*** on $(\Omega,\mathcal{F},P)$ is a map $X:\mathbb{R}\times\Omega\to\mathbb{R}$ such that the maps $\omega\mapsto X(t,\omega)$ are measurable from $(\Omega,\mathcal{F})$ to $(\mathbb{R},\mathcal{B})$ for all $t\in\mathbb{R}$; i.e., $X=(X_t: t\in\mathbb{R})$ is a family of random variables indexed by $\mathbb{R}$. The dependence on ω will frequently be suppressed when working with stochastic processes. When we say a process is left-continuous, right-continuous, decreasing, increasing, has limits from the left or right, or is of bounded variation, then we mean that the sample paths $X(\cdot,\omega)$ have these properties a.s. Let $\mathbb{E}$ be the expectation operator.

Sample-path integrals

The integrals considered here will be of the form $\int_{[a,b]}X\,dY$ where X and Y are stochastic processes that are of bounded variation on a closed interval containing [a,b] in its interior. We refer to such integrals as ***Lebesgue-Stieltjes integrals*** by analogy with the terminology used when the integrator is increasing (see (A1)). Under these conditions, the integral

$$\Big(\int_{[a,b]} X\,dY\Big)(\omega) = \int_{[a,b]} X(s,\omega)\,Y(ds,\omega) \tag{4}$$

refers to the Lebesgue-Stieltjes integral along the sample paths of the two processes X and Y. Under conditions which will generally hold in this thesis, $\int_{[a,b]}X\,dY$ is a random variable, see (A2).

Hazard functions

Although the focus of this thesis is the product-limit estimator, we will deal mainly with the associated hazard function because it provides both a motivation for the use of the product-limit estimator and a convenient framework for analyzing the product-limit estimator. A *life distribution* is a distribution function F with $F(0^-)=0$ and $F(\infty)\leq 1$. The *cumulative hazard function* Λ, associated with a life distribution F, is defined by the Lebesgue-Stieltjes integral

$$\Lambda(t) = \int_{[0,t]} \frac{1}{1-F(s^-)}\, dF(s) \,. \tag{5}$$

In the following, we will refer to the cumulative hazard function as simply the hazard function, dropping the modifier *cumulative* as is usually done with its counterpart, the cumulative distribution function.

Example 2.1 Let F be the d.f. corresponding to a uniform probability density on [0,1], that is,

$$F(t) = \begin{cases} 0 & \text{if } t<0 \\ t & \text{if } 0\leq t\leq 1 \\ 1 & \text{if } t>1 \end{cases} .$$

Then the corresponding hazard function is

$$\Lambda(t) = \begin{cases} 0 & \text{if } t<0 \\ -\log(1-t) & \text{if } 0\leq t<1 \\ \infty & \text{if } t\geq 1 \end{cases} .$$

Consequently we see that Λ can be infinite for finite t.

Example 2.2 Let F be the following d.f. which includes a jump of e^{-1} at 1:

$$F(t) = \begin{cases} 0 & \text{if } t<0 \\ 1-e^{-t} & \text{if } 0\leq t<1 \\ 1 & \text{if } t\geq 1 \end{cases} .$$

Then the corresponding hazard function is

$$\Lambda(t) = \begin{cases} 0 & \text{if } t<0 \\ t & \text{if } 0 \le t < 1 \\ 2 & \text{if } t \ge 1. \end{cases}$$

Consequently we see that Λ can have finite jumps and need not be unbounded.

Let F be a life distribution, with hazard function Λ, then define

$$\tau = \sup\{t \in \mathbb{R} \colon \Lambda(t) < \infty\} \tag{6}$$

and, for $t<\tau$, let

$$\Lambda^c(t) = \Lambda(t) - \sum_{s \le t} \Delta\Lambda(s) \tag{7}$$

be the continuous part of the hazard function where the sum is over the countable number of s where $\Delta\Lambda(s)>0$. Define the transformation Φ of the hazard function of a life distribution by

$$\Phi(\Lambda(t)) = \begin{cases} 1 - \exp\{-\Lambda^c(t)\} \cdot \prod_{s \le t} (1 - \Delta\Lambda(s)) & \text{if } -\infty < t < \tau \\ 1 & \text{if } \tau < \infty \text{ and } t \ge \tau \end{cases} \tag{8}$$

where $\prod_{s \le t} \phi(s) = \exp\left\{\sum_{s \le t} \log \phi(s)\right\}$, with $\log 0 = -\infty$ and $e^{-\infty} = 0$. Then Φ is the inverse of the transformation (5) (see [G2], [S1], or [W2]); i.e.,

$$F(t) = \Phi(\Lambda(t)), \qquad \text{for } -\infty < t < \infty. \tag{9}$$

The following examples illustrate the correspondence between hazard functions and distribution functions.

Example 2.3 For a partial proof of (9), consider a d.f. F which is continuously differentiable on $[0,\tau)$. For such F, note that

$$\Lambda(t) = \int_{[0,t]} \frac{dF(s)}{1-F(s^-)} = \int_{[0,t]} \frac{F'(s)}{1-F(s)}\, ds$$

is continuous on $[0,\tau)$ by the fundamental theorem of calculus, and hence

$$\Phi(\Lambda(t)) = 1 - \exp\{-\Lambda(t)\}\,.$$

Let $G(s) = -\log(1-F(s))$; then $G'(s)=F'(s)/(1-F(s))$; consequently $\Lambda(t)=G(t)$ and

$$F(t) = 1 - \exp\{\log(1-F(t)\} = 1- \exp\{-\Lambda(t)\} = \Phi(\Lambda(t)).$$

Example 2.4 If the distribution function F is a step function with jumps at $0 \le t_1 < t_2 < ... < t_k$ and $X \sim F$, then Λ is a step function with jumps

$$\Delta\Lambda(t) = \frac{P[X=t_i]}{P[X\ge t_i]} \qquad \text{if } t=t_i \text{ and } 0 \text{ otherwise.}$$

Furthermore, with $t_j = \max\{t_i : t_i \le t\}$

$$\prod_{s\le t}(1-\Delta\Lambda(s)) = \prod_{t_i\le t} \frac{P[X>t_i]}{P[X\ge t_i]} = \frac{P[X>t_1]}{P[X\ge t_1]} \cdot \frac{P[X>t_2]}{P[X\ge t_2]} \cdots \frac{P[X>t_j]}{P[X\ge t_j]}$$

$$= \frac{P[X\ge t_2]}{P[X\ge t_1]} \cdot \frac{P[X\ge t_3]}{P[X\ge t_2]} \cdots \frac{P[X>t]}{P[X\ge t_j]} = P[X>t]$$

$$= 1 - F(t)\,.$$

This verifies (9) when the distribution function is a step function.

Example 2.5 If the hazard function Λ is a step function with jumps at $0 \le t_1 < t_2 < ... < t_k$ then the inversion formula reduces to

$$F(t) = 1 - \prod_{t_i \le t} (1 - \Delta\Lambda(t_i)). \tag{10}$$

The observant reader will note the similarity between (10) of Example 2.5 and the product-limit estimator of (1-5). The advantage of the hazard function approach is that it suggests some natural estimators for distribution functions. In particular, (5) implies that

$$\Delta\Lambda(s) = \frac{\Delta F(s)}{1-F(s^-)}$$

and this suggests that $\Delta\Lambda(s)$ might be estimated by

$$\frac{\#\{\text{inter-renewal times of exactly s}\} \;/\; \#\{\text{objects observed}\}}{\#\{\text{objects observed at age s}\} \;/\; \#\{\text{objects observed}\}}$$

where $\#\{...\}$ is the cardinality of $\{...\}$. This empirical estimate of $\Delta\Lambda(s)$ is the ratio of the number of observed inter-renewal times of exactly s to the number of objects observed at age s. The latter is often referred to as the *number at risk* at time s. This approach provides a discrete estimate (step function approximation) of the hazard function using the observations at hand.

Example 2.6 If the distribution function F is a step function with jumps at $0 \leq t_1 < t_2 < ... < t_k$ and $X \sim F$, then $\Lambda(t)$ can also be written as

$$\Lambda(t) = \sum_{t_i \leq t} \frac{P[X=t_i]}{P[X \geq t_i]} = \sum_{t_i \leq t} P[X=t_i \,|\, X \geq t_i] \,.$$

This form suggests an estimator of the form

$$\hat{\Lambda}(t) = \sum_{s \leq t} \frac{\Delta D(s)}{R(s)} \tag{11}$$

where $\Delta D(s)$ is the number of inter-renewal times of exactly s, and $R(s)$ is the number of objects observed at age s. An estimator of this form was first proposed by Nelson [N1], [N2] and is often referred to as the Nelson hazard estimator.

Example 2.7 If we apply the inverse transformation (8) to the hazard estimator of Example 2.6, we obtain

$$\begin{aligned}\hat{F}(t) &= 1 - \prod_{t_i \le t} (1 - \Delta\Lambda(t_i)) \\ &= 1 - \prod_{t_i \le t} (1 - \frac{\Delta D(t_i)}{R(t_i)}) \qquad (12)\end{aligned}$$

as an estimator of the distribution function, a type of product-limit estimator.

As the last two examples indicate, hazard estimators and product-limit estimators are two sides of the same coin. They are related by the transformation

$$\hat{F} = \Phi(\hat{\Lambda}) \,. \qquad (13)$$

The hazard function is inherently easier to analyze because of the linear way in which it treats the observations, as opposed to the nonlinear form of the product-limit estimator. This advantage will become evident in the following chapters and, as an aside, we note that the relationship between hazard functions and product-limit estimators is one instance of a larger class of corresponding sum- and product-integrals [G3].

CHAPTER 3: Product-limit estimator – single observation of a stabilized renewal process

In this chapter we study a form of the problem of estimating the distribution function associated with a stabilized renewal process which was considered by Winter and Földes [W1]. We consider a sequence of such processes, independent of each other, each with underlying distribution F; we suppose that each process is observed only until the first renewal or the censoring time T, whichever occurs first. For the r*th* process $(Y^{(r)}, Z^{(r)}, X_1^{(r)}, X_2^{(r)}, \ldots)$, $Y^{(r)}$ represents the age of the r*th* among the n objects under observation at the time the observations commenced, and $Z^{(r)}$ represents the remaining life of that object. The random variable $\tilde{Z}^{(r)} = \min\{Z^{(r)}, T\}$ represents the length of time, possibly censored, that the object was under observation. Define $\Gamma^{(r)} = I[Z^{(r)} \leq T]$ to indicate whether an observation was censored or not. Then the ***single observation*** made on the r*th* process is represented by the random triple $(Y^{(r)}, \tilde{Z}^{(r)}, \Gamma^{(r)})$. This information is used to estimate the underlying distribution, F, using a product-limit estimator, and to study the asymptotic behaviour of the estimate as $n \to \infty$.

Assumptions

We assume there exists a $\gamma > 0$ such that $F(\gamma^-) = 0$. We also assume that $T > \gamma$ because otherwise the situation becomes trivial. The constant γ is akin to a warranty period over which the object is guaranteed not to fail. This constant γ is needed to show uniform convergence over an interval starting at zero. As the lifetime of an object decreases, the probability of observing it also decreases in the scenario described above. In the limiting case it is impossible to observe an object with a lifetime of zero. In practice the constant γ could be as small as one wishes, and if it is much smaller than the time scale of the process, it will for practical purposes be equivalent to zero.

Estimation equations

The time axis for each stabilized renewal process is defined such that $t=0$ corresponds to the beginning of the life of the first object observed. The number of observed inter-renewal times of t or less is represented by

$$D_n(t) = \sum_{r=1}^{n} I[Y^{(r)}+\tilde{Z}^{(r)} \leq t,\ \Gamma^{(r)}=1]\ . \tag{1}$$

This process is clearly right-continuous and increasing. The number of objects under observation (number at risk) at age t is represented by the r.v.

$$R_n(t) = \sum_{r=1}^{n} I[Y^{(r)} < t \leq Y^{(r)}+\tilde{Z}^{(r)}]\ . \tag{2}$$

(We define $R_n(t)$ with $Y^{(r)}<t$ rather than, as in [W1], $Y^{(r)}\leq t$. This is of no practical consequence since $F_Y(t)=\int_{[0,t]}\{1-F(s)\}ds$ [W1], and thus $P[Y=t]=0$. However, this modification is useful because it makes $R_n(\cdot)$ left-continuous.)

To estimate Λ, we use a Nelson hazard function estimator, analogous to (2-11):

$$\Lambda_n(t) = \sum_{s\leq t} \frac{\Delta D_n(s)}{R_n(s)}\ . \tag{3}$$

It is useful to note that Λ_n can also be represented by the stochastic integral

$$\Lambda_n(t) = \int_{[0,t]} \frac{1}{R_n(s)}\, dD_n(s)\ . \tag{4}$$

If we note that (4) can also be written as

$$\Lambda_n(t) = \int_{[0,t]} \frac{I[R_n(s)>0]}{R_n(s)} \, dD_n(s)$$

then it can be shown that the integrand in this expression is of bounded variation and consequently, by (A2), $\Lambda_n(t)$ is a random variable.

To obtain an estimator of F, we apply (2-13) to the hazard function estimator in (4), and that yields the product-limit estimator

$$F_n(t) = 1 - \prod_{s \leq t} \left(1 - \frac{\Delta D_n(s)}{R_n(s)} \right) . \tag{5}$$

Example 3.1 Suppose the following observations were made on four stabilized renewal processes: $Y^{(1)}=1$, $\tilde{Z}^{(1)}=4$, $Y^{(2)}=2$, $\tilde{Z}^{(2)}=1$, $Y^{(3)}=4$, $\tilde{Z}^{(3)}=4$, $Y^{(4)}=6$, $\tilde{Z}^{(4)}=1$, and only the third observation was censored. Figures 3.1 and 3.2 show different representations of these observations and the functions D_4, R_4, Λ_4, and F_4, for this outcome. Note that $F_4(\infty)<1$ in this example.

Alternative representation of the hazard function

Several results which are useful in showing the convergence of the hazard estimator (4) to the true hazard function can be extracted directly from [W1]. The first is obtained from Lemma 3.8 of [W1] with $\theta_n=0$; when adjusted for the redefined R_n it states that the expected number of objects under observation at age t satisfies

$$E\tfrac{1}{n}R_n(t) = \tfrac{\tilde{t}}{\mu}\, \bar{F}(t) \qquad \text{for } 0 \leq t < \infty . \tag{6}$$

Since $\bar{F}(t)$ is a decreasing function, (6) implies

$$E\tfrac{1}{n}R_n(t) \geq \frac{\gamma \bar{F}(\sigma)}{\mu} \qquad \text{if } \gamma \leq t \leq \sigma . \tag{7}$$

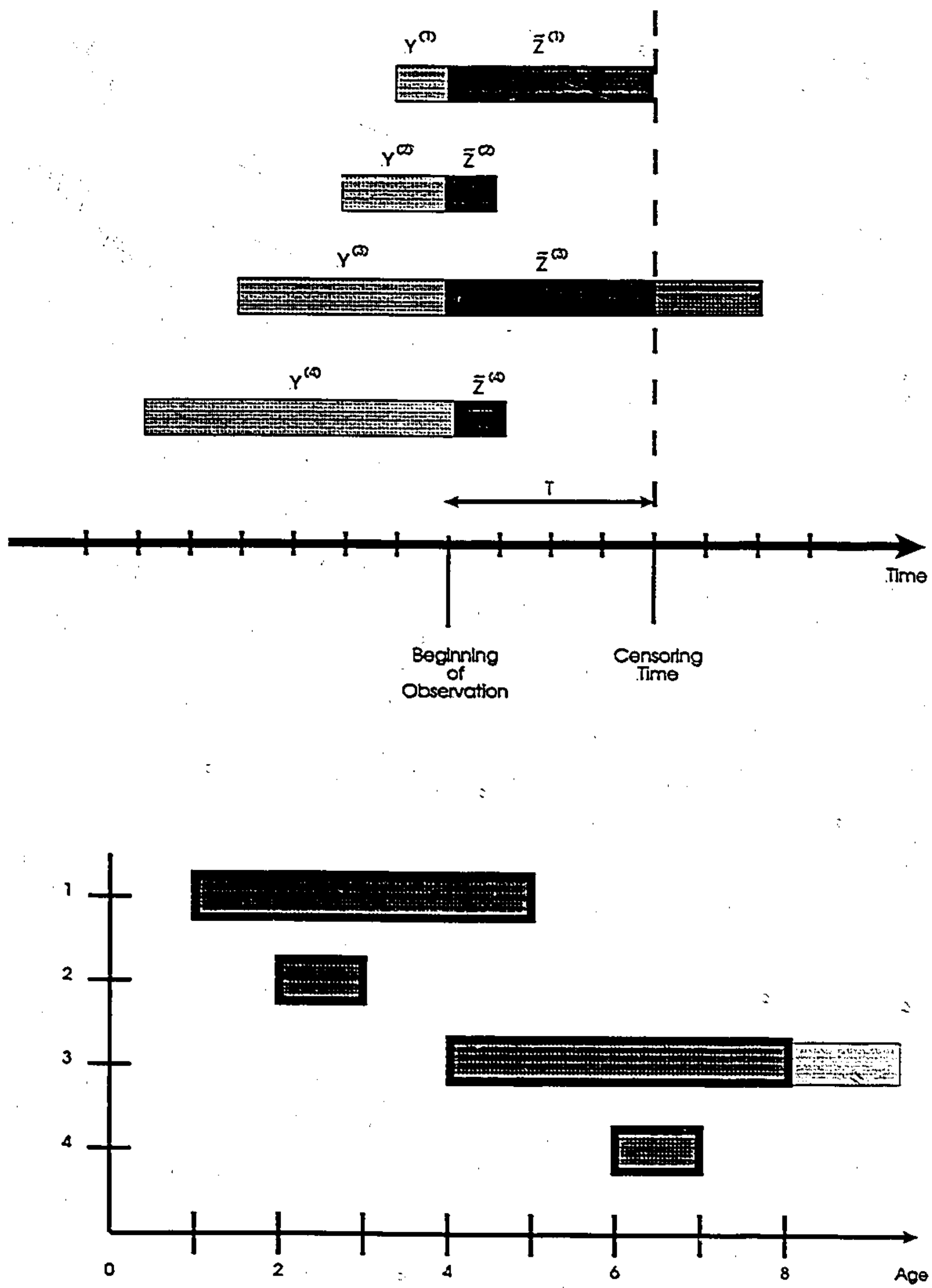

Figure 3.1 Representation of observations of object lifetimes for Example 3.1.

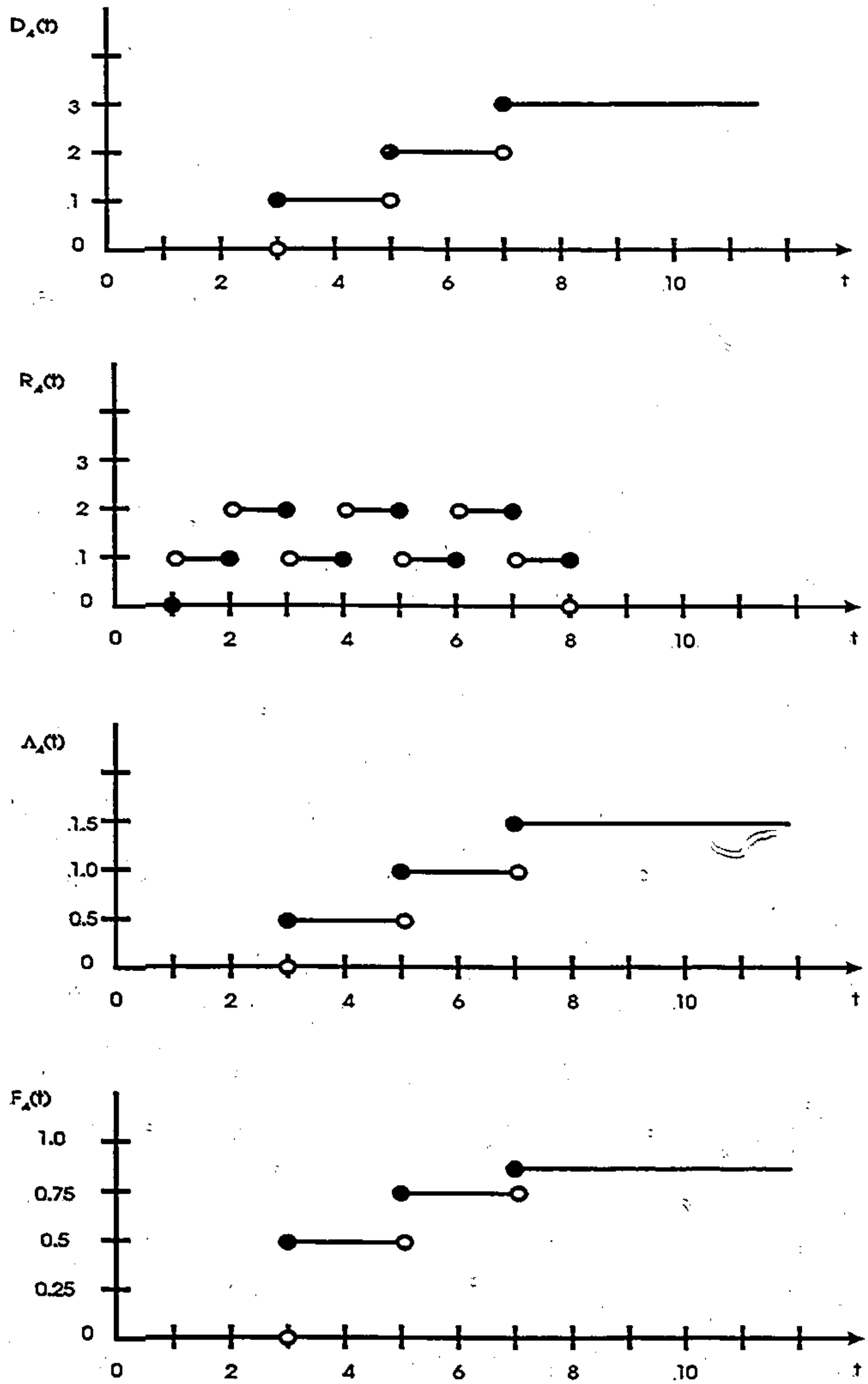

Figure 3.2 Plots of D, R, Λ, and F for Example 3.1.

The second useful result is obtained from Lemma 3.9 of [W1]; it states that the expected number of renewals observed at ages up to and including t satisfies

$$\mathbb{E}\tfrac{1}{n}D_n(t) = \frac{1}{\mu}\int_{[0,t]} \bar{s}\ \ dF(s) \qquad \text{for } 0\leq t<\infty\ . \tag{8}$$

It follows from (8) and the assumption that $F(\gamma^-)=0$ that

$$\mathbb{E}\tfrac{1}{n}D_n(t) = 0 \qquad\qquad \text{if } t<\gamma \tag{9}$$

since, for such t,

$$0 \leq \mathbb{E}\tfrac{1}{n}D_n(t) = \frac{1}{\mu}\int_{[0,t]} s\ dF(s) \leq \frac{\gamma}{\mu}\left(F(\gamma^-) - F(0)\right) = 0\ . \tag{10}$$

The combination of these two results leads to the following alternative representation of the hazard function.

Lemma 3.1 For every $t\in\mathbb{R}$

$$\Lambda(t) = \int_{[0,t]} \frac{1}{\mathbb{E}\frac{1}{n}R_n(s)}\ d(\mathbb{E}\tfrac{1}{n}D_n)(s)\ . \tag{11}$$

Proof By (6) and (8), and since $F(0)=0$ by assumption,

$$\int_{[0,t]} \frac{1}{\mathbb{E}\frac{1}{n}R_n(s)}\ d(\mathbb{E}\tfrac{1}{n}D_n)(s) = \int_{(0,t]} \frac{1}{\frac{\bar{s}}{\mu}\,\bar{F}(s)}\ \frac{\bar{s}}{\mu}\ dF(s)$$

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

$$= \Lambda(t). \qquad \square$$

Variation properties of the number at risk

If f is a function of bounded variation on the bounded interval [a,b], such that $f(x) \geq c > 0$ on [a,b], then $\frac{1}{f}$ is also of bounded variation and the total variation, V_a^b, satisfies

$$V_a^b(\tfrac{1}{f}) \leq \frac{1}{c^2} V_a^b(f) . \tag{12}$$

Furthermore, if f and g are functions of bounded variation on [a,b], then so is fg and

$$V_a^b(fg) \leq V_a^b(f) \sup_{a \leq x \leq b} |g(x)| + V_a^b(g) \sup_{a \leq x \leq b} |f(x)| . \tag{13}$$

The proofs of these results are straightforward and are provided in (A3) and (A4) of the Appendix.

Lemma 3.2 For σ such that $\gamma \leq \sigma < \infty$ and $\bar{F}(\sigma) > 0$,

$$V_\gamma^\sigma((\mathbb{E}\tfrac{1}{n}R_n)^{-1}) \leq \frac{2\sigma\mu}{(\gamma\bar{F}(\sigma))^2} . \tag{14}$$

Proof From (13) and (6) it follows that

$$V_\gamma^\sigma(\mathbb{E}\tfrac{1}{n}R_n) \leq \frac{\sigma}{\mu}\cdot 1 + \frac{\sigma}{\mu}\cdot 1 = \frac{2\sigma}{\mu} .$$

Consequently, (12) and (7) imply that

$$V_\gamma^\sigma((\mathbb{E}\tfrac{1}{n}R_n)^{-1}) \leq \left(\frac{\gamma\bar{F}(\sigma)}{\mu}\right)^{-2} \cdot \frac{2\sigma}{\mu} = \frac{2\sigma\mu}{(\gamma\bar{F}(\sigma))^2} . \qquad \square$$

Kiefer's result and its consequences

In [K2] Keifer obtains a result which can be shown (see (A5)) to imply the following. For $d \in \mathbb{N}^+$, let $\left((X_i^{(1)}, \ldots , X_i^{(d)}) : i = 1,2,\ldots\right)$ be an independent identically distributed sequence of

$\mathbb{R}^d$-valued random vectors on a probability space $(\Omega',\mathcal{F}',P')$. Then there exists $\Omega_0\in\mathcal{F}'$ with $P'(\Omega_0)=1$ such that if $\omega\in\Omega_0$ then one can find $n_0(\omega)$ for which $n\geq n_0(\omega)$ implies

$$\sup_{a\in\mathbb{R}^d}\left|Q_n^*(a,\omega) - Q(a)\right| \leq \sqrt{n^{-1}\log n} \tag{15}$$

where, for $a= (a_1,a_2, \dots a_d)$,

$$Q_n^*(a) = \frac{1}{n}\sum_{r=1}^{n} I[X_r^{(1)}\leq a_1, \dots, X_r^{(d)}\leq a_d] \quad \text{and} \quad Q(a)= \mathbb{E}Q_n^*(a)\,, \tag{16}$$

i.e., $Q(a) = P'[X_1^{(1)}\leq a_1, \dots, X_1^{(d)}\leq a_d]$. The inequality (15) implies the multidimensional Glivenko-Cantelli theorem, and it is stronger in the sense that it provides information on the rate of convergence. In our application it leads to the following lemma.

Lemma 3.3 There exists $\Omega_0\in\mathcal{F}$ with $P(\Omega_0)=1$ and where, for each $\omega\in\Omega_0$, there is a $n_0(\omega)$ such that if $n\geq n_0(\omega)$ then

$$\sup_{t\in\mathbb{R}}\left|\tfrac{1}{n}D_n(t,\omega) - \mathbb{E}\tfrac{1}{n}D_n(t)\right| \leq \sqrt{n^{-1}\log n} \tag{17}$$

and

$$\sup_{t\in\mathbb{R}}\left|\tfrac{1}{n}R_n(t,\omega) - \mathbb{E}\tfrac{1}{n}R_n(t)\right| \leq 2\sqrt{n^{-1}\log n}\,. \tag{18}$$

Proof From (1),

$$\tfrac{1}{n}D_n(t) = \frac{1}{n}\sum_{r=1}^{n} I[Z^{(r)}\leq T,\ Y^{(r)}+Z^{(r)}\leq t]\,.$$

As this is the value at $(T,t)\in\mathbb{R}^2$ of the two-dimensional empirical distribution function of $(Z^{(r)},Y^{(r)}+Z^{(r)})$, (15) implies (17).

From (2),

$$\tfrac{1}{n}R_n(t) = \frac{1}{n}\sum_{r=1}^{n} I[\,Y^{(r)}<t\leq Y^{(r)}+\tilde{Z}^{(r)}\,]$$

$$= \frac{1}{n}\sum_{r=1}^{n} I[Y^{(r)}<t] - \frac{1}{n}\sum_{r=1}^{n} I[\,Y^{(r)}<t,\ Y^{(r)}+\tilde{Z}^{(r)}<t]\,.$$

As (15) clearly remains valid when some "$\leq$"s are replaced by "$<$" in (16), (18) follows. □

Convergence of the hazard function estimator

The above results are used in the proof of the following theorem to show the strong uniform consistency of the hazard function estimator and to provide a bound on the rate of convergence.

Theorem 3.4 Consider $\gamma \leq \sigma < \infty$ with $F(\sigma^-) < 1$. There exists $\Omega^* \in \mathfrak{F}$ with $P(\Omega^*)=1$ such that for each $\omega \in \Omega^*$

$$\sup_{0 \leq t \leq \sigma} |\Lambda_n(t,\omega) - \Lambda(t)| \leq c_0 \sqrt{n^{-1} \log n} \tag{19}$$

for some constant c_0 depending on F and σ but not on ω and all sufficiently large n.

Proof Consider $t \in [0,\sigma]$. From (4) and (11),

$$|\Lambda_n(t) - \Lambda(t)| = \left| \int_{[0,t]} \frac{1}{\frac{1}{n}R_n(s)} \, d(\tfrac{1}{n}D_n)(s) - \int_{[0,t]} \frac{1}{\mathbb{E}\frac{1}{n}R_n(s)} \, d(\mathbb{E}\tfrac{1}{n}D_n)(s) \right|$$

$$\leq \left| \int_{[0,t]} \left(\frac{1}{\frac{1}{n}R_n(s)} - \frac{1}{\mathbb{E}\frac{1}{n}R_n(s)} \right) d(\tfrac{1}{n}D_n)(s) \right|$$

$$+ \left| \int_{[0,t]} \frac{1}{\mathbb{E}\frac{1}{n}R_n(s)} \, d(\tfrac{1}{n}D_n - \mathbb{E}\tfrac{1}{n}D_n)(s) \right|$$

$$\leq A_n(t) + B_n(t)$$

where

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

and

$$B_n(t) = \left| \int_{[0,t]} \frac{1}{\mathbb{E}\frac{1}{n}R_n(s)} \, d(\tfrac{1}{n}D_n - \mathbb{E}\tfrac{1}{n}D_n)(s) \right| .$$

To obtain a bound for $A_n(t)$ we make the following observations.

• Let $0 \le t_j < \gamma$ and $t_j \to \gamma$ as $j \to \infty$. Then (9) and the fact that $D_n(\cdot,\omega)$ is non-negative and increasing imply that $\frac{1}{n}D_n(t,\omega)=0$ for $0 \le t \le t_j$ for all $\omega \in \Omega_j$ for some $\Omega_j \in \mathcal{F}$ with $P(\Omega_j)=1$. Let $\Omega_0' = \cap \Omega_j$; then $P(\Omega_0')=1$ and, for $\omega \in \Omega_0'$, $D_n(t,\omega)=0$ if $t \in [0,\gamma)$. Therefore $A_n(\gamma^-) = 0$ on Ω_0'. Similarly, $B_n(\gamma^-)=0$ on some Ω_0'' where $P(\Omega_0'')=1$.

In the following assume that $\omega \in \Omega^* = \Omega_0' \cap \Omega_0'' \cap \Omega_0$, where Ω_0 is as in Lemma 3.3, let $n_0(\omega)$ be as in that lemma, and suppose that $\gamma \le t \le \sigma$.

• Let n_1 be the smallest n satisfying $2\sqrt{n^{-1}\log n} \le \frac{\gamma\bar{F}(\sigma)}{2\mu}$. If $n \ge \max\{n_0(\omega), n_1\}$, then (7) and (18) imply

$$\tfrac{1}{n}R_n(t,\omega) \ge \mathbb{E}\tfrac{1}{n}R_n(t) - 2\sqrt{n^{-1}\log n} \ge \frac{\gamma\bar{F}(\sigma)}{2\mu} .$$

• $\frac{1}{n}D_n(\cdot,\omega)$ is a probability subdistribution function.

Combining these observations with (7) and (18) we have, at ω and for $n \ge \max\{n_0(\omega), n_1\}$,

$$A_n(t) \le A_n(\gamma^-) + \frac{2\sqrt{n^{-1}\log n}}{\frac{\gamma\bar{F}(\sigma)}{2\mu}\,\frac{\gamma\bar{F}(\sigma)}{\mu}} \int_{[\gamma,t]} d(\tfrac{1}{n}D_n)(s)$$

$$\le \frac{4\mu^2\sqrt{n^{-1}\log n}}{(\gamma\bar{F}(\sigma))^2} .$$

To obtain a bound for $B_n(t)$ we make the following observations.

• Let $f(s) = (\mathbb{E}\frac{1}{n}R_n(s))^{-1}$ and $g(s) = \frac{1}{n}D_n(s,\omega) - \mathbb{E}\frac{1}{n}D_n(s)$. Note that f is left-continuous, see (6), and g is right-continuous. By (14), f is of bounded variation over $[\gamma,\sigma]$ and, since g is the difference of two monotonic functions, it is of bounded variation as well. Applying the general formula for integration by parts, see (A6),

$$B_n(t,\omega) \le B_n(\gamma^-,\omega) + \left| \int_{\{\gamma\}} f(s)\,dg(s) + \int_{(\gamma,t)} f(s^-)\,dg(s) + \int_{\{t\}} f(s)\,dg(s) \right|$$

$$= \left| \int_{\{\gamma\}} f(s)\,dg(s) + f(t^-)g(t^-) - f(\gamma^+)g(\gamma^+) - \int_{(\gamma,t)} g(s^+)\,df(s) + \int_{\{t\}} f(s)\,dg(s) \right| .$$

From (7) and (17), for all $n \ge n_0(\omega)$,

$$|f(t^-)g(t^-)| \le \frac{\sqrt{n^{-1}\log n}}{\frac{\gamma\bar{F}(\sigma)}{\mu}}, \quad |f(\gamma^+)g(\gamma^+)| \le \frac{\sqrt{n^{-1}\log n}}{\frac{\gamma\bar{F}(\sigma)}{\mu}} .$$

From (14) and (17), for all $n \ge n_0(\omega)$,

$$\left| \int_{(\gamma,t)} g(s^+)\,df(s) \right| \le \sup_{\gamma \le s \le t} |g(s)| \cdot V_\gamma^t(f) \le \sqrt{n^{-1}\log n}\, \frac{2\sigma\mu}{(\gamma\bar{F}(\sigma))^2} .$$

From (7) and (17), for $n \ge n_0(\omega)$,

$$\left| \int_{\{t\}} f(s)\,dg(s) \right| \le f(t)\,|g(t^+) - g(t^-)| \le \frac{\mu}{\gamma\bar{F}(\sigma)}\, 2\sqrt{n^{-1}\log n}$$

and, similarly,

$$\left| \int_{\{\gamma\}} f(s)\,dg(s) \right| \le \frac{\mu}{\gamma\bar{F}(\sigma)}\, 2\sqrt{n^{-1}\log n} .$$

Combining these observations we obtain

$$B_n(t,\omega) \le \frac{\mu}{\gamma\bar{F}(\sigma)} \left(6 + \frac{2\sigma}{\gamma\bar{F}(\sigma)}\right) \sqrt{n^{-1}\log n} .$$

Combining the bounds for $A_n(t)$ and $B_n(t)$ shows that

$$|\Lambda_n(t) - \Lambda(t)| \le \frac{\mu}{\gamma\bar{F}(\sigma)} \left(6 + \frac{2\sigma+4\mu}{\gamma\bar{F}(\sigma)}\right) \sqrt{n^{-1}\log n}$$

for each $\omega \in \Omega^*$ and for all $n \geq \max\{n_0(\omega), n_1\}$. The claim of the theorem follows. □

Convergence of the product-limit estimator

The previous theorem shows the strong uniform consistency of the hazard function estimator over the interval $[0,\sigma]$ when $\bar{F}(\sigma)>0$. It remains to show that this implies the convergence of the product-limit estimator to the true d.f. Recall that Φ is the inverse transformation of the hazard function, defined in (2-8), and ρ_t is the sup-norm metric defined in (2-2). Then we have the following result.

Theorem 3.5 For any two hazard functions Λ' and Λ'', with $F''=\Phi(\Lambda'')$ and $1-F''(\sigma^-)>0$,

$$\rho_t(\Phi(\Lambda'),\Phi(\Lambda'')) \leq \frac{7}{\left(1-F''(\sigma^-)\right)^2} \cdot \rho_t(\Lambda',\Lambda'') \tag{20}$$

for $0 \leq t \leq \sigma$.

Proof It is shown in [W2] that if F' and F'' are life distributions, Λ' and Λ'' are their hazard functions, and $t \in [0,\infty)$ is such that $F''(t)<1$ and $\Lambda'(t)<\infty$, then

$$\rho_t(F',F'') \leq \frac{5}{\left(1-F''(t)\right)^2}\, \rho_t(\Lambda',\Lambda'').$$

For $t<\sigma$, (20) follows immediately from this result. With $t\uparrow\sigma$ this result yields

$$\left|\Phi(\Lambda')(\sigma^-)-\Phi(\Lambda'')(\sigma^-)\right| \leq \frac{5}{\left(1-F''(\sigma^-)\right)^2}\, \rho_\sigma(\Lambda',\Lambda'')$$

which is used in the following.

If $t<\sup\{t\in R:\ \Lambda'(t)<\infty\}$ then $1-\Phi(\Lambda'(t))=\{1-\Phi(\Lambda')(t^-)\}(1-\Delta\Lambda'(t))$ from the inversion formula (2-8); this can be shown [W2] to be true for all $t \in \mathbb{R}$ provided that $\Delta\Lambda'(t)$ is defined to be zero when $\Lambda'(t^-)=\infty$; analogous remarks apply to Λ''. Consequently,

$$\begin{aligned}
\left|(\Phi(\Lambda')-\Phi(\Lambda''))(\sigma)\right| &= \left|\ \{1-\Phi(\Lambda'')(\sigma^-)\}(1-\Delta\Lambda''(\sigma)) - \{1-\Phi(\Lambda')(\sigma^-)\}(1-\Delta\Lambda'(\sigma))\ \right| \\
&\leq \max\ \{1-\Phi(\Lambda'')(\sigma^-),\ 1-\Phi(\Lambda')(\sigma^-)\}\cdot \left|\Delta\Lambda''(\sigma)-\Delta\Lambda'(\sigma)\right| \\
&\qquad + \max\{1-\Delta\Lambda''(\sigma),\ 1-\Delta\Lambda'(\sigma)\}\cdot\left|\Phi(\Lambda'')(\sigma^-) - \Phi(\Lambda')(\sigma^-)\right| \\
&\leq \left|\Delta\Lambda''(\sigma)-\Delta\Lambda'(\sigma)\right| + \left|\Phi(\Lambda'')(\sigma^-) - \Phi(\Lambda')(\sigma^-)\right| \\
&\leq 2\cdot\rho_\sigma(\Lambda',\Lambda'') + \frac{5}{\left(1-F''(\sigma^-)\right)^2}\cdot\rho_\sigma(\Lambda',\Lambda'') \\
&\leq \frac{7}{\left(1-F''(\sigma^-)\right)^2}\cdot\rho_\sigma(\Lambda',\Lambda'')
\end{aligned}$$

where the second line follows from the relation

$$|ab-a'b'| \leq \max\{|a|,|a'|\}\cdot|b-b'| + \max\{|b|,|b'|\}\cdot|a-a'|. \qquad \square$$

This result combined with Theorem 3.4 on the convergence of the hazard function estimator proves the following theorem which bounds the rate of convergence of the product-limit estimator to the distribution function F.

Theorem 3.6 Consider $\gamma\leq\sigma<\infty$ with $F(\sigma^-)<1$. There exists $\Omega^*\in\mathfrak{F}$ with $P(\Omega^*)=1$ and some constant c_2 depending on F and σ but not on ω, such that if $\omega\in\Omega^*$ then

$$\sup_{0\leq t\leq\sigma}\left|F_n(t,\omega) - F(t)\right| \leq c_2\sqrt{n^{-1}\log n} \tag{21}$$

for all n greater than some $n_0(\omega)$.

This theorem clearly implies that

$$\lim_{n\to\infty} \rho_\sigma(F_n, F) = 0 \quad \text{a.s.;} \tag{22}$$

that is, there is strong uniform consistency of the product-limit estimator over the interval $[0,\sigma]$. The following theorem shows that, as in the Glivenko-Cantelli theorem, this property holds over the whole real line.

Theorem 3.7

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| F_n(t) - F(t) \right| = 0 \quad \text{a.s.} \tag{23}$$

Proof For $k\in\mathbb{N}^+$, let $\sigma_k=\inf\{s\in\mathbb{R}: F(s) > 1-\frac{1}{k}\}$. Then $F(\sigma_k^-)\leq 1-\frac{1}{k}$ and therefore there exists Ω_k, with $P(\Omega_k)=1$, on which $\lim_{n\to\infty} \rho_{\sigma_k}(F_n, F)=0$.

Consider $\omega\in\cap_1^\infty\Omega_k$ and an arbitrary $\epsilon>0$, and k such that $\frac{1}{k}<\frac{\epsilon}{2}$. Then for $t\leq\sigma_k$ there exists $n_0(\omega)$ such that

$$\sup_{0\leq t\leq\sigma_k} \left| F_n(t,\omega) - F(t) \right| \leq \frac{1}{k} \quad \text{when } n\geq n_0(\omega).$$

For $t>\sigma_k$ and $n\geq n_0(\omega)$

$$1 \geq F(t) > 1 - \frac{1}{k}$$

and

$$1 \geq F_n(t,\omega) \geq F_n(\sigma_k,\omega) \geq F(\sigma_k) - \frac{1}{k} \geq (1-\frac{1}{k}) - \frac{1}{k},$$

hence $|F_n(t,\omega)-F(t)|< \frac{2}{k} < \epsilon$. □

This theorem shows that the product-limit estimator for the stabilized renewal process with single observations is strongly uniformly consistent on $\mathbb{R}$. The proofs of Theorems 3.5 and 3.7 do not depend on the form of the product-limit estimator, but rely only on the transformation Φ which links the hazard estimator with the product-limit estimator. Thus any hazard estimator which has the property $\rho_\sigma(\Lambda_n,\Lambda)\to 0$ a.s. when $\bar{F}(\sigma)>0$ will result in a strongly uniformly consistent product-limit estimator through the application of these two theorems. In particular, we have the following lemma.

Lemma 3.8 For any sequence of hazard functions $(\Lambda_n)_0^\infty$, and their corresponding life distributions F_n obtained via (2-8), if $\rho_\sigma(\Lambda_n,\Lambda_0)\to 0$ whenever $\bar{F}_0(\sigma)>0$ then

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} | F_n(t) - F_0(t) | = 0 \tag{24}$$

Related results

The approach taken in this chapter closely parallels that of [W1]. Winter and Földes use a hazard function estimator of the form

$$\Lambda_n(t) = \sum_{s\le t} \frac{\Delta D_n(s)}{\theta_n + R_n(s)}$$

where (θ_n) is a sequence of positive reals tending to zero, and show that the corresponding product-limit estimator is a strongly uniformly consistent estimator of the underlying distribution function for $0\le t<\sigma$ where $F(\sigma^-)<1$. Here, we have replaced the sequence (θ_n) with the condition $F(\gamma^-)=0$ for some $\gamma>0$ and have obtained similar results. The substantially new result in this chapter is the extension of the range of convergence to the whole real line. This extension relies on a result from [W2] relating the error in the d.f. estimate to that in the hazard estimate.

CHAPTER 4: Product-limit estimator — multiple observations of a stabilized renewal process

In Chapter 3 we looked at the case where one observation is made of each stabilized renewal process. In this chapter we will consider the case where observations continue until the censoring time T. Recall that for a stabilized renewal process $((Y,Z),X_1,X_2, \dots)$ the random variable Y represents the age of the first object at the time when the observations begin, Z represents the remaining life of the first object, and $X_1,X_2,\dots$ represent the subsequent inter-renewal times. In the following we carefully distinguish between age and time: the former refers to the age of an individual object, while the latter represents the total length of time observations have been made on the process. The censoring time, T, refers to the total length of time the process is observed; and during this time zero, one, two, up to any number of renewals could have been observed. All observed ages, after the first, must surely be less than T. We begin the investigation of this problem by presenting the following useful result from renewal theory.

Expected number of renewals

For the stabilized renewal process $(Y,Z,X_1,X_2,\dots)$ put

$$S_n = Z + \sum_{k=1}^{n} X_k, \qquad n \geq 0 \tag{1}$$

and

$$N(t) = \sum_{k=0}^{\infty} I[S_k \leq t] . \tag{2}$$

The partial sum S_n represents the waiting time to the $(n+1)th$ renewal (the first renewal occurs at $S_0=Z$), and $N(t)$ represents the number of renewals observed up to and including time t. Observe that $(Z,X_1,X_2,...)$ is an independent sequence of r.v.s and, see (1-6),

$$P[Z\leq t] = \int_{[0,t]} \frac{1-F(s)}{\mu} \, ds . \tag{3}$$

Thus $(Z,X_1,X_2,...)$ is a *stationary renewal process* [F1] with underlying d.f. F. As $N(t)$ also represents the number of renewals (up to and including time t) in this stationary renewal process, it follows that

$$\begin{aligned} E[N(t)] &= \sum_{k=0}^{\infty} P[S_k\leq t] \\ &= \frac{t}{\mu} \qquad \text{if } 0\leq t<\infty ; \end{aligned} \tag{4}$$

see, for example, §XI.4 in [F1].

Estimation equations

When observing one stabilized renewal process the number of observed inter-renewal times of t or less is represented by the r.v.

$$\begin{aligned} D(t) &= I[Y+\tilde{Z}\leq t, Z\leq T] + \sum_{k=1}^{\infty} I[X_k\leq t, S_k\leq T] \\ &= I[Y+Z\leq t, S_0\leq T] + \sum_{k=1}^{\infty} I[X_k\leq t, S_k\leq T] . \end{aligned} \tag{5}$$

The number of objects which are under observation at age t is represented by

$$R(t) = I[Y<t\leq Y+\tilde{Z}] + \sum_{k=1}^{\infty} I[X_k\geq t, S_{k-1}\leq T-t] \tag{6}$$

$$= I[Y+\tilde{Z}\geq t] - I[Y\geq t] + \sum_{k=1}^{\infty} I[X_k\geq t,\ T-S_{k-1}\geq t]. \tag{7}$$

Observe that, for any age t, $D(t)\leq N(T)$ and $R(t)\leq 1+N(T-t)$. Since $\mathbb{E}N(T)=\frac{T}{\mu}<\infty$, it follows that $P[N(T)<\infty]=1$ and, for every $\omega\in[N(T)<\infty]$, $D(\cdot,\omega)$ is right-continuous and increasing, and $R(\cdot,\omega)$ is left-continuous. Equation (7) is useful because it expresses $R(\cdot,\omega)$ as a difference of decreasing functions and shows that $R(\cdot,\omega)$ is of bounded variation on $\mathbb{R}$ when $\omega\in[N(T)<\infty]$.

As in Chapter 3 we shall assume that observations are made on a sequence of independent copies of this process, the r*th* one being denoted by $(Y^{(r)},Z^{(r)},X_1^{(r)},X_2^{(r)}, \ldots)$. Define $D^{(r)}(t)$ and $R^{(r)}(t)$ in a manner analogous to (5) and (6), respectively, and let

$$\tfrac{1}{n}D_n(t) = \tfrac{1}{n}\sum_{r=1}^{n} D^{(r)}(t) \tag{8}$$

and

$$\tfrac{1}{n}R_n(t) = \tfrac{1}{n}\sum_{r=1}^{n} R^{(r)}(t)\ . \tag{9}$$

To simplify notation in the following we will make use of the observation that $\mathbb{E}D(t)=\mathbb{E}\frac{1}{n}D_n(t)$ and $\mathbb{E}R(t)=\mathbb{E}\frac{1}{n}R_n(t)$.

To estimate the hazard function Λ, we use a Nelson hazard estimator, analogous to (2-11):

$$\Lambda_n(t) = \sum_{s\leq t} \frac{\Delta D_n(s)}{R_n(s)}\ . \tag{10}$$

Again, it is useful to note that Λ_n is given by the sample-path integral

$$\Lambda_n(t) = \int_{[0,t]} \frac{1}{R_n(s)}\ dD_n(s)\ . \tag{11}$$

In this case, the product-limit estimator of F is again, from the inversion formula:

$$F_n(t) = 1 - \prod_{s\leq t} \left(1 - \frac{\Delta D_n(s)}{R_n(s)} \right). \tag{12}$$

Clearly there is no difference in form between the estimators of this chapter and those of Chapter 3. The difference lies in the processes D_n and R_n. Also note that in this chapter we do not assume the existence of $\gamma>0$ such that $F(\gamma^-)=0$ or even that $F(0)=0$; we only require that $F(0^-)=0$.

Alternative representation of the hazard function

Following the approach taken in Chapter 3, we first determine the expected values of the number at risk and the number of renewals.

Lemma 4.1 $\qquad \mathbb{E}R(t) = \frac{T}{\mu}\,\bar{F}(t) \qquad \text{for } t\geq 0.$ (13)

Proof Using (3-2) and (3-6) with n=1, as well as (4),

$$\begin{aligned}
\mathbb{E}R(t) &= P[Y<t\leq Y+\tilde{Z}] + \sum_{k=1}^{\infty} P[X_k\geq t,\, S_{k-1}\leq T-t] \\
&= \frac{\tilde{t}}{\mu}\,\bar{F}(t) + \sum_{k=1}^{\infty} P[X_k\geq t]\, P[S_{k-1}\leq T-t] \\
&= \frac{\tilde{t}}{\mu}\,\bar{F}(t) + \bar{F}(t)\sum_{k=1}^{\infty} P[S_{k-1}\leq T-t] \\
&= \frac{\tilde{t}}{\mu}\,\bar{F}(t) + \bar{F}(t)\sum_{k=0}^{\infty} P[S_k\leq T-t] \\
&= \frac{\tilde{t}}{\mu}\,\bar{F}(t) + \bar{F}(t)\,\mathbb{E}N(T-t) \\
&= \frac{\tilde{t}}{\mu}\,\bar{F}(t) + \bar{F}(t)\cdot\frac{T-t}{\mu}\, I[t\leq T] \\
&= \bar{F}(t)\,\frac{T}{\mu}\;.
\end{aligned}$$

□

As the proof of the Lemma 4.1 indicates, care must be taken to restrict all ages except the first, as well as the total observation time, to be less than T. For example, the situation where the first renewal does not occur before observation ceases at time T is quite possible.

Lemma 4.2

$$\mathbb{E}D(t) = \frac{T}{\mu} F(t). \tag{14}$$

Proof Using (3-8) with n=1, and Theorem 20.3 of [B1],

$$\mathbb{E}D(t) = P[Y+\tilde{Z}\leq t, Z\leq T] + \sum_{k=1}^{\infty} P[X_k\leq t, S_k\leq T]$$

$$= \frac{1}{\mu} \int_{[0,t]} \bar{s}\ dF(s) + \sum_{k=1}^{\infty} \int_{[0,t]} P[S_{k-1}\leq T-s]\ dF(s) \ .$$

Since the integrand is nonnegative, the order of summation and integration can be interchanged in the above expression. Using (4), this results in

$$\mathbb{E}D(t) = \frac{1}{\mu} \int_{[0,t]} \bar{s}\ dF(s) + \int_{[0,t]} \sum_{k=0}^{\infty} P[S_k\leq T-s]\ dF(s)$$

$$= \int_{[0,t]} \frac{\bar{s}}{\mu}\ dF(s) + \int_{[0,t]} \frac{(T-s)}{\mu} I_{[0,T]}(s)\ dF(s)$$

$$= \int_{[0,t]} \frac{T}{\mu}\ dF(s) = \frac{T}{\mu} F(t). \qquad \square$$

The two preceding lemmas lead to an alternative representation of the hazard function analogous to Lemma 3.1.

Lemma 4.3

$$\mathbb{E}D(t) = \int_{[0,t]} \mathbb{E}R(s)\ d\Lambda(s) \tag{15}$$

and

$$\Lambda(t) = \int_{[0,t]} \frac{1}{\mathbb{E}R(s)} \, d(\mathbb{E}D)(s) \, . \tag{16}$$

Proof From Lemmas 4.1 and 4.2,

$$\int_{[0,t]} \mathbb{E}R(s) \, d\Lambda(s) = \int_{[0,t]} \tfrac{T}{\mu} \, \bar{F}(s) \cdot \frac{1}{\bar{F}(s)} \, dF(s) = \tfrac{T}{\mu} \, F(t) = \mathbb{E}D(t).$$

Cancelling $\bar{F}(s)$ is permitted when $\bar{F}(t)>0$, and it is easily shown that the result is in fact true for all t as a consequence of the convention that $0 \cdot \frac{1}{0} = 0$.

From (13) and (14),

$$\int_{[0,t]} \frac{1}{\mathbb{E}R(s)} \, d(\mathbb{E}D)(s) = \int_{[0,t]} \frac{1}{\frac{T}{\mu}\bar{F}(s)} \, \tfrac{T}{\mu} \, dF(s) = \int_{[0,t]} \frac{1}{\bar{F}(s)} \, dF(s) = \Lambda(t) \, . \qquad \square$$

Convergence of the hazard function estimator

We now consider the problem of showing the convergence of the hazard function estimator to the true hazard function. In [G2] Gill examined a similar problem in the case of an ordinary renewal process, and the following lemma is analogous to a result he establishes in the first part of the proof of Lemma 2 of [G2]. A proof of this lemma can be found in (A7) of the Appendix.

Lemma 4.4 Assume $0<\sigma<\infty$ and $\bar{F}(\sigma)>0$. If $\frac{T}{\mu}\bar{F}(\sigma) - \rho_\sigma(\mathbb{E}R, \frac{1}{n}R_n)>0$ then, on $[N(T)<\infty]$

$$\rho_\sigma(\Lambda, \Lambda_n) \le 2\frac{\rho_\sigma(\mathbb{E}D, \frac{1}{n}D_n)}{\frac{T}{\mu}\bar{F}(\sigma)} + \frac{\rho_\sigma(\mathbb{E}R, \frac{1}{n}R_n)\,(\frac{T}{\mu} + \rho_\sigma(\mathbb{E}D, \frac{1}{n}D_n))}{\frac{T}{\mu}\bar{F}(\sigma)(\frac{T}{\mu}\bar{F}(\sigma) - \rho_\sigma(\mathbb{E}R, \frac{1}{n}R_n))} \, . \tag{17}$$

This lemma shows that $\rho_\sigma(\Lambda, \Lambda_n) \rightarrow 0$ whenever $\rho_\sigma(\frac{1}{n}D_n, \mathbb{E}D)$ and $\rho_\sigma(\frac{1}{n}R_n, \mathbb{E}R)$ converge to 0. Clearly to use Lemma 4.4 to show the strong uniform consistency of the hazard estimator for a stabilized renewal processes, we must first show that $\rho_\sigma(\frac{1}{n}R_n, \mathbb{E}R) \rightarrow 0$ and $\rho_\sigma(\frac{1}{n}D_n, \mathbb{E}D) \rightarrow 0$ a.s. The following lemma provides this result. A proof of this lemma is given in (A8) of the Appendix.

Lemma 4.5 Let $h_i:\mathbb{R}\times\Omega\rightarrow\mathbb{R}$ for i=1,2,... and let σ be a positive real. Suppose that either

(a) for every fixed ω, $h_i(\cdot,\omega)$ is a right-continuous, nonnegative, increasing function and, for fixed $t\in[0,\sigma]$, $(h_i(t^-,\cdot)$: i=1,...) and $(h_i(t,\cdot)$: i=1,...) are independent sequences of nonnegative, identically distributed r.v.s of finite mean; or

(b) for every fixed ω, $h_i(\cdot,\omega)$ is a left-continuous, nonnegative, decreasing function and, for fixed $t\in[0,\sigma]$, $(h_i(t^+,\cdot)$: i=1,...) and $(h_i(t,\cdot)$: i=1,...) are independent sequences of nonnegative, identically distributed r.v.s of finite mean

holds. Let $H(t) = \mathbb{E}h_1(t)$ and

$$H_n(t,\omega) = \frac{1}{n}\sum_{k=1}^{n} h_k(t,\omega) . \tag{18}$$

Then $\rho_\sigma(H_n, H) \rightarrow 0$ a.s.

Part (a) of Lemma 4.5 can clearly be applied to D_n, and if we note by (7) that R_n can be written as a difference of two functions satisfying the conditions (b), this lemma can also be applied to R_n. These observations combined with Lemma 4.4 lead to the following hazard function consistency result.

Theorem 4.6 Consider $\sigma<\infty$ with $\bar{F}(\sigma)>0$. The hazard estimator for the stabilized renewal process with multiple observations is strongly uniformly consistent over $[0,\sigma]$; that is,

$$\lim_{n\rightarrow\infty} \rho_\sigma(\Lambda_n,\Lambda) = 0 \text{ a.s.} \tag{19}$$

Proof Since

$$D(t^-,\omega) = I[Y+Z<t,\, S_0 \le T] + \sum_{k=1}^{\infty} I[X_k<t,\, S_k \le T]$$

and the series becomes a finite sum on $[N(T)<\infty]$, the requirements on $(h_i\colon i=1,\ldots)$, H_n, and H in Lemma 4.5 are met on $[N(T)<\infty]$ by $(D^{(i)}\colon i=1,\ldots)$, $\frac{1}{n}D_n$, and $\mathbb{E}D$, respectively. Thus, $\rho_\sigma(\frac{1}{n}D_n,\mathbb{E}D)\rightarrow 0$ a.s.

Let f^1 and f^2 be the two terms in the decomposition of R_n into left-continuous decreasing functions given in (7),

$$f^1 = I[Y+\bar{Z}\ge t] + \sum_{k=1}^{\infty} I[X_k\ge t,\, S_{k-1}\le T-t] \quad \text{and} \quad f^2 = I[Y\ge t],$$

and define $\frac{1}{n}f_n^i$ in a manner analogous to (9). Then noting that remarks analogous to those made about $D(t^-,\omega)$ also apply to $R(t^+,\omega)$, we have

$$\rho_\sigma(\tfrac{1}{n}R_n,\mathbb{E}R) \le \rho_\sigma(\tfrac{1}{n}f_n^1\,,\,\mathbb{E}f^1) + \rho_\sigma(\tfrac{1}{n}f_n^2\,,\,\mathbb{E}f^2) \rightarrow 0 \text{ a.s.} \tag{20}$$

since each of the two functions in the decomposition satisfies the conditions of Lemma 4.5 on $[N(T)<\infty]$. Substituting these results in (17) proves (19). □

While this result proves convergence of the estimate to the hazard function a.s., it does not guarantee a certain rate of convergence as the results of Chapter 3 did. This can be attributed to the fact that Kiefer's result does not seem to be applicable in this situation.

Convergence of the product-limit estimator

At this point we can appeal to some of the results in Chapter 3 which do not depend on the particular hazard or product-limit estimator being used, but only rely on the transformation Φ linking the two. In particular, Lemma 3.8 can be applied directly to obtain the following:

Theorem 4.7 The product-limit estimator for multiple observations of a stabilized renewal process is strongly uniformly consistent; that is,

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \mid F_n(t) - F(t) \mid = 0 \quad \text{a.s.} \tag{21}$$

Proof From Theorem 4.6, $\rho_\sigma(\Lambda_n,\Lambda)\to 0$ a.s. whenever $\bar{F}(\sigma)>0$, and under this condition Lemma 3.8 implies (21). □

This is a very simple yet important result, and it comes, almost free using the tools of Chapter 3, from the convergence result for the hazard function estimator.

Related results

This is the first treatment of stabilized renewal processes with multiple observations of which we are aware. In that sense all results concerning these processes are new, but clearly the approach taken parallels that of Chapter 3 and [W1]. In [G2] Gill considered the related case of estimating the d.f. of a renewal process when the observations commence with the process. He was able to show the strong uniform consistency of the product-limit estimator over the interval $[0,\sigma]$ where $\bar{F}(\sigma)>0$. We have drawn Lemma 4.4 from Gill's work. However, because of the nature of our process (its open-endedness on the left), we are able to obtain the Glivenko-Cantelli-like result over the whole line.

Concluding remarks

This thesis has shown the strong uniform consistency of the product-limit estimator when it is applied to the censored observations of a stabilized renewal process. It has extended the results of [W1] in the case of a single observation per stabilized renewal process by showing that, as in the Glivenko-Cantelli theorem, the strong uniform consistency holds over the whole real line. This thesis is believed to be the first instance where the case of multiple observations of a stabilized renewal process terminated by a deterministic censoring time is investigated and the strong uniform consistency of the product-limit estimator is shown.

One area which we would have liked to have treated in this thesis was the weak convergence properties of the product-limit estimator when applied to the above situations. Investigations lead us to believe that useful results in this area are achievable, but we were unable to rigourously develop these results for presentation here.

Appendix

(A1) To define Lebesgue-Stieltjes integrals when the integrator F is not necessarily increasing but is of bounded variation over [c,d] we let the variation of F over [c,t] for $c \leq t \leq d$ be denoted by $V_c^t(F)$, and define F^1 and F^2 by

$$F^1(t) = \begin{cases} 0 & t<c \\ V_c^t(F) & c \leq t \leq d \\ V_c^d(F) & t>d \end{cases}$$

and

$$F^2(t) = \begin{cases} -F(c) & t<c \\ V_c^t(F)-F(t) & c \leq t \leq d \\ V_c^d(F)-F(d) & t>d. \end{cases}$$

Then, for $A \in (c,d) \cap \mathfrak{B}$ and measurable $\psi: \mathbb{R} \to \mathbb{R}$, if $\int_A \psi \, dF^1$ is finite, $\int_A \psi \, dF$ is considered defined and equals

$$\int_A \psi \, dF = \int_A \psi \, dF^1 - \int_A \psi \, dF^2$$

where these latter integrals are ordinary Lebesgue-Stieltjes integrals.

(A2) Lemma Let $X,Y:\mathbb{R}\times\Omega\to\mathbb{R}$ be increasing functions which are left- and right-continuous, respectively. Assume both $X(t,\cdot)$ and $Y(t,\cdot)$ are $\mathcal{G}$-measurable for $a-\delta\leq t\leq b$, for some $\delta>0$ and $-\infty<a\leq b<\infty$. Then the function

$$\omega\mapsto\Big(\int_{[a,b]} X\,dY\Big)(\omega) \tag{1}$$

is $\mathcal{G}/\mathcal{B}$ measurable.

Proof Let P_n be a sequence of partitions of $[a,b]$ defined by

$$P_n:\ a = s_0^{(n)}<s_1^{(n)}< \dots <s_{2^n}^{(n)} = b$$

where $s_i^{(n)}= a + i\frac{b-a}{2^n}$. Let $\Delta_1^{(n)}= [s_0^{(n)},s_1^{(n)}]$ and $\Delta_i^{(n)}= (s_i^{(n)},s_{i+1}^{(n)}]$ for $2\leq i\leq 2^n$. Define the sequence of step functions $X^{(n)}(\cdot,\omega)$ by

$$X^{(n)}(t,\omega) = \sum_{i=1}^{2^n} X(s_{i-1}^{(n)}, \omega)\, I_{\Delta_i^{(n)}}(t)\,.$$

For fixed t and ω, it is straightforward to show that $X^{(n)}(t,\omega)\uparrow X(t,\omega)$ as $n\to\infty$.

From the definition of the Lebesgue-Stieltjes integral, the left-continuity of X, and the right-continuity of Y we have

$$\Big(\int_{[a,b]} X^{(n)}dY\Big)(\omega) = \sum_{i=1}^{2^n} X(s_{i-1}^{(n)},\omega)\Big(Y(s_i^{(n)},\omega) - Y(s_{i-1}^{(n)},\omega)\Big) + X(a,\omega)\Big(Y(a,\omega) - Y(a^-,\omega)\Big).$$

Since the functions $\omega\mapsto X(s,\omega)$, $\omega\mapsto Y(s,\omega)$, and $\omega\mapsto Y(a^-,\omega)$ are $\mathcal{G}$-measurable, it follows that the mapping

$$\omega\mapsto\Big(\int_{[a,b]} X^{(n)}\,dY\Big)(\omega)$$

is $\mathcal{G}$-measurable. By the monotone convergence theorem , for fixed ω,

$$\Big(\int_{[a,b]} X^{(n)}\, dY\Big)(\omega) \to \Big(\int_{[a,b]} X\, dY\Big)(\omega)\ ;$$

consequently [B1, Thm 13.4], $\int_{[a,b]} X\, dY$ is $\mathcal{G}$-measurable. □

Corollary Let $X,Y:\mathbb{R}\times\Omega\to\mathbb{R}$ be, respectively, left- and right-continuous functions which are of bounded variation on a closed interval containing [a,b] in its interior. Assume both $X(t,\cdot)$ and $Y(t,\cdot)$ are $\mathcal{G}$-measurable for $a-\delta\le t\le b$, for some $\delta>0$. Then $\int_{[a,b]} X\, dY$ is $\mathcal{G}$-measurable.

(A3) Lemma Let f be a function of bounded variation on the bounded interval [a,b], such that $f(x)\ge c>0$. Then $\frac{1}{f}$ is also of bounded variation on [a,b] and the total variation, V_a^b, satisfies

$$V_a^b(\tfrac{1}{f}) \le \frac{1}{c^2}\, V_a^b(f)\ . \tag{2}$$

Proof If $a=x_0< \dots < x_m=b$ is a partition of [a,b] then

$$\begin{aligned}\sum_{i=1}^{m} \left|f^{-1}(x_i) - f^{-1}(x_{i-1})\right| &= \sum_{i=1}^{m}\left|\frac{f(x_i) - f(x_{i-1})}{f(x_i)\,f(x_{i-1})}\right| \\ &\le \Big(\inf_{a\le x\le b} |f(x)|\Big)^{-2} \Big(\sum_{i=1}^{m} |f(x_i)-f(x_{i-1})|\Big) \\ &\le c^{-2}\ V_a^b(f).\end{aligned}$$

Since this holds for any partition of the interval [a,b], the stated inequality follows. □

(A4) Lemma If f and g are functions of bounded variation on [a,b], then so is fg and

$$V_a^b(fg) \le V_a^b(f) \sup_{a\le x\le b}|g(x)| \ + \ V_a^b(g) \sup_{a\le x\le b}|f(x)|\ . \tag{3}$$

Proof If $a=x_0< \dots < x_m=b$ is a partition of [a,b] then

$$\sum_{i=1}^{m} |(fg)(x_i) - (fg)(x_{i-1})| = \sum_{i=1}^{m} |\{f(x_i)-f(x_{i-1})\} g(x_i) + \{g(x_i)-g(x_{i-1})\} f(x_{i-1})|$$

$$\leq \sup_{a\leq x\leq b} |g(x)| \sum_{i=1}^{m} |f(x_i)-f(x_{i-1})|$$

$$+ \sup_{a\leq x\leq b} |f(x)| \sum_{i=1}^{m} |g(x_i)-g(x_{i-1})|$$

$$\leq \sup_{a\leq x\leq b} |g(x)|\, V_a^b(f) + \sup_{a\leq x\leq b} |f(x)|\, V_a^b(g) . \qquad \square$$

(A5) For $d\in\mathbb{N}^+$, let $\left((X_i^{(1)}, \dots , X_i^{(d)}) : i= 1,2,\dots\right)$ be an independent identically distributed sequence of $\mathbb{R}^d$-valued random vectors on a probability space $(\Omega',\mathcal{F}',P')$. It is known [K2] that if $1<\alpha<2$ then there exists a constant K, dependent on α and the dimension d but not on the distribution of the random vectors, such that, for any $n\in\mathbb{N}^+$ and any $\delta_n>0$,

$$P'(A_n) < Ke^{-\alpha n\delta_n^2} , \qquad (4)$$

where

$$A_n = \left\{\omega\in\Omega': \sup_{a\in\mathbb{R}^d} |Q_n^*(a,\omega) - Q(a)|>\delta_n\right\} \qquad (5)$$

and, for $a= (a_1, a_2, \dots a_d)$,

$$Q_n^*(a) = \tfrac{1}{n} \textstyle\sum_{i=1}^{n} I[X_i^{(1)}\leq a_1, \dots, X_i^{(d)}\leq a_d] \text{ and } Q(a) = \mathbb{E}Q_n^*(a). \qquad (6)$$

If $\delta_n=\sqrt{n^{-1}\log n}$ then $\Sigma P'(A_n) < \infty$ so by the Borel-Cantelli lemma $P'(\liminf A_n)=0$. This implies that there exists a set $\Omega_0\in\mathcal{F}'$ with $P'(\Omega_0)=1$, such that if $\omega\in\Omega_0$ then, for some $n_0(\omega)$,

$$\sup_{a\in\mathbb{R}^d} |Q_n^*(a,\omega) - Q(a)| \leq \sqrt{n^{-1}\log n} \qquad \text{if } n\geq n_0(\omega). \qquad (7)$$

(A6) A general form of integration by parts is the following. If $-\infty<a<\alpha\leq\beta<b<\infty$ and $f:\mathbb{R}\to\mathbb{R}$ and $g:\mathbb{R}\to\mathbb{R}$ are of bounded variation on [a,b] then

$$\int_{[\alpha,\beta]} f(s^-)\, dg(s) + \int_{[\alpha,\beta]} g(s^+)\, df(s) = f(\beta^+)g(\beta^+) - f(\alpha^-)g(\alpha^-)\,; \tag{8}$$

see, for example, (21.67) of [H1]. It follows that if f and g are of bounded variation on $[\alpha,\beta]$

$$\int_{(\alpha,\beta)} f(s^-)\, dg(s) + \int_{(\alpha,\beta)} g(s^+)\, df(s) = f(\beta^-)g(\beta^-) - f(\alpha^+)g(\alpha^+)\,, \tag{9}$$

and if f and g are of bounded variation on $[a,\beta]$

$$\int_{[\alpha,\beta)} f(s^+)\, dg(s) + \int_{[\alpha,\beta)} g(s^-)\, df(s) = f(\beta^-)g(\beta^-) - f(\alpha^-)g(\alpha^-)\,. \tag{10}$$

(A7) Lemma 4.4 Assume $0<\sigma<\infty$ and $\bar{F}(\sigma)>0$. If $\frac{T}{\mu}\bar{F}(\sigma) - \rho_\sigma(\mathbb{E}R, \frac{1}{n}R_n)>0$ then, on $[N(T)<\infty]$

$$\rho_\sigma(\Lambda, \Lambda_n) \le 2\frac{\rho_\sigma(\mathbb{E}D, \frac{1}{n}D_n)}{\frac{T}{\mu}\bar{F}(\sigma)} + \frac{\rho_\sigma(\mathbb{E}R, \frac{1}{n}R_n)\,(\frac{T}{\mu} + \rho_\sigma(\mathbb{E}D, \frac{1}{n}D_n))}{\frac{T}{\mu}\bar{F}(\sigma)(\frac{T}{\mu}\bar{F}(\sigma) - \rho_\sigma(\mathbb{E}R, \frac{1}{n}R_n))} \tag{11}$$

Proof Consider a fixed $t\le\sigma$ and a fixed $\omega\in[N(T)<\infty]$. For that ω,

$$\Lambda(t) - \Lambda_n(t) = \int_{[0,t]} \frac{1}{\mathbb{E}R(s)}\, d(\mathbb{E}D)(s) - \int_{[0,t]} \frac{1}{\frac{1}{n}R_n(s)}\, d(\tfrac{1}{n}D_n)(s)$$

$$= \int_{[0,t)} \frac{1}{\mathbb{E}R(s)}\, d(\mathbb{E}D(s) - \tfrac{1}{n}D_n)(s) + \frac{\Delta\mathbb{E}D(t) - \Delta\frac{1}{n}D_n(t)}{\mathbb{E}R(t)}$$

$$+ \int_{[0,t]} \Big(\frac{1}{\mathbb{E}R(s)} - \frac{1}{\frac{1}{n}R_n(s)} \Big)\, d(\tfrac{1}{n}D_n)(s)\,.$$

From (4-13), $\mathbb{E}R(s)= \frac{T\bar{F}(s)}{\mu} \ge \frac{T\bar{F}(\sigma)}{\mu}>0$ when $0\le s\le t$, and thus $(\mathbb{E}R)^{-1}$ is of bounded variation on $(-\infty,\sigma]$ and left-continuous. By Lemma 4.2 and the remark following (4-7), $\mathbb{E}D$

and $\frac{1}{n}D_n(\cdot,\omega)$ are of bounded variation on $\mathbb{R}$ and right-continuous. Therefore we can integrate the first term by parts to obtain, see (A5),

$$\int_{[0,t)} \frac{1}{\mathbb{E}R(s)}\, d(\mathbb{E}D(s) - \tfrac{1}{n}D_n)(s) = \frac{\mathbb{E}D(t^-) - \frac{1}{n}D_n(t^-)}{\mathbb{E}R(t^-)} - \frac{\mathbb{E}D(0^-) - \frac{1}{n}D_n(0^-)}{\mathbb{E}R(0^-)}$$

$$- \int_{[0,t)} (\mathbb{E}D(s) - \tfrac{1}{n}D_n(s))\, d(\tfrac{1}{\mathbb{E}R})(s)\,.$$

Noting that

$$\frac{\mathbb{E}D(t^-) - \frac{1}{n}D_n(t^-)}{\mathbb{E}R(t^-)} + \frac{\Delta\mathbb{E}D(t) - \Delta\frac{1}{n}D_n(t)}{\mathbb{E}R(t)} = \frac{\mathbb{E}D(t) - \frac{1}{n}D_n(t)}{\mathbb{E}R(t)}\,,$$

we obtain

$$\Lambda(t) - \Lambda_n(t) = \frac{\mathbb{E}D(t) - \frac{1}{n}D_n(t)}{\mathbb{E}R(t)} - \int_{[0,t)} (\mathbb{E}D(s) - \tfrac{1}{n}D_n(s))\, d(\tfrac{1}{\mathbb{E}R})(s)$$

$$+ \int_{[0,t]} \frac{\frac{1}{n}R_n(s) - \mathbb{E}R(s)}{\mathbb{E}R(s)\, \frac{1}{n}R_n(s)}\, d(\tfrac{1}{n}D_n)(s)\,.$$

This expression can be bounded as follows.

$$\left|\Lambda(t) - \Lambda_n(t)\right| \le \frac{\rho_\sigma(\mathbb{E}D, \frac{1}{n}D_n)}{\mathbb{E}R(t)} + \int_{[0,t)} \rho_\sigma(\mathbb{E}D, \tfrac{1}{n}D_n)\, d(\tfrac{1}{\mathbb{E}R})(s)$$

$$+ \int_{[0,t]} \frac{\rho_\sigma(\frac{1}{n}R_n, \mathbb{E}R)}{\mathbb{E}R(s)\, \frac{1}{n}R_n(s)}\, d(\tfrac{1}{n}D_n)(s)\,.$$

Since $\mathbb{E}R(t) = \frac{T}{\mu}\bar{F}(t) \ge \frac{T}{\mu}\bar{F}(\sigma)$,

$$\frac{\rho_\sigma(\mathbb{E}D, \frac{1}{n}D_n)}{\mathbb{E}R(t)} \le \frac{\rho_\sigma(\mathbb{E}D, \frac{1}{n}D_n)}{\frac{T}{\mu}\bar{F}(\sigma)}$$

and

$$\int_{[0,t)} \rho_\sigma(\mathbb{E}D, \tfrac{1}{n}D_n)\, d(\tfrac{1}{\mathbb{E}R})(s) \le \rho_\sigma(\mathbb{E}D, \tfrac{1}{n}D_n) \left(\frac{1}{\frac{T}{\mu}\bar{F}(\sigma)} - \frac{1}{\frac{T}{\mu}}\right) \le \frac{\rho_\sigma(\mathbb{E}D, \frac{1}{n}D_n)}{\frac{T}{\mu}\bar{F}(\sigma)};$$

and, if $\frac{T}{\mu}\bar{F}(\sigma) - \rho_\sigma(\mathbb{E}R, \frac{1}{n}R_n)>0$, then

$$\int_{[0,t]} \frac{\rho_\sigma(\frac{1}{n}R_n,\mathbb{E}R)}{\mathbb{E}R(s)\,\frac{1}{n}R_n(s)}\, d(\tfrac{1}{n}D_n)(s) \le \int_{[0,t]} \frac{\rho_\sigma(\frac{1}{n}R_n, \mathbb{E}R)}{\mathbb{E}R(s)\left|\mathbb{E}R(s) - \left|\mathbb{E}R(s) - \frac{1}{n}R_n(s)\right|\right|}\, d(\tfrac{1}{n}D_n)(s)$$

$$\le \frac{\rho_\sigma(\frac{1}{n}R_n, \mathbb{E}R)}{\frac{T}{\mu}\bar{F}(\sigma)\left(\frac{T}{\mu}\bar{F}(\sigma) - \rho_\sigma(\frac{1}{n}R_n, \mathbb{E}R)\right)}\, \tfrac{1}{n}D_n(\sigma)$$

$$\le \frac{\rho_\sigma(\frac{1}{n}R_n, \mathbb{E}R)}{\frac{T}{\mu}\bar{F}(\sigma)\left(\frac{T}{\mu}\bar{F}(\sigma) - \rho_\sigma(\frac{1}{n}R_n, \mathbb{E}R)\right)}\, (\mathbb{E}D(\sigma)+\rho_\sigma(\mathbb{E}D,\tfrac{1}{n}D_n)).$$

Inequality (11) follows. □

(A8) Lemma 4.5 Let $h_i:\mathbb{R}\times\Omega\to\mathbb{R}$ for i=1,2,... and let σ be a positive real. Suppose that either

(a) for every fixed ω, $h_i(\cdot,\omega)$ is a right-continuous, nonnegative, increasing function and, for fixed $t\in[0,\sigma]$, $(h_i(t,\cdot)$: i=1,...) and $(h_i(t^-,\cdot)$: i=1,...) are independent sequences of nonnegative, identically distributed r.v.s of finite mean; or

(b) for every fixed ω, $h_i(\cdot,\omega)$ is a left-continuous, nonnegative, decreasing function and, for fixed $t\in[0,\sigma]$, $(h_i(t,\cdot)$: i=1,...) and $(h_i(t^+,\cdot)$: i=1,...) are independent sequences of nonnegative, identically distributed r.v.s of finite mean,

holds. Let $H(t) = \mathbb{E}h_1(t)$ and

$$H_n(t,\omega) = \frac{1}{n}\sum_{k=1}^{n} h_k(t,\omega)\,. \qquad (12)$$

Then $\rho_\sigma(H_n, H) \to 0$ a.s.

Proof Consider case (a). For each t, $H_n(t,\cdot)$ is a r.v. and, by right continuity,

$$\Delta_n(\omega) = \rho_\sigma(H_n(\cdot,\omega), H) = \sup_{0\le t\le\sigma} \left|H_n(t,\omega) - H(t)\right|$$

is unchanged if $t\in[0,\sigma)$ is restricted to the rationals. Therefore Δ_n is a r.v. For fixed t, $H_n(t) \to H(t)$ a.s. by the strong law of large numbers. Furthermore, for fixed $t\in(0,\sigma]$, $H_n(t^-) \to H(t^-)$ a.s. since

$$\mathbb{E}\, h_1(t^-) = \lim_{s\uparrow t} \mathbb{E}\, h_1(s) = \lim_{s\uparrow t} H(s) = H(t^-)$$

by the monotone convergence theorem. Therefore, for each t there is at most a set A_t of probability 0 on which $H_n(t)$ does not converge to $H(t)$ or $H_n(t^-)$ does not converge to $H(t^-)$. The objective of the following is to show that this uncountable number of sets of measure 0 can be included in a countable number of sets of measure 0.

For $m\in N^+$ and $k\in\{1,\dots,m\}$ define

$$B_{m,k} = \{ t\in\mathbb{R} : \frac{(k-1)H(\sigma)}{m} \le H(t) < \frac{kH(\sigma)}{m} \}$$

and for $k=m+1$ define

$$B_{m,m+1} = \{ t\in[0,\sigma] : H(t) = H(\sigma) \}.$$

Note that $B_{m,k}$ could be empty for some k. For $B_{m,k}$ non-empty, let $t_{m,k} = \sup B_{m,k}$ and $s_{m,k} = \inf B_{m,k}$. Note that, if nonempty, the set $B_{m,k}$ is an interval of the form $[s_{m,k}, t_{m,k})$, except that $B_{m,m+1} = [s_{m,m+1}, \sigma]$.

Let

$$\Delta_{m,n}(\omega) = \max_{k:B_{m,k}\neq\phi} \{ \left|H_n(s_{m,k},\omega) - H(s_{m,k})\right| , \left|H_n(t_{m,k}^-,\omega) - H(t_{m,k}^-)\right| \}.$$

Then, for $k \leq m$ and $t \in B_{m,k}$

$$H_n(t,\omega) \leq H_n(t^-_{m,k}, \omega) \leq H(t^-_{m,k}) + \Delta_{m,n}(\omega) \leq H(t) + \frac{1}{m} + \Delta_{m,n}(\omega) \tag{13}$$

where the first inequality follows because H_n is increasing, the second follows from the definition of $\Delta_{m,n}$ and the third because $| H(x) - H(y) | < \frac{1}{m}$ when $x,y \in B_{m,k}$. Similarly,

$$H_n(t,\omega) \geq H_n(s_{m,k}, \omega) \geq H(s_{m,k}) - \Delta_{m,n}(\omega) \geq H(t) - \frac{1}{m} - \Delta_{m,n}(\omega). \tag{14}$$

Furthermore, let

$$\Delta_{*,n}(\omega) = \max \{ \left|H_n(s_{m,m+1}, \omega) - H(s_{m,m+1})\right|, \left|H_n(\sigma, \omega) - H(\sigma)\right| \}.$$

Then, for $t \in B_{m,m+1}$

$$H_n(t,\omega) \leq H_n(\sigma, \omega) \leq H(\sigma) + \Delta_{*,n}(\omega) \leq H(t) + \Delta_{*,n}(\omega) \tag{15}$$

Similarly,

$$H_n(t,\omega) \geq H_n(s_{m,k}, \omega) \geq H(s_{m,k}) - \Delta_{*,n}(\omega) \geq H(t) - \Delta_{*,n}(\omega). \tag{16}$$

The expressions (13), (14), (15) and (16) imply that

$$\Delta_n(\omega) \leq \frac{1}{m} + \Delta_{m,n}(\omega) + \Delta_{*,n}(\omega).$$

Let $A_{s_{m,k}}$ be the set of probability 0 on which $H_n(s_{m,k}, \cdot)$ does not converge to $H(s_{m,k})$, and define $A_{t^-_{m,k}}$ and A_σ analogously. If ω lies outside $A = A_\sigma \cup \bigcup_k (A_{s_{m,k}} \cup A_{t^-_{m,k}})$, then $\lim_n \Delta_{m,n}(\omega) = 0$ and $\lim_n \Delta_{*,n}(\omega) = 0$. Hence $\lim_n \Delta_n(\omega) = 0$. But A has probability 0, so the lemma follows.

The result for case (b) can be proven analogously. □

References

[A1] O.O.Aalen and S.Johansen, "An empirical transition matrix for nonhomogeneous Markov chains based on censored observations", *Scand. J. Statist.*, pp.141-150, 1978.

[B1] P.Billingsley, *Probability and Measure*, 2nd ed., New York, NY: Wiley, 1986.

[F1] W.Feller, *An Introduction to Probability Theory and Its Applications, Vol. II, 2nd ed.*, New York, NY: Wiley, 1971.

[F2] T.R.Fleming and D.P.Harrington, *Counting Process Approach to Survival Analysis*, unpublished manuscript, University of Washington and Harvard University, 1987.

[G1] R.D.Gill, *Censoring and Stochastic Integrals*, Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam, 1980.

[G2] R.D.Gill, "Testing with replacement and the product limit estimator", *Ann. Statist.*, vol.9, pp.853-860, 1981.

[G3] R.D.Gill and S.Johansen, "Product-Integrals and Counting Processes", Report MS-R8707, Centre for Mathematics and Computer Science, Amsterdam, 1987.

[H1] E.Hewitt and K.Stromberg, *Real and Abstract Analysis*, New York, NY: Springer, 1965.

[K1] E.L.Kaplan and P.Meier, "Nonparametric estimation from incomplete observations", *J. Am. Statist. Assoc.*, vol.53, pp.457-481, 1958.

[K2] J.Kiefer, "On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm", *Pacific J. Math.*, vol.11, pp.649-660, 1961.

[L1] M.Loève, *Probability Theory, 4th ed.*, New York, NY: Springer-Verlag, 1977.

[N1] W.Nelson, "Hazard plotting for incomplete failure data", *J. Qual. Technol.*, vol.1, pp.27-52, 1969.

[N2] W.Nelson, "Theory and application of hazard plotting for censored failure data", *Technometrics*, vol.14, pp.945-965, 1972.

[S1] G.R.Shorack and J.A.Wellner, *Empirical Processes with Applications to Statistics*, New York, NY: Wiley, 1986.

[W1] B.B.Winter and A.Földes, "A product-limit estimator for use with length-biased data", *Can. J. Statist.*, vol.16, pp.337-355, 1988.

[W2] B.B.Winter, "Hazard Functions", technical report in preparation, Lab. Res. Statist. Probab., Ottawa.