Weak Convergence of the Aalen Estimator For a Censored Renewal Process

by

Thomas Sellke
Purdue University

Technical Report #86-51

Department of Statistics
Purdue University

October 1986

WEAK CONVERGENCE OF THE AALEN ESTIMATOR FOR A CENSORED RENEWAL PROCESS

Thomas Sellke
Department of Statistics
Purdue University
West Lafayette, Indiana, U.S.A.

1. INTRODUCTION

Suppose that $X_1, X_2, \ldots$ are the lifetimes of a sequence of machine components. We assume that the X_i's are i.i.d. with unknown continuous c.d.f. F and cumulative hazard function $\Lambda(v) = -\log\{1-F(v)\}$. The components will be put on test one at a time and will be replaced with a new component either upon failure or at some preplanned replacement (or "censoring") time, whichever comes first. The censoring time for the n^{th} component is allowed to depend upon observations of the previous $n-1$ components. Bather (1977) considers this situation under the assumption that replacing a failed component costs c_1 dollars, while the planned replacement of a working component costs c_2 dollars, with $c_1 > c_2 > 0$. It is intuitively plausible that the expected cost of replacement per unit time over a long period should be minimized by a policy of replacing components whenever they reach a certain fixed age a_0. That this is so is shown by Berg (1976). However, the optimal age of replacement a_0 clearly depends upon the unknown F as well as on c_1 and c_2. Bather's (1977) goal is to minimize long term average cost in the absence of prior knowledge of F. Since it is necessary to have information about the tail of F in order to estimate a_0 consistently, there is at each stage a conflict between the choice of trying to minimize short term costs by replacing the current component at an estimate of a_0 and the choice of not replacing the current component until it fails in order to gain information about the tail of F, thereby making possible better estimation of a_0 in the future. Bather's procedure specifies a sequence of numbers $1 = p_1 \geq p_2 \geq p_3 \ldots$ such that $p_n \downarrow 0$ and $\Sigma p_n = \infty$. The n^{th} component is allowed to continue on until failure with probability p_n, and is otherwise censored at a current estimate of a_0. Since infinitely many components will not be subject to replacement until they fail, a crude policy is to consistently estimate F based only on observations from these components and to use this estimate of F to estimate a_0. This approach discards all of the information available from the majority of components, which are either censored or potentially censored. The sequential replacement policy actually proposed by Bather (1977) uses data from all previous components to estimate a_0, but for technical reasons he is forced to modify the past data by imposing a "fake" censoring scheme which does throw out some of the available information. He specifies a fixed sequence $b_0 > b_1 > \ldots$, $b_n \downarrow 0$, and if the actual censoring age for the n^{th} component was a_n, then he pretends that the n^{th} component was censored instead at age b_j, where $b_j \leq a_n < b_{j+1}$.

It would be more natural to use the Kaplan-Meier estimator (or the asymptotically equivalent Aalen estimator) based on all past data to estimate F. However, the peculiar nature of the censoring here prevents the usual proofs of consistency and weak convergence from applying. When censoring times are independent of component lifetimes, there is an obvious martingale structure which makes possible straightforward proofs of weak convergence to a Gaussian process. (See, for example, Gill (1980).) The fact that censoring here is allowed to depend on previous lifetimes destroys this martingale structure.

This paper will prove a weak convergence result which applies to censored renewal processes of the above form. The proof is based on the fact that an approximation of the Aalen estimator of $\Lambda(v)$ has a certain orthogonal martingale structure when this approximation is viewed as a process in a two-dimensional time scale. The theorem is not the most general possible, but generalizations in various directions (noncontinuous F's, triangular arrays, etc.) seem straightforward.

2. NOTATION AND THE THEOREM

In what follows, calendar time will be denoted by t and u, and age time will be denoted by s and v. The assumption that components are necessarily put on test only one at a time will be dropped, and it will be permitted for components to be put *back* on test after having been removed from use. Repair of failed components will also be allowed, with the assumption that the distribution of remaining time on test until the next failure is the same as if no previous failure had occurred, i.e., past failures of repaired components do not affect future performance.

The situation of interest here is most easily modelled using the machinery of counting processes. The reference is Gill (1980). Let $Y_i(t)$ be the indicator of the event that the i^{th} component is on test at time t. Define

$$Z_i(t) = \int_0^t Y_i(u)\, du, \tag{1}$$

so that $Z_i(t)$ is the length of time that component i has been on test by calendar time t. Let $N_i(t)$ be the number of failures that component i has suffered at or before time t. Let $\mathcal{F}_t$, $t \geq 0$, be a complete, right-continuous filtration, where $\mathcal{F}_t$ represents knowledge at time t. For each i, $Y_i(t)$ is assumed to be left-continuous and $\mathcal{F}_t$-predictable. Thus, for each ω, $Y_i(t)$ is the indicator function of a union of disjoint intervals which are open on the left and closed on the right. The counting processes $N_i(t)$ are assumed to be $\mathcal{F}_t$-adapted with right-continuous paths. They are zero at time zero, integer valued, nondecreasing, and have jumps of size +1. The process $N_i(t)$ are connected with the continuous cumulative hazard function Λ by the assumption that

$$M_i(t) = N_i(t) - \Lambda\{Z_i(t)\} \tag{2}$$

is an $\mathcal{F}_t$-martingale. Independence of the different components is reflected in the assumption that, with probability one, no two counting processes jump simultaneously. Let $A_i(t) = \Lambda\{Z_i(t)\}$ = amount of accumulated hazard to which component i has been subjected by time t. It follows from Theorem 2.3.1, of Gill (1980) that

$$\langle M_i, M_j \rangle(t) = A_i(t) 1\{i = j\}. \tag{3}$$

Here, $\langle M_i, M_j \rangle$ is the "predictable covariation process" of M_i and M_j, so that the product $M_i M_j$ is a martingale for $i \neq j$, i.e., M_i and M_j are orthogonal martingales. The "predictable variance process" of M_i is

$$\langle M_i \rangle = \langle M_i, M_i \rangle = A_i, \tag{4}$$

so that $M_i^2 - A_i$ is a martingale.

The Aalen estimator of Λ based on the data available at time t is a function of the amount of time that each component has been in use and of the ages at which failures have occurred. Define

$$N_i(t,s) = \int_0^s 1\{Z_i(u) \leq s\}\, N_i(du) \tag{5}$$

to be the number of failures suffered by component i before calendar time t and age s. Set $N(t,s) = \sum_i N_i(t,s)$. Define

$$R(t,s) = \sum_i 1\{Z_i(t) \geq s\} \tag{6}$$

to be the number of components which have been on test for at least s units of time by calendar time t. Thus, $R(t,s)$ is the size of the "risk set" at age s and time t. The Aalen estimator of $\Lambda(s)$ based on data available at time t is defined by

$$\hat{\Lambda}(t,s) = \int_0^s \frac{N(t,dv)}{R(t,v)}. \tag{7}$$

(Interpret $\frac{0}{0}$ as 0.)

Note that $N_i(t,s)$ is, as a function of t for fixed s, simply a stopped version of $N_i(t)$ and is therefore itself a counting process. Define

$$A_i(t,s) = \Lambda\{Z_i(t) \ \wedge \ s\} \quad \text{and} \quad A(t,s) = \sum_i A_i(t,s),$$

so that $A_i(t,s)$ is the amount of accumulated hazard to which component i has been exposed before age s and time t. Then

$$M_i(t,s) = N_i(t,s) - A_i(t,s) \tag{8}$$

is, as a function of t for fixed s, a stopped version of $M_i(t)$ and therefore also an $\mathcal{F}_t$-martingale. It follows from (3) and (8) that for $0 \leq s_1 \leq s_2$,

$$\langle M_i(\cdot, s_1), M_j(\cdot, s_2) \rangle (t) = A_i(t, s_1) 1\{i = j\}. \tag{9}$$

If we define

$$M(t,s) = \sum_i M_i(t,s) = N(t,s) - A(t,s),$$

then it follows from (9) that

$$\langle M(\cdot, s_1), M(\cdot, s_2)\rangle(t) = A(t, s_1 \wedge s_2). \tag{10}$$

and that, for $0 \leq s_1 \leq s_2 \leq s_3$

$$\langle M(\cdot, s_1), M(\cdot, s_3) - M(\cdot, s_2)\rangle(t) = 0. \tag{11}$$

Thus, $M(t, s_1)$ and $M(t, s_3) - M(t, s_2)$ are orthogonal $\mathcal{F}_t$-martingales. This orthogonality structure of $M(t, s)$ will play only a minor direct role in the next section. However, a related two-dimensional process $L(t, s)$ will turn out to have the same sort of orthogonal martingale structure as $M(t, s)$, and this will be central to the proof of the following theorem.

THEOREM. *Let r: $[0, a] \to (0, \infty)$ be a fixed, left-continuous, nonincreasing function. Suppose that*

$$\frac{R(t, s)}{t} \xrightarrow{P} r(s) \tag{12}$$

as $t \to \infty$, for continuity points s of r, and for $s = 0$ and $s = a$. Assume also that the cumulative hazard function Λ is continuous on $[0, a]$ and that $\Lambda(a) < \infty$. Then as $t \to \infty$,

$$t^{\frac{1}{2}}\{\hat{\Lambda}(t, \cdot) - \Lambda(\cdot)\} \xrightarrow{w} B\left(\int_0^{(\cdot)} \frac{\Lambda(dv)}{r(v)}\right), \tag{13}$$

where B is standard Brownian motion and the weak convergence is in the space of functions $D[0, a]$ with the Skorokhod topology.

3. PROOF OF THE THEOREM

I will assume for convenience that for each t under consideration, $R(t, 0)$ (and therefore $R(t, s)$, $0 \leq s \leq a$) is bounded above by $2r(0)t$, a.s. Condition (12) does not imply this, but the general case is easily reduced to this one by a trivial localization argument.

Note that

$$A(t, dv) = R(t, v)\Lambda(dv). \tag{14}$$

If $R(t, s) > 0$, it follows from (14) that

$$\begin{aligned} t^{\frac{1}{2}}\{\hat{\Lambda}(t, s) - \Lambda(s)\} &= t^{\frac{1}{2}} \int_0^s \frac{N(t, dv) - A(t, dv)}{R(t, v)} \\ &= t^{-\frac{1}{2}} \int_0^s \{t/R(t, v)\}\, M(t, dv). \end{aligned} \tag{15}$$

Define

$$L(t, s) = \int_0^s \{r(v)\}^{-1}\, M(t, dv). \tag{16}$$

Assumption (12) in the Theorem makes it plausible that (15) should, with high probability, be approximately equal to $t^{-\frac{1}{2}}L(t,s)$ for all $s \in [0,a]$ when t is large. The proof that this is so is postponed until the end of this section. Modulo this asymptotic equivalence, it suffices to show that

$$t^{-\frac{1}{2}} L(t,\cdot) \xrightarrow{w} B\{\int_0^{(\cdot)} r(v)^{-1} \Lambda(dv)\} \tag{17}$$

as $t \to \infty$. The proof of convergence of finite dimensional distributions in (17) depends upon the orthogonal martingale structure alluded to in the previous section.

LEMMA 1. For each fixed $s \in [0,a]$, $L(t,s)$ is a square integrable $\mathcal{F}_t$-martingale, and for $0 \le s_1 \le s_2 \le a$

$$\langle L(\cdot,s_1), L(\cdot,s_2)\rangle(t) = \int_0^{s_1} \frac{R(t,v)}{r^2(v)} \Lambda(dv).$$

Proof of Lemma 1.

$$\begin{aligned} L(t,s) &= \sum_i \int_0^s r(v)^{-1} M_i(t,dv) \\ &= \sum_i \int_0^t r\{Z_i(u)\}^{-1} M_i(du,s) \end{aligned}$$

(This last equality is "obvious", but its proof is a bit tedious and therefore omitted.) Since the integrands $r\{Z_i(u)\}^{-1}$ are bounded and predictable, it follows from Gill (1980), page 10, that $L(t,s)$ is a square-integrable martingale in t, and that for $0 \le s_1 \le s_2 \le a$

$$\begin{aligned} \langle L(\cdot,s_1), L(\cdot,s_2)\rangle(t) &= \sum_i \int_0^t r\{Z_i(u)\}^{-2} A_i(du,s_1) \\ &= \sum_i \int_0^{s_1} r(v)^{-2} A_i(t,dv) \quad (\text{"obvious"}) \\ &= \int_0^{s_1} r(v)^{-2} A(t,dv). \end{aligned}$$

Now use (14). □

The convergence of finite dimensional distributions in (17) is an immediate consequence of Lemma 2.

LEMMA 2. Fix $0 = s_0 < s_1 < \ldots < s_k = a$, and set $c_i = \int_{s_{i-1}}^{s_i} r(v)^{-1}\Lambda(dv)$. Let $B_1, \ldots, B_k$ be independent standard Brownian motions, and let

$W^\infty(u) = [c_1B_1(u), \ldots, c_kB_k(u)]$ be a k-dimensional diffusion. For each t, let $W^t(u)$ be the k-dimensional stochastic process with time index u and with i^{th} coordinate function

$$W_i^t(u) = t^{-\frac{1}{2}}\{L(tu, s_i) - L(tu, s_{i-1})\}. \tag{18}$$

Then as $t \to \infty$, $W^t \xrightarrow{w} W^\infty$ in the Skorokhod topology on $(D[0,\infty))^k$.

Proof of Lemma 2. By Lemma 1,

$$\langle W_i^t, W_j^t\rangle(u) = 1\{i=j\} \int_{s_{i-1}}^{s_i} \frac{R(tu,v)}{tr^2(v)}\, \Lambda(dv), \tag{19}$$

so that by (12),

$$\langle W_i^t, W_j^t\rangle(u) \xrightarrow{P} 1\{i=j\}c_iu \tag{20}$$

as $t \to \infty$ for each i, j, u. Also, the jumps of W_i^t are bounded in size by $t^{-\frac{1}{2}}r(a)^{-1}$. The claim follows from the martingale central limit theorem 2.4.1. of Gill (1980). $\square$

Tightness in (17) will be proven using the tightness criterion (12.51) of Billingsley (1968). This tightness criterion only applies to continuous random elements in $D[0,a]$, however, so it will be necessary to define a continuous (in s) modification $\tilde{L}(u,s)$ of $L(u,s)$. The criterion (12.51) will be verified for $t^{-\frac{1}{2}}\tilde{L}(t,s)$, and then we'll see that

$$\sup_{0\le s\le a} t^{-\frac{1}{2}}|\tilde{L}(t,s) - L(t,s)| \xrightarrow{P} 0.$$

In order to conform with Billingsley's notation, define

$$\tilde{F}(s) = \int_0^s r(v)^{-1}\, \Lambda(dv).$$

Now fix t, and partition $[0,a]$ into $k = k_t = [t^{\frac{2}{3}}]$ pieces with partition points $0 = v_0 < v_1 < \ldots < v_k = a$ so that

$$2^{-\frac{1}{2}}\tilde{F}(a)t^{-\frac{2}{3}} \le \tilde{F}(v_{i+1}) - \tilde{F}(v_i) \le 2\tilde{F}(a)t^{-\frac{2}{3}}. \tag{21}$$

For $v_i \le s < v_{i+1}$, define

$$\tilde{L}(u,s) = L(u,v_i) + \frac{\tilde{F}(s) - \tilde{F}(v_i)}{\tilde{F}(v_{i+1}) - \tilde{F}(v_i)}\{L(u,v_{i+1}) - L(u,v_i)\}. \tag{22}$$

LEMMA 3.

$$E[\{t^{-\frac{1}{2}}\ \tilde{L}(t,s_2) - t^{-\frac{1}{2}}\ \tilde{L}(t,s_1)\}^4] \le C\{\tilde{F}(s_2) - \tilde{F}(s_1)\}^2 \tag{23}$$

for $0 \leq s_1 \leq s_2 \leq a$, C a suitable constant not depending on t, and t sufficiently large.

(23) is a version of the tightness criterion (12.51) in Billingsley (1968). The tool used to prove Lemma 3 will be the Burkholder-Davis-Gundy inequality for continuous-time martingales quoted in the Appendix.

Proof of Lemma 3. It follows from Lemma 1 that if $v_i < s_1 \leq v_{i+1} \leq v_j \leq s_2 < v_{j+1}$, then

$$\langle \tilde{L}(\cdot, s_2) - \tilde{L}(\cdot, s_1) \rangle (t) = \int_{v_{i+1}}^{v_j} \frac{R(t,v)}{r^2(v)} \Lambda(dv) \tag{24}$$
$$+ \left\{ \frac{\tilde{F}(s_2) - \tilde{F}_{(v_j)}}{\tilde{F}(v_{j+1}) - \tilde{F}(v_j)} \right\}^2 \int_{v_j}^{v_{j+1}} \frac{R(t,v)}{r^2(v)} \Lambda(dv)$$
$$+ \left\{ \frac{\tilde{F}(v_{i+1}) - \tilde{F}(s_1)}{\tilde{F}(v_{i+1}) - \tilde{F}(v_i)} \right\}^2 \int_{v_i}^{v_{i+1}} \frac{R(t,v)}{r^2(v)} \Lambda(dv).$$

The assumption that $R(t,0) \leq 2r(0)t$ implies

$$\int_x^y \frac{R(t,v)}{r^2(v)} \Lambda(dv) \leq 2t \frac{r(0)}{r(a)} \{\tilde{F}(y) - \tilde{F}(x)\}.$$

Applying this to (24) yields

$$\langle \tilde{L}(\cdot, s_2) - \tilde{L}(\cdot, s_1) \rangle (t) \leq 2t \frac{r(0)}{r(a)} \{\tilde{F}(s_2) - \tilde{F}(s_1)\}. \tag{25}$$

A very similar calculation shows that (25) also holds if $v_i \leq s_1 \leq s_2 \leq v_{i+1}$.

In order to apply the Burkholder-Davis-Gundy inequality, we need to get a bound on the jumps of the martingale (in u)

$$t^{-\frac{1}{2}} \{\tilde{L}(u, s_2) - \tilde{L}(u, s_1)\}. \tag{26}$$

CLAIM: The jumps of (26) are bounded in size by $\{\tilde{F}(s_2) - \tilde{F}(s_1)\}^{\frac{1}{2}}$ for t sufficiently large.

Proof of the Claim: The jumps of (26) are always bounded by $t^{-\frac{1}{2}} r(a)^{-1}$. This is more than enough if $\tilde{F}(s_2) - \tilde{F}(s_1) > t^{-\frac{3}{4}}$. If $\tilde{F}(s_2) - \tilde{F}(s_1) \leq t^{-\frac{3}{4}}$ and $v_i \leq s_1 < s_2 < v_{i+1}$, then the interpolation used in the definition of $\tilde{L}(u, s)$ reduces the bound to

$$t^{-\frac{1}{2}} r(a)^{-1} \frac{\tilde{F}(s_2) - \tilde{F}(s_1)}{\tilde{F}(v_2) - \tilde{F}(v_1)} < 2t^{\frac{1}{6}} r(a)^{-1} \{\tilde{F}(s_2) - \tilde{F}(s_1)\}. \tag{27}$$

(The inequality follows from (21).) The right-hand side of (27) is less than $\{\tilde{F}(s_2)-\tilde{F}(s_1)\}^{\frac{1}{2}}$ if $\tilde{F}(s_2)-\tilde{F}(s_1) \leq t^{-\frac{3}{4}}$ and t is sufficiently large. If $\tilde{F}(s_2)-\tilde{F}(s_1) \leq t^{-\frac{3}{4}}$ and $s_1 \leq v_i \leq s_2$, then a similar argument applies. □

By the Burkholder-Davis-Gundy inequality, (25) and the Claim imply that Lemma 3 holds with $C = c_0 \left[\left\{2\frac{r(0)}{r(a)}\right\}^{\frac{1}{2}} + 1\right]^4$ for t large enough. □

LEMMA 4. As $t \to \infty$,

$$\sup_{0 \leq s \leq a} t^{-\frac{1}{2}} |\tilde{L}(t,s) - L(t,s)| \xrightarrow{P} 0.$$

Proof of Lemma 4: For any i,

$$\begin{aligned}\langle L(\cdot, v_{i+1}) - L(\cdot, v_i)\rangle(t) &= \int_{v_i}^{v_{i+1}} \frac{R(t,v)}{r^2(v)}\, \Lambda(dv) \\ &\leq 2 + \frac{r(0)}{r(a)}\, \{\tilde{F}(v_{i+1}) - \tilde{F}(v_i)\} \\ &\leq 4\frac{r(0)}{r(a)} t^{\frac{1}{3}}.\end{aligned}$$

By Freedman's (1975) tail inequality for bounded-jump martingales (quoted in the Appendix),

$$P\{|L(t,v_{i+1}) - L(t,v_i)| \geq t^{\frac{2}{5}}\} \leq 2\exp\left[\frac{-t^{\frac{4}{5}}}{2\{r(a)^{-1}t^{\frac{2}{5}} + 4r(0)t^{\frac{1}{3}}/r(a)\}}\right].$$

(Take $a = t^{\frac{2}{5}}$, $b = 4r(0)t^{\frac{1}{3}}/r(a)$, and $K = r(a)^{-1}$ in Freedman's theorem.) Since $t^{\frac{2}{3}}$ times the bound on the right-hand side goes to 0 as $t \to \infty$, we have

$$P\{\sup_{1 \leq i \leq [t^{\frac{2}{3}}]} |L(t,v_{i+1}) - L(t,v_i)| \leq t^{\frac{2}{5}}\} \longrightarrow 1 \tag{28}$$

as $t \to \infty$. Since for $v_i \leq s \leq v_{i+1}$,

$$\begin{aligned}L(t,s) - L(t,v_i) &= \int_{v_i}^{s} r(v)^{-1} M(t,dv) \\ &= \int_{v_i}^{s} r(v)^{-1} N(t,dv) - \int_{v_i}^{s} r(v)^{-1} R(t,v)\Lambda(dv),\end{aligned}$$

$L(t,s) - L(t,v_i)$ is bounded below by the negative of

$$\begin{aligned}\int_{v_i}^{v_{i+1}} r(v)^{-1} R(t,v)\Lambda(dv) &< 2\; t\; r(0)\{\tilde{F}(v_{i+1}) - F(v_i)\} \\ &\leq 4\; r(0)\; t^{\frac{1}{3}}, \quad \text{by (21).}\end{aligned}$$

If the event in (28) holds, then

$$\int_{v_i}^{v_{i+1}} r(v)^{-1}N(t,dv) = L(t,v_{i+1}) - L(t,v_i) + \int_{v_i}^{v_{i+1}} r(v)^{-1}R(t,v)\Lambda(dv)$$
$$< t^{\frac{2}{5}} + 4r(0)t^{\frac{1}{3}}. \tag{29}$$

Since the left-hand side of (29) is an upper bound for $L(t,s) - L(t,v_i)$, $v_i \le s \le v_{i+1}$, this implies that

$$\sup_{0\le s\le a} t^{-\frac{1}{2}}|\tilde{L}(s,t) - L(s,t)| \le 2t^{-\frac{1}{2}}\{t^{\frac{2}{5}} + 4\, r(0)t^{\frac{1}{3}}\}$$

whenever the event in (28) holds. This proves Lemma 4. □

Lemma 4 and Lemma 2 imply the convergence of finite dimensional distributions in

$$t^{-\frac{1}{2}}\tilde{L}(t,\cdot) \xrightarrow{w} B\{\int_0^{(\cdot)} r(v)^{-1}\Lambda(dv)\}, \tag{30}$$

and Lemma 3 implies tightness. Together, (30) and Lemma 4 imply (17).

It remains to prove

LEMMA 5. As $t \to \infty$,

$$\sup_{0\le s\le a} |t^{-\frac{1}{2}}L(t,s) - t^{\frac{1}{2}}\{\hat{\Lambda}(t,s) - \Lambda(s)\}| \xrightarrow{P} 0.$$

Proof of Lemma 5. It follows from (16) and (17) that

$$t^{-\frac{1}{2}}M(t,\cdot) \xrightarrow{w} B\{\int_0^{(\cdot)} r(v)\Lambda(dv)\} \tag{31}$$

on $D[0,a]$. Alternatively, this can be proven by repeating the previous arguments of this section, with $M(t,s)$ in place of $L(t,s)$ and with formula (10) taking the place of Lemma 1.

Define, for $0 \le s \le a$,

$$G(s) = r(s)^{-1}, \quad G_t(s) = t/R(t,s), \quad \text{and} \quad \tilde{B}_t(s) = t^{-\frac{1}{2}}M(t,s).$$

If $R(t,a) > 0$, it follows from (15) and (16) that

$$t^{-\frac{1}{2}}L(t,s) - t^{\frac{1}{2}}\{\hat{\Lambda}(t,s) - \Lambda(s)\} = \int_0^s \{r(v)^{-1} - t/R(t,v)\}t^{-\frac{1}{2}}M(t,dv)$$
$$= \int_0^s \{G(v) - G_t(v)\}\tilde{B}_t(dv). \tag{33}$$

The salient properties of G, G_t, and $\tilde{B}_t$ needed to prove Lemma 5 are:

(a) G is an increasing, left-continuous function with $0 < G(0) \leq G(a) = r(a)^{-1}$.

(b) G_t is a non-negative, left-continuous random function for which

$$P\{G_t(a) < 2r(a)^{-1}\} \longrightarrow 1 \text{ as } t \to \infty, \tag{34}$$

and

$$G_t(s) \xrightarrow{P} G(s) \text{ as } t \to \infty \tag{35}$$

for continuity points s of G, and for $s = 0$, $S = a$.

(c) For each t, $\tilde{B}_t(\cdot)$ is a right-continuous process of almost surely bounded variation. Let

$$w_t(\delta) = \sup_{|s_2 - s_1| < \delta} |\tilde{B}_t(s_2) - \tilde{B}_t(s_1)|$$

be the modulus of continuity of $\tilde{B}_t$. Then by (31), for each $\varepsilon > 0$, there exists t_0 and $\delta > 0$ such that $t \geq t_0$ implies

$$P\{w_t(\delta) > \varepsilon\} < \varepsilon. \tag{36}$$

Let $\varepsilon > 0$. Choose t_0 and δ so that (36) holds. Then choose continuity points $s_1, \ldots, s_{k-1}$ of G so that

$$0 = s_0 < s_1 < \ldots < s_k = a \qquad \text{and} \qquad s_{i+1} - s_i < \delta.$$

By (35) and (34), $t \geq t_0$ will imply

$$P\{\sup_{0 \leq i \leq k} |G_t(s_i) - G(s_i)| \geq 1/k\} < \varepsilon \tag{37}$$

and

$$P\{G_t(a) < 2r(a)^{-1}\} < \varepsilon \tag{38}$$

if t_0 is made sufficiently larger. Note that the event in (38) implies $R(t,a) > 0$.

Assume that the events in (36), (37), and (38) hold. (This is obviously true with probability $> 1 - 3\varepsilon$ for $t \geq t_0$). A bound for the absolute value of (33) will now be found. Fix $s \in [0, a]$, and let $s_j \leq s < s_{j+1}$. Rewrite (33) as

$$\sum_{i=0}^{j-1} \int_{s_i}^{s_{i+1}} \{G(v) - G_t(v)\} d\{\tilde{B}_t(v) - \tilde{B}_t(s_i)\} + \int_{s_j}^{s} \{G(v) - G_t(v)\} d\{\tilde{B}_t(v) - \tilde{B}_t(s_j)\}. \tag{39}$$

The events in (36) and (37) imply that (40) is bounded in absolute value by

$$\sum_{i=0}^{j-1} \varepsilon/k + 3\{G(a) + G_t(a)\}\varepsilon$$

which by (38) is bounded by $\varepsilon + 9r(a)^{-1}\varepsilon$. Since this bound is valid for all s with probability $> 1 - 3\varepsilon$ whenever $t \geq t_0$, Lemma 5 follows. $\square$

APPENDIX

The following Burkholder-Davis-Gundy inequality for continuous-time martingales can be found in Lenglart, Lepingle, and Pratelli (1980).

Burkholder-Davis-Gundy Inequality

Let $t \geq 0$, be a continuous-time, square integrable martingale satisfying $M(0) = 0$ and with predictable variance process $\langle M\rangle(t)$. Suppose the jumps of M are bounded in size by a real number $K \geq 0$, and let $M^*(t) = \sup_{u \leq t} |M(u)|$. Then there exists a universal constant c_0 such that

$$E[\{M^*(\infty)\}^4] \leq c_0 \; E[\{\langle M\rangle(\infty)^{\frac{1}{2}} + K\}^4].$$

The following result is a continuous-time version of Freedman's (1975) tail inequality for martingales (Proposition 2.1). Although Freedman only considers discrete-time martingales, his proof is easily adapted to the continuous-time case.

Tail Inequality for Bounded-jump Martingales

Suppose $M(t)$, $t \geq 0$, is a locally square integrable martingale with jumps bounded in size by K. Then for all positive real numbers a and b,

$$P\{|M(t)| \geq a \quad \text{and} \quad \langle M\rangle(t) \leq b \quad \text{for some} \quad t \geq 0\} \leq 2\exp\left[-\frac{a^2}{2(Ka+b)}\right].$$

ACKNOWLEDGMENT

The techniques of this paper were originally developed to study the evolution of the partial likelihood function for the proportional hazards model in staggered-entry clinical trials. I am very grateful to Richard Gill for pointing out to me that these methods apply here.

REFERENCES

[1] Bather, J. A. (1977). On the sequential construction of an optimal age replacement policy. *Proc. 41st Session ISI*, New Delhi, 253–266.

[2] Berg, M. (1976). A proof of optimality for age replacement policies. *J. Appl. Prob.* **13**, 751–759.

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

[4] Freedman, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3**, 100–118.

[5] Gill, R. D. (1975). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

[6] Lenglart, E., Lepingle, D., Pratelli, M. (1980). Présentation unifiée de certaines inéqualités de la théorie des martingales. *Seminaire de Probabilités XIV 1978/79*, Springer Verlag Lecture Notes in Mathematics, No. 784, 26–48.