

CONSISTENCY IN A PROPORTIONAL HAZARDS MODEL INCORPORATING A RANDOM EFFECT¹

BY S. A. MURPHY

Pennsylvania State University

The frailty model is a generalization of Cox's proportional hazards model which includes a random effect. Nielsen, Gill, Andersen and Sørensen proposed an EM algorithm to estimate the cumulative baseline hazard and the variance of the random effect. Here existence and consistency of the estimators are proved. An example using truncated and censored data is considered.

0. Introduction. The frailty model, which is a multivariate generalization of the proportional hazards model, allows for heterogeneity of hazard by incorporating a random effect. Estimation in the frailty model has received much attention; for example, see Clayton and Cuzick (1985), Self and Prentice (1986) and Nielsen, Gill, Andersen and Sørensen (1992). The parameters include regression coefficients, parameters describing the distribution of the random effect; and the cumulative baseline hazard. In the usual proportional hazards model, the maximum likelihood estimator (MLE) of the regression coefficients and the nonparametric MLE (the Nelson–Aalen estimator) of the cumulative baseline hazard have been shown to be consistent and asymptotically efficient [Greenwood and Wefelmeyer (1990)]. This paper addresses the problem of consistency for the one-sample frailty model. The method of proof should generalize to the regression setting. To form MLE's, the counting process approach outlined in Nielsen, Gill, Andersen and Sørensen (1992) and Andersen, Borgan, Gill and Keiding (1993) is used. In Section 1 a brief introduction is given. The consistency theorem and proof are in Section 2. The Appendix contains technical details.

1. Statistical model. As mentioned above, much of the following description is a review of the counting process approach presented in Nielsen, Gill, Andersen and Sørensen (1992). Using their notation, the frailty or random effect is defined on the probability space (Ω', G', P'_θ) and is denoted by $\mathbf{Z} = (Z_1, \dots, Z_n)$. Let $(\Omega'', \{G''_{t \in [0, \tau]}\}, P''_{\alpha\mathbf{z}})$ be a filtered probability space for each $\mathbf{Z} = \mathbf{z}$, so that under $P''_{\alpha\mathbf{z}}$ (i.e., conditionally on $\mathbf{Z} = \mathbf{z}$) the multivariate counting process $\mathbf{N} = (N_i; i = 1, \dots, n)$ has intensity process λ given by

$$\lambda_i(u) = z_i Y_i(u) \alpha(u).$$

Received December 1991; revised February 1993.

¹This material is based in part upon work supported by the North Atlantic Treaty Organization under a grant awarded in 1990 and conducted at the Mathematical Institute of the University of Utrecht, The Netherlands.

AMS 1991 subject classifications. 62M09, 62G05.

Key words and phrases. Counting processes, survival analysis, frailty model, random effects.

The N_i can be thought of as the aggregate of the counting processes for group i , so that each N_i can have more than one jump. The members of the i th group share the same frailty, Z_i . The Z_i are assumed to be independent random variables, each distributed according to a gamma distribution with mean 1 and variance θ . Of course this means that if $\theta = 0$, then the random effect Z_i is identically equal to 1 and the random effect does not induce a dependency between members of group i . The Y_i are observable, nonnegative, predictable processes, and α is an unknown baseline hazard rate. Note that this model would have to be reformulated to allow for a discrete hazard. In a discrete model the intensity $Z_i\alpha(u)$ is bounded by 1, which certainly does not allow Z_i to have a gamma distribution.

The goal is to estimate θ and the cumulative baseline hazard $A(t) = \int_0^t \alpha(u) du$ based on observation of (\mathbf{N}, \mathbf{Y}) only and via maximum likelihood estimation. There are two ways to form the likelihood of (\mathbf{N}, \mathbf{Y}) . The first method is to write the likelihood of $(\mathbf{N}, \mathbf{Y}, \mathbf{Z})$ as the density of (\mathbf{N}, \mathbf{Y}) given $\mathbf{Z} = \mathbf{z}$ times the density of \mathbf{Z} and to integrate out over the variable \mathbf{z} . Actually, only a partial conditional likelihood of (\mathbf{N}, \mathbf{Y}) given $\mathbf{Z} = \mathbf{z}$ is specified, and it is assumed that the remaining term in the conditional likelihood does not involve \mathbf{z} [Nielsen, Gill, Andersen and Sørensen (1992) state this in Assumption 2, "Conditional on $\mathbf{Z} = \mathbf{z}$, censoring is noninformative of \mathbf{z} "]. The partial likelihood for (\mathbf{N}, \mathbf{Y}) given $\mathbf{Z} = \mathbf{z}$ is

$$\prod_{i=1}^n \left\{ \prod_t (z_i Y_i(t) \alpha(t))^{\Delta N_i(t)} \exp \left\{ -z_i \int_0^\tau Y_i dA \right\} \right\}.$$

Multiplying by the density of \mathbf{Z} and then integrating over \mathbf{z} yields the partial likelihood

$$(1.1) \quad \prod_{i=1}^n \frac{\prod_t \left((1 + \theta N_i(t-)) Y_i(t) \alpha(t) \right)^{\Delta N_i(t)}}{\left(1 + \theta \int_0^\tau Y_i(t) dA(t) \right)^{1/\theta + N_i(\tau)}}.$$

It is also straightforward to see that the distribution of Z_i given (\mathbf{N}, \mathbf{Y}) is a gamma distribution with mean

$$\frac{1 + \theta N_i(\tau)}{1 + \theta \int_0^\tau Y_i dA}$$

and variance

$$\theta \frac{1 + \theta N_i(\tau)}{\left(1 + \theta \int_0^\tau Y_i dA \right)^2}.$$

A second method of forming the partial likelihood of (\mathbf{N}, \mathbf{Y}) is to use the innovation theorem [Bremaud (1981)]; that is, in order to derive the intensity of \mathbf{N} with respect to the observed history (i.e., the product of the trivial sigma field on Ω' with G''), Z_i is replaced by its conditional mean relative to this history. Therefore the intensity of \mathbf{N} is

$$\lambda_i(u) = \frac{1 + \theta N_i(u-)}{1 + \theta \int_0^{u-} Y_i(s) dA(s)} Y_i(u) \alpha(u).$$

Note that if $\theta = 0$, then the intensity of \mathbf{N} takes the multiplicative form with Z_i

identically equal to 1. The partial likelihood function is given by

$$(1.2) \quad \prod_{i=1}^n \left\{ \prod_u \left(\frac{1 + \theta N_i(u-)}{1 + \theta \int_0^{u-} Y_i(s) dA(s)} Y_i(u) \alpha(u) \right)^{\Delta N_i(u)} \right. \\ \left. \times \exp \left(- \int_0^\tau \frac{1 + \theta N_i(t-)}{1 + \theta \int_0^{t-} Y_i(s) dA(s)} Y_i(t) dA(t) \right) \right\}.$$

Since A is continuous, (1.1) and (1.2) can be shown to be equivalent via integration by parts [see (1.3)]. Both are full likelihoods for (θ, A) if the omitted term does not depend on (θ, A) [Nielsen, Gill, Andersen and Sørensen (1992) call this noninformative censoring for the parameter (θ, A)]. The true values of the parameters [say, (θ_0, A_0)] lie in $[0, \infty) \times \{\text{absolutely continuous cumulative hazards}\}$. However, maximization of the log-likelihood over this parameter space leads to the same difficulties as in estimation of a density function. The problem is that there is no absolutely continuous estimator \hat{A} which will maximize the likelihood. There appear to be two approaches to this problem. If an estimator of the hazard α is desired, then it is necessary to restrict the parameter space further for a finite sample, for example, employ the method of sieves or penalized likelihood estimation. The interest here is in the estimation of the cumulative hazard, and a second approach is used. This second approach extends the parameter space so that the estimator \hat{A} is allowed to be discrete. The parameter space is then $[0, \infty) \times \{\text{cumulative hazards}\}$. This is the type of extension of parameter spaces which allows one to consider the empirical distribution function as a nonparametric maximum likelihood estimator of a continuous distribution function. To allow for a discrete estimator, replace $\alpha(u)$ by $\Delta A(u)$, the jump of A at the point u , in (1.1) and (1.2). Unfortunately (1.1) and (1.2) are no longer equivalent and will not lead to the same estimators. The natural logarithm of (1.1) is given by

$$nL_n(\theta, A) = \sum_{i=1}^n \int_0^\tau \ln(1 + \theta N_i(u-)) dN_i(u) \\ - (\theta^{-1} + N_i(\tau)) \ln \left(1 + \theta \int_0^\tau Y_i(u) dA(u) \right) \\ + \int_0^\tau \ln(Y_i(u) \Delta A(u)) dN_i(u).$$

If $\theta = 0$, the second term above is defined by its right-hand limit at 0, that is, $\int_0^\tau Y_i dA + N_i(\tau)$. The natural logarithm of (1.2) is

$$nL'_n(\theta, A) = \sum_{i=1}^n \int_0^\tau \ln \left(\frac{1 + \theta N_i(u-)}{1 + \theta \int_0^{u-} Y_i(s) dA(s)} \right) dN_i(u) \\ - \int_0^\tau \frac{1 + \theta N_i(u-)}{1 + \theta \int_0^{u-} Y_i(s) dA(s)} Y_i(u) dA(u) \\ + \int_0^\tau \ln(Y_i(u) \Delta A(u)) dN_i(u).$$

In fact,

$$(1.3) \quad nL'_n(\theta, A) = nL_n(\theta, A) + \sum_{i=1}^n \int_0^\tau (\theta^{-1} + N_i(u)) d \left[\ln \left(1 + \theta \int_0^u Y_i dA \right) \right] - \sum_{i=1}^n \int_0^\tau \frac{1 + \theta N_i(u-)}{1 + \theta \int_0^{u-} Y_i dA} Y_i(u) dA(u).$$

Note that the last two terms would cancel if A were continuous. Nielsen, Gill, Andersen and Sørensen (1992) maximize L_n via the EM algorithm. In this paper, only estimators derived by maximizing L_n are considered.

Taking the derivative of L_n with respect to the jump sizes of A and setting the derivative equal to zero yields the following equation for \hat{A} :

$$\hat{A}(t) = \int_0^t \left(n^{-1} \sum_i \frac{1 + \hat{\theta} N_i(\tau)}{1 + \hat{\theta} \int_0^\tau Y_i(s) d\hat{A}(s)} Y_i(u) \right)^{-1} d\bar{N}(u),$$

where $\bar{N}(u) = n^{-1} \sum_{i=1}^n N_i(u)$. This equation also results from the M step of the EM algorithm; see Nielsen, Gill, Andersen and Sørensen (1992) for details.

2. Existence and consistency. Assume that the (N_i, Y_i) are i.i.d. copies of (N, Y) , where Y is a.s. left-continuous with right-hand limits and takes on nonnegative integer values. Both N and Y are bounded in supremum norm. The counting process N satisfies

$$(2.1) \quad E \int_0^\tau C(u) dN(u) = E \int_0^\tau C(u) \frac{1 + \theta_0 N(u-)}{1 + \theta_0 \int_0^{u-} Y(s) dA_0(s)} Y(u) dA_0(u),$$

for C left-continuous and adapted to the filtration $\sigma\{N(s); Y(s), s \leq t\}; t \in [0, \tau]$. The variance parameter θ_0 lies in a known interval, say, $[0, M]$. The cumulative baseline hazard A_0 is strictly increasing and is continuous on $[0, \tau]$ for $\tau < \infty$. Call the first jump of N , T_1 .

THEOREM 1. *If $\max_i N_i(\tau) > 1$, then a maximizer of $L_n(\theta, A)$, $(\theta, A) = (\hat{\theta}, \hat{A})$, exists and is finite.*

The proof of this theorem is in the Appendix.

THEOREM 2. *Assume the following:*

(a) *Y is a nonincreasing step function, and $P[Y(t) \geq 1]$ has at most a finite number of discontinuities in $t \in [0, \tau]$ [or (a') Y is a step function with at most a bounded number of steps and an upper bound on $A_0(\tau)$ is known];*

(b) $\inf_{u \in (0, \tau]} EY(u) > 0$;

(c) $P[Y(T_1+) \geq 1] > 0$.

Then

$$\sup_{t \in (0, \tau]} |\widehat{A}(t) - A_0(t)| \rightarrow 0 \quad a.s.$$

and

$$|\widehat{\theta} - \theta_0| \rightarrow 0 \quad a.s.$$

REMARKS. Assumption (a) is used to prove that \widehat{A} does not diverge to infinity, but it should not be necessary. Note that it excludes applications requiring Y to be nonmonotonic. However, if one is willing to make the assumption that $A_0(\tau)$ belongs to a finite range, then Y is allowed to be nonmonotonic. It is unclear whether the assumption of a finite range for θ_0 is necessary. If this assumption is necessary, then this may indicate that there is a sequence of large values of θ which maximize the partial likelihood and are inconsistent. Assumption (b) ensures that N has sufficient activity on the entire interval so as to estimate the parameters. Note that (c) excludes the possibility of N having at most only one jump. Some version of this assumption should be necessary because, as pointed out by Nielsen, Gill, Andersen and Sørensen (1992), the model is unidentifiable if all of the N_i have only one jump. The method used in this proof should extend to the regression setting as long as an assumption which excludes colinearity of the covariates is made.

EXAMPLE (Survival analysis with left-truncated, right-censored data). As in the above, assume that the frailty Z has a gamma distribution with mean 1 and variance θ_0 . Given Z , let (X_1, \dots, X_k) be i.i.d. survival times with hazard $Z\alpha_0(\cdot)$. Let (T_1, \dots, T_k) and (C_1, \dots, C_k) be truncation and censoring times which are independent of both Z and the X_j 's. Define

$$G_t = \sigma\{Z, I\{X_j \leq s\}, I\{T_j < s \leq C_j\}, s \leq t, j = 1, \dots, k\}.$$

Then $N_j(t) = I\{X_j \leq t\}$, $j = 1, \dots, k$, is a multivariate counting process with intensities $I\{X_j \geq t\}Z\alpha_0(t)$, $j = 1, \dots, k$. Now define $N(t) = \sum_{j=1}^k \int_0^t I\{T_j < s \leq C_j\} dN_j(s)$. Because $I\{T_j < s \leq C_j\}$ is left-continuous (considered as a function of s) and adapted, N has intensity $Y(t)Z\alpha_0(t)$, where $Y(t) = \sum_{j=1}^k I\{T_j < s \leq C_j\} I\{X_j \geq t\}$. Since the conditional distribution of (T_1, \dots, T_k) and (C_1, \dots, C_k) given $Z = z$ is certainly independent of z , the innovation theorem [Bremaud (1981)] can be used to derive the intensity of N with respect to the observed history $F_t = \sigma\{N(s), s \leq t; Y(s), s \leq t\}$ to obtain (2.1). The filtering approach to truncation follows Keiding and Gill (1990). Note that the survival times are indeed truncated in that X_j is observed only if $X_j \wedge C_j \geq T_j$ and $X_j \leq C_j$.

As in the beginning of this section, assume the observations of (N_i, Y_i) , $i = 1, \dots, n$ are i.i.d. copies of (N, Y) . Let $(\widehat{\theta}, \widehat{A})$ be the maximizer of L_n . A_0 will be strictly increasing if $\alpha_0 \neq 0$ a.s. Since Y will not be monotone, assumption (a) cannot be satisfied; but if there is a known p , $p \in (0, 1)$, for which $P(X_1 \leq \tau) < p$,

then (a') will be satisfied. It is intuitively clear that it must be possible to observe a failure time in any subinterval of $(0, \tau]$ in order estimate A_0 on that subinterval. To ensure this, assume that $\inf_{t \in [0, \tau]} \sum_{j=1}^k P\{T_j < t \leq C_j\} > 0$. It is also clear that one needs to ensure the possibility on two or more failure times occurring in $(0, \tau]$ in order to estimate the variance of the random effect. Here a stronger assumption is made; assume that $P[\inf_{t \in [0, \tau]} \sum_{j=1}^k I\{T_j < t \leq C_j\} \geq 2] > 0$. This is enough to satisfy both assumptions (b) and (c).

An outline of the proof of consistency. Under (a'), \hat{A} is not allowed to diverge to infinity. However, if (a) is assumed, then the first step, and the hardest, is to show that \hat{A} does not diverge to infinity. A natural approach is to show that since \hat{A} maximizes L_n it cannot diverge. Because $(\hat{\theta}, \hat{A})$ maximizes the likelihood, $L_n(\hat{\theta}, \hat{A})$ minus $L_n(\bar{\theta}, \bar{A})$ must be nonnegative for any $(\bar{\theta}, \bar{A})$ in the parameter space. The idea is to show that if \hat{A} diverges, then the difference in the log-likelihoods must be negative eventually. This will be a contradiction. Unfortunately if \bar{A} is continuous, $L_n(\bar{\theta}, \bar{A})$ will be infinite for finite n , thus excluding the choice of $\bar{A} = A_0$. However, as long as \bar{A} has jumps at the jump times of \mathbf{N} , $L_n(\hat{\theta}, \hat{A}) - L_n(\bar{\theta}, \bar{A})$ will be finite. A possibility for \bar{A} is

$$(2.2) \quad \bar{A}(\cdot) = \int_0^\cdot \left(n^{-1} \sum_{i=1}^n \frac{1 + \theta_0 N_i(u-)}{1 + \theta_0 \int_0^{u-} Y_i dA_0} Y_i(u) \right)^{-1} d\bar{N}(u),$$

which can be shown to converge to A_0 . If θ_0 and A_0 were used as the initial values in the EM algorithm [Nielsen, Gill, Andersen and Sørensen (1992) explain how to use the EM algorithm], then the one-step estimator of the cumulative baseline hazard is very similar to (2.2). In the proof it is shown that, for \hat{A} diverging to infinity, $L_n(\hat{\theta}, \hat{A}) - L_n(\bar{\theta}, \bar{A})$ diverges to negative infinity. This rules out a diverging \hat{A} as a maximizer of L_n .

Since \hat{A} is not allowed to diverge, Helly's selection theorem can be used to prove the existence of a convergent subsequence of $(\hat{\theta}, \hat{A})$. The second step is to show that any such convergent subsequence of $(\hat{\theta}, \hat{A})$ must converge to (θ_0, A_0) . The approach taken here is classical, in that it depends on the positivity of the Kullback–Leibler information. This approach has been used with some success in sieve estimation [see Karr (1987) and Grenander (1981)]. The idea is to characterize the limit of a subsequence of $(\hat{\theta}, \hat{A})$ by using the fact that $L_n(\hat{\theta}, \hat{A}) - L_n(\theta_0, A_0) \geq 0$ for finite n , yet the limiting version of L_n (say, L) is maximized at (θ_0, A_0) . These two facts will yield $L(\theta, A) - L(\theta_0, A_0) = 0$, where (θ, A) is the limit of the convergent subsequence of $(\hat{\theta}, \hat{A})$. Now the problem reduces to identifiability of the parameters; that is, the parameters are identifiable if and only if $L(\theta, A) - L(\theta_0, A_0) = 0$ implies that $(\theta, A) = (\theta_0, A_0)$. Note that $L(\theta, A) - L(\theta_0, A_0)$ is minus the Kullback–Leibler information. As before, $L_n(\theta, A)$ will be infinite at $A = A_0$; instead use $A = \bar{A}$ from (2.2). Since \bar{A} converges to A_0 the proof goes through essentially as outlined above. The identifiability equation,

$L(\theta, A) - L(\theta_0, A_0) = 0$, implies that

$$\sup_{t \in [0, \tau]} \left| \int_0^t \frac{1 + \theta N(u-)}{1 + \theta \int_0^{u-} Y dA} Y(u) dA - \int_0^t \frac{1 + \theta_0 N(u-)}{1 + \theta_0 \int_0^{u-} Y dA_0} Y(u) dA_0 \right| = 0 \quad \text{a.s.}$$

This is indeed a question of identifiability, as these are the integrated intensities when (θ, A) and when (θ_0, A_0) are the true parameter values, respectively. The question becomes “Does equality of the integrated intensities imply equality of the parameters?” In the proof it is shown that sufficient conditions are (b) and (c).

PROOF OF THEOREM 2. Since this is a proof of a.s. consistency, most of the proof will be for ω fixed in a set of probability 1. This set (say, Δ) is the intersection of sets each of probability 1. Each of these is a set on which a strong law of large numbers holds for some average. Instead of listing Δ explicitly, the component sets will be obvious as the proof proceeds. It is important to be careful that the intersection only include a countable number of sets of probability 1 since an uncountable intersection can have probability less than 1. See Lemma 1 for an example in which such care is taken.

Step 1. Fix $\omega \in \Delta$ and suppose that, for some subsequence, $\lim_{n_k \rightarrow \infty} \hat{A}(\tau) = \infty$. If necessary, choose a further subsequence for which $\hat{\theta}$ converges (say, to θ). Choose $\bar{\theta} = \theta_0$. The goal is to show that $L_{n_k}(\hat{\theta}, \hat{A}) - L_{n_k}(\theta_0, \bar{A})$ will be negative for large n_k . In the following, any terms which are bounded away from positive infinity will be represented by $O(1)$, and, in an abuse of notation, n will be used instead of n_k . Recall that

$$\begin{aligned} 0 &\leq L_n(\hat{\theta}, \hat{A}) - L_n(\theta_0, \bar{A}) \\ &= n^{-1} \sum_{i=1}^n \int_0^\tau \ln \left[\frac{1 + \hat{\theta} N_i(u-)}{1 + \theta_0 N_i(u-)} \right] dN_i(u) \\ &\quad + (\theta_0^{-1} + N_i(\tau)) \ln \left(1 + \theta_0 \int_0^\tau Y_i d\bar{A} \right) \\ &\quad - (\hat{\theta}^{-1} + N_i(\tau)) \ln \left(1 + \hat{\theta} \int_0^\tau Y_i d\hat{A} \right) \\ &\quad + \int_0^\tau \ln \left(\frac{\Delta \hat{A}(u)}{\Delta \bar{A}(u)} \right) dN_i(u). \end{aligned}$$

Since $\hat{\theta}$ and θ_0 constrained to lie in $[0, M]$, only the last two terms above are important to consider,

$$\begin{aligned} 0 &\leq L_n(\hat{\theta}, \hat{A}) - L_n(\theta_0, \bar{A}) \\ &\leq O(1) - n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau)) \ln \left(1 + \hat{\theta} \int_0^\tau Y_i d\hat{A} \right) \\ &\quad + \int_0^\tau \ln \left(\frac{\Delta \hat{A}(u)}{\Delta \bar{A}(u)} \right) d\bar{N}(u). \end{aligned}$$

Substituting \bar{A} from (2.2) yields

$$O(1) - n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau)) \ln \left(1 + \hat{\theta} \int_0^\tau Y_i dA \right) - \int_0^\tau \ln \left(n^{-1} \sum_{i=1}^n \frac{1 + \hat{\theta} N_i(\tau)}{1 + \hat{\theta} \int_0^\tau Y_i dA} Y_i(u) \right) d\bar{N}(u).$$

Let K be a common upper bound on $N(\tau)$ and $\sup_{u \in [0, \tau]} Y(u)$. Since

$$(2.3) \quad K^{-1} \leq \frac{1 + \hat{\theta} K^{-1} \int_0^\tau Y_i d\hat{A}}{1 + \hat{\theta} \int_0^\tau Y_i d\hat{A}} \leq 1,$$

$$0 \leq L_n(\hat{\theta}, \hat{A}) - L_n(\theta_0, \bar{A})$$

$$(2.4) \quad \leq O(1) - n^{-1} \sum_{i=1}^n \left\{ (\hat{\theta}^{-1} + N_i(\tau)) \ln \left(1 + \hat{\theta} \int_0^\tau Y_i d\hat{A} \right) - \int_0^\tau \ln \left(n^{-1} \sum_{j=1}^n \frac{1 + \hat{\theta} N_j(\tau)}{1 + \hat{\theta} K^{-1} \int_0^\tau Y_j d\hat{A}} Y_j(u) \right) dN_i(u) \right\}.$$

Intuitively, as \hat{A} diverges to infinity, the second term above will diverge to negative infinity and the third term will diverge to positive infinity. The idea is to show that the rate of the second term is faster than the rate of the third term, so that eventually the difference is negative and a contradiction ensues. To understand the following argument better, combine the last two terms above to obtain

$$-n^{-1} \sum_{i=1}^n \int_0^\tau \ln \left(n^{-1} \sum_{j=1}^n \left(1 + \hat{\theta} \int_0^\tau Y_i d\hat{A} \right)^{(\hat{\theta} N_i(\tau))^{-1} + 1} \times \frac{1 + \hat{\theta} N_j(\tau)}{1 + \hat{\theta} K^{-1} \int_0^\tau Y_j d\hat{A}} Y_j(u) \right) dN_i(u).$$

In the sum over j , the terms with $\int_0^\tau Y_j d\hat{A} \gg \int_0^\tau Y_i d\hat{A}$ will be approximately zero if \hat{A} is large. So the only terms to pay attention to are the terms that are smaller or of comparable size to $\int_0^\tau Y_i d\hat{A}$. The key is to show that there are enough of these terms so that their average times $1 + \hat{\theta} \int_0^\tau Y_i d\hat{A}$ diverges to positive infinity. For simplicity assume that $P[Y(t) \geq 1]$ is continuous in t .

Partition both of the terms in (2.4) according to a nonnegative strictly decreasing sequence $\tau = s_0 > s_1 > \dots \geq 0$ to obtain

$$\begin{aligned} & -n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau)) \ln \left(1 + \hat{\theta} \int_0^\tau Y_i d\hat{A} \right) \\ & \leq -n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau)) I\{Y_i(\tau) \geq 1\} \ln(1 + \hat{\theta} \hat{A}(\tau)) \\ & \quad - \sum_{p=1}^N n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau)) I\{Y_i(s_p) \geq 1, Y_i(s_{p-1}) = 0\} \ln(1 + \hat{\theta} \hat{A}(s_p)) \\ & \quad - n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau)) I\{Y_i(s_N) = 0\} \ln \left(1 + \hat{\theta} \int_0^\tau Y_i d\hat{A} \right) \end{aligned}$$

and

$$\begin{aligned} & - \int_0^\tau \ln \left(n^{-1} \sum_{j=1}^n \frac{1 + \hat{\theta} N_j(\tau)}{1 + \hat{\theta} K^{-1} \int_0^\tau Y_j d\hat{A}} Y_j(u) \right) d\bar{N}(u) \\ & \leq -n^{-1} \sum_{i=1}^n I\{Y_i(s_1) \geq 1\} \int_0^\tau \ln \left(n^{-1} \sum_{j=1}^n \frac{1 + \hat{\theta} N_j(\tau)}{1 + \hat{\theta} \hat{A}(\tau)} Y_j(u) \right) dN_i(u) \\ & \quad - \sum_{p=1}^N n^{-1} \sum_{i=1}^n I\{Y_i(s_{p+1}) \geq 1, Y_i(s_p) = 0\} \\ & \quad \quad \times \int_0^\tau \ln \left(n^{-1} \sum_{j=1}^n \frac{1 + \hat{\theta} N_j(\tau)}{1 + \hat{\theta} \hat{A}(s_p)} I\{Y_j(s_p) = 0\} Y_j(u) \right) dN_i(u) \\ & \quad - n^{-1} \sum_{i=1}^n I\{Y_i(s_{N+1}) = 0\} \\ & \quad \quad \times \int_0^\tau \ln \left(n^{-1} \sum_{j=1}^n \frac{1 + \hat{\theta} N_j(\tau)}{1 + \hat{\theta} K^{-1} \int_0^\tau Y_j d\hat{A}} Y_j(u) \right) dN_i(u). \end{aligned}$$

Combining these two sums term by term yields

$$\begin{aligned} & 0 \leq L_n(\hat{\theta}, \hat{A}) - L_n(\theta_0, \bar{A}) \\ & \leq O(1) - \ln[1 + \hat{\theta} \hat{A}(\tau)] \left[n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau)) I\{Y_i(\tau) \geq 1\} \right. \\ & \quad \left. - n^{-1} \sum_{i=1}^n N_i(\tau) I\{Y_i(s_1) \geq 1\} \right] \end{aligned}$$

$$\begin{aligned}
 & - \sum_{p=1}^N \ln[1 + \hat{\theta} \hat{A}(s_p)] \left[n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau)) I\{Y_i(s_p) \geq 1, Y_i(s_{p-1}) = 0\} \right. \\
 & \qquad \qquad \qquad \left. - n^{-1} \sum_{i=1}^n N_i(\tau) I\{Y_i(s_{p+1}) \geq 1, Y_i(s_p) = 0\} \right] \\
 (2.5) \quad & - n^{-1} \sum_{i=1}^n I\{Y_i(s_{N+1}) = 0\} \int_0^\tau \ln \left(n^{-1} \sum_{j=1}^n \frac{1 + \hat{\theta} N_j(\tau)}{1 + \hat{\theta} K^{-1} \int_0^\tau Y_j d\hat{A}} Y_j(u) \right) dN_i(u) \\
 & - \sum_{p=1}^N n^{-1} \sum_{i=1}^n I\{Y_i(s_{p+1}) \geq 1, Y_i(s_p) = 0\} \\
 & \qquad \qquad \qquad \times \int_0^\tau \ln \left[n^{-1} \sum_{j=1}^n (1 + \hat{\theta} N_j(\tau)) Y_j(u) I\{Y_j(s_p) = 0\} \right] dN_i(u).
 \end{aligned}$$

The sequence $\{s_i\}_{i \geq 0}$ needs to satisfy three conditions:

- (a) the last term above does not diverge to positive infinity;
- (b) $s_{N+1} = 0$ so that the second-to-last term is zero;
- (c) the coefficients of $\ln[1 + \hat{\theta} \hat{A}(s_p)]$ are positive for large n .

Recall that $\hat{\theta}$ converges to θ . Choose $U > 1$ and $s_0 = \tau$. If $\theta = 0$, choose $s_1 = 0$. In this case it is easily shown that $\lim \hat{A}(\tau) = \infty$ implies that $L_n(\hat{\theta}, \hat{A}) - L_n(\theta_0, \bar{A})$ diverges to negative infinity, and the first step is completed. Assume that $\theta > 0$. Choose s_1 to be the smallest value (greater than or equal to 0) for which

$$E((U\theta)^{-1} + N(\tau))I\{Y(\tau) \geq 1\} \geq E(N(\tau)I\{Y(s_1) \geq 1\}).$$

Unless $s_1 = 0$, continuity of $P[Y(T) \geq 1]$ implies equality above. Then given $s_p, p \geq 1$, choose s_{p+1} to be the smallest value (greater than or equal to 0) for which

$$\begin{aligned}
 & E((U\theta)^{-1} + N(\tau))I\{Y(s_p) \geq 1, Y(s_{p-1}) = 0\} \\
 & \geq E(N(\tau)I\{Y(s_{p+1}) \geq 1, Y(s_p) = 0\}).
 \end{aligned}$$

Once again, unless $s_{p+1} = 0$, continuity of $P[Y(t) \geq 1]$ implies equality above. If there exists an $N < \infty$ for which $s_{N+1} = 0$, stop. Suppose this does not occur. Then the k th partial sum is given by

$$\begin{aligned}
 & E((U\theta)^{-1} + N(\tau))I\{Y(\tau) \geq 1\} + \sum_{p=1}^k E((U\theta)^{-1} + N(\tau))I\{Y(s_p) \geq 1, Y(s_{p-1}) = 0\} \\
 & = E(N(\tau)I\{Y(s_1) \geq 1\}) + \sum_{p=1}^k E(N(\tau)I\{Y(s_{p+1}) \geq 1, Y(s_p) = 0\}),
 \end{aligned}$$

which implies that

$$E((U\theta)^{-1} + N(\tau))I\{Y(s_k) \geq 1\} = E(N(\tau)I\{Y(s_{k+1}) \geq 1\}).$$

Since the sequence $\{s_p\}$ is decreasing and is nonnegative, it converges (say, to s^0). Taking the limit as k goes to infinity of the above, results in

$$E((U\theta)^{-1} + N(\tau))I\{Y(s^0+) \geq 1\} = E(N(\tau)I\{Y(s^0+) \geq 1\}).$$

This is a contradiction; so there exists a finite N for which $s_{N+1} = 0$ and both conditions (b) and (c) will be satisfied.

All that is left is to verify condition (a), that is, that the last term in (2.5) does not diverge to positive infinity. This term is bounded above by

$$\begin{aligned} & - \sum_{p=1}^N n^{-1} \sum_{i=1}^n I\{Y_i(s_{p+1}) \geq 1, Y_i(s_p) = 0\} \\ & \quad \times \int_0^\tau \ln \left[n^{-1} \sum_{j=1}^n (1 + \hat{\theta}N_j(\tau))I\{Y_j(u) \geq 1, Y_j(s_p) = 0\} \right] dN_i(u). \end{aligned}$$

On $I\{Y_j(s_{p+1}) \geq 1, Y_j(s_p) = 0\}(N_j(s_p) - N_j(u-)) > 0$,

$$\frac{I\{Y_j(u) \geq 1, Y_j(s_p) = 0\}(1 + \hat{\theta}N_j(\tau))}{I\{Y_j(s_{p+1}) \geq 1, Y_j(s_p) = 0\}(N_j(s_p) - N_j(u-))} > 1.$$

Inequality (2.5) becomes

$$\begin{aligned} & 0 \leq L_n(\hat{\theta}, \hat{A}) - L_n(\theta_0, \bar{A}) \\ & \leq O(1) - \ln [1 + \hat{\theta}\hat{A}(\tau)] \left[n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau))I\{Y_i(\tau) \geq 1\} \right. \\ & \quad \left. - n^{-1} \sum_{i=1}^n I\{Y_i(s_1) \geq 1\}N_i(\tau) \right] \\ & - \sum_{p=1}^N \ln [1 + \hat{\theta}\hat{A}(s_j)] \left[n^{-1} \sum_{i=1}^n (\hat{\theta}^{-1} + N_i(\tau))I\{Y_i(s_p) \geq 1, Y_i(s_{p-1}) = 0\} \right. \\ & \quad \left. - n^{-1} \sum_{i=1}^n N_i(\tau)I\{Y_i(s_{p+1}) \geq 1, Y_i(s_p) = 0\} \right] \\ & - \sum_{p=1}^N n^{-1} \sum_{i=1}^n I\{Y_i(s_{p+1}) \geq 1, Y_i(s_p) = 0\} \\ & \quad \times \int_0^\tau \ln \left[n^{-1} \sum_{j=1}^n I\{Y_j(s_{p+1}) \geq 1, Y_j(s_p) = 0\} (N_j(s_p) - N_j(u-)) \right] dN_i(u). \end{aligned}$$

This last term is just $-\sum_{p=1}^N \sum_{i=1}^{nN(p)} i^{-1} (i/n) \ln(i/n)$, where

$$N(p) = n^{-1} \sum_{j=1}^n I\{Y_j(s_{p+1}) \geq 1, Y_j(s_p) = 0\} N_j(s_p).$$

Since N is finite and each $N(p)$ converges to a positive value, the last term does not diverge to positive infinity.

If $P\{Y(t) \geq 1\}$ has a finite number of discontinuities, then a similar argument involving both the discontinuities and the s_p 's will achieve the same result.

Step 2. Once again fix $\omega \in \Delta$. Step 1 implies that $\limsup \widehat{A}(\tau) < \infty$, that is, there exists some a , depending on ω , which serves as an upper bound on \widehat{A} . Any infinite sequence in the product of the set $[0, M]$ with the set of bounded cumulative hazards has a pointwise convergent subsequence, say, for the subsequence $\{n_k\}$. In an abuse of notation, denote the convergent subsequence of $(\widehat{\theta}, \widehat{A})$ by $(\widehat{\theta}, \widehat{A})$ and use n instead of n_k . Let (θ, A) be the limit point. In Lemma 1 it is shown that A is continuous. The goal is to show that $L(\theta, A) - L(\theta_0, A_0) = 0$ and then to conclude that $(\theta, A) = (\theta_0, A_0)$. Recall that \widehat{A} is given by (2.2).

To begin, consider

$$\begin{aligned} & 0 \leq L_n(\widehat{\theta}, \widehat{A}) - L_n(\theta_0, \bar{A}) \\ &= n^{-1} \sum_{i=1}^n \int_0^\tau \ln \left[\frac{\left([1 + \widehat{\theta} N_i(u-)] / [1 + \widehat{\theta} \int_0^{u-} Y_i d\widehat{A}] \right) Y_i(u) \Delta \widehat{A}(u)}{\left([1 + \theta_0 N_i(u-)] / [1 + \theta_0 \int_0^{u-} Y_i dA_0] \right) Y_i(u) \Delta \bar{A}(u)} \right] \\ & \quad \times \left[dN_i(u) - \frac{1 + \theta_0 N_i(u-)}{1 + \theta_0 \int_0^{u-} Y_i dA_0} Y_i(u) d\bar{A}(u) \right] \\ (2.6) \quad &+ n^{-1} \sum_{i=1}^n \int_0^\tau \left(\ln \left[\frac{\left([1 + \widehat{\theta} N_i(u-)] / [1 + \widehat{\theta} \int_0^{u-} Y_i d\widehat{A}] \right) Y_i(u) \Delta \widehat{A}(u)}{\left([1 + \theta_0 N_i(u-)] / [1 + \theta_0 \int_0^{u-} Y_i dA_0] \right) Y_i(u) \Delta \bar{A}(u)} \right] \right. \\ & \quad \left. - \left[\frac{\left([1 + \widehat{\theta} N_i(u-)] / [1 + \widehat{\theta} \int_0^{u-} Y_i d\widehat{A}] \right) Y_i(u) d\widehat{A}(u)}{\left([1 + \theta_0 N_i(u-)] / [1 + \theta_0 \int_0^{u-} Y_i dA_0] \right) Y_i(u) d\bar{A}(u)} - 1 \right] \right) \\ & \quad \times \frac{1 + \theta_0 N_i(u-)}{1 + \theta_0 \int_0^{u-} Y_i dA_0} Y_i(u) d\bar{A}(u) \\ &+ [L_n(\widehat{\theta}, \bar{A}) - L_n(\theta_0, \bar{A}) - L'_n(\widehat{\theta}, \widehat{A}) + L'_n(\theta_0, \bar{A})]. \end{aligned}$$

Note that, for $x \geq 0$, $\ln(x) - (x - 1) \leq 0$, implying that the second term above is nonpositive. In Lemma 2, it is shown that the first and third terms above converge to zero; this implies that the limit of the second term is zero. Recall

that \widehat{A} satisfies

$$\begin{aligned} \widehat{A}(t) &= \int_0^t \left(n^{-1} \sum_{i=1}^n \frac{1 + \widehat{\theta} N_i(\tau)}{1 + \widehat{\theta} \int_0^\tau Y_i d\widehat{A}} Y_i(u) \right)^{-1} d\widehat{N}(u) \\ &= \int_0^t \left[\frac{n^{-1} \sum_{i=1}^n \left([1 + \theta_0 N_i(u-)] / [1 + \theta_0 \int_0^{u-} Y_i dA_0] \right) Y_i(u)}{n^{-1} \sum_{i=1}^n \left([1 + \widehat{\theta} N_i(\tau)] / [1 + \widehat{\theta} \int_0^\tau Y_i d\widehat{A}] \right) Y_i(u)} \right] d\widehat{A}(u). \end{aligned}$$

In Lemma 1, it is shown that the above integrand, $d\widehat{A}/d\bar{A}$, converges in supremum norm to γ , where

$$\gamma(u) = \frac{E \left[\left([1 + \theta_0 N(u-)] / [1 + \theta_0 \int_0^{u-} Y dA_\sigma] \right) Y(u) \right]}{E \left[\left([1 + \theta N(\tau)] / [1 + \theta \int_0^\tau Y dA] \right) Y(u) \right]}$$

and $A(t) = \int_0^t \gamma dA_0$. Similarly, the second term in (2.6) converges to minus the Kullback–Leibler information $[L(\theta, A) - L(\theta_0, A_0)]$,

$$\begin{aligned} (2.7) \quad & E \int_0^\tau \left(\ln \left[\frac{\left([1 + \theta N(u-)] / [1 + \theta \int_0^{u-} Y dA] \right) Y(u)}{\left([1 + \theta_0 N(u-)] / [1 + \theta_0 \int_0^{u-} Y dA_0] \right) Y(u)} \gamma(u) \right] \right. \\ & \left. - \left[\frac{\left([1 + \theta N(u-)] / [1 + \theta \int_0^{u-} Y dA] \right) Y(u)}{\left([1 + \theta_0 N(u-)] / [1 + \theta_0 \int_0^{u-} Y dA_0] \right) Y(u)} \gamma(u) - 1 \right] \right) \\ & \times \frac{1 + \theta_0 N(u-)}{1 + \theta_0 \int_0^{u-} Y dA_0} Y(u) dA_0(u). \end{aligned}$$

The above is equal to zero as stated earlier. Formula (2.7) depends on ω and the subsequence $\{n_k\}$ only through the choice of (θ, A) .

Next show that $(\theta, A) = (\theta_0, A_0)$. As Grenander [(1981), page 398] demonstrates

$$\ln(x) - (x - 1) \leq \begin{cases} -\eta(x - 1)^2, & \text{for } \frac{1}{2} < x < \frac{3}{2}, \\ -\eta'|x - 1|, & \text{for } 0 \leq x \leq \frac{1}{2}, \quad x \geq \frac{3}{2}, \end{cases}$$

where η and η' are positive constants. Applying the above equation to (2.7) yields, after some easy manipulation, that

$$(2.8) \quad E \int_0^\tau \left| \frac{1 + \theta N(u-)}{1 + \theta \int_0^{u-} Y \gamma dA_0} \gamma(u) - \frac{1 + \theta_0 N(u-)}{1 + \theta_0 \int_0^{u-} Y dA_0} \right| Y(u) dA_0(u) = 0.$$

If the parameters are identifiable then the above equation should imply that $(\theta, A) = (\theta_0, A_0)$. This equation depends on ω and the subsequence n_k only

through the choice of θ and γ . Lemma 1 implies that γ is a bounded left-continuous function with right-hand limits. Therefore, γ has at most a countable number of discontinuities. Because of the left-continuity of all functions involved, (2.8) implies that

$$\frac{1 + \theta N(u-)}{1 + \theta \int_0^{u-} Y \gamma dA_0} \gamma(u) Y(u) = \frac{1 + \theta_0 N(u-)}{1 + \theta_0 \int_0^{u-} Y dA_0} Y(u),$$

for all u and a.e. P . Let T_1 be the time of the first jump of N . Note that $P[Y(T_1) \geq 1] = 1$ and the probability that T_1 is not equal to any of the discontinuity points of γ is 1. Take the intersection of all of these sets of probability 1 (including the a.e. set in the last equation) to get a further set of probability 1. Intersect this set with $\{Y(T_1+) \geq 1\}$. This last set will have probability greater than zero by assumption (c). On this set, one has that both of the following hold:

$$\frac{\gamma(T_1)}{1 + \theta \int_0^{T_1} Y \gamma dA_0} Y(T_1) = \frac{1}{1 + \theta_0 \int_0^{T_1} Y dA_0} Y(T_1)$$

and

$$\frac{(1 + \theta)\gamma(T_1)}{1 + \theta \int_0^{T_1} Y \gamma dA_0} Y(T_1+) = \frac{1 + \theta_0}{1 + \theta_0 \int_0^{T_1} Y dA_0} Y(T_1+).$$

Noting that $Y(T_1)$ and $Y(T_1+)$ are positive, take the difference of the above equations [leaving out $Y(T_1)$ and $Y(T_1+)$] to obtain

$$\frac{\theta_\gamma(T_1)}{1 + \theta \int_0^{T_1} Y \gamma dA_0} = \frac{\theta_0}{1 + \theta_0 \int_0^{T_1} Y dA_0}.$$

This implies that $\theta = \theta_0$.

All that is left to this step is to show that $A = A_0$. Equation (2.8) yields

$$\gamma(u) \left(1 + \theta_0 \int_0^u Y dA_0 \right) Y(u) = \left(1 + \theta_0 \int_0^u Y \gamma dA_0 \right) Y(u)$$

a.e. $dA_0 \times dP$, which, combined with the left-continuity of Y implies that

$$(\gamma(u) - 1) EY(u) + \theta_0 \int_0^u (\gamma(u) - \gamma(v)) EY(u) Y(v) dA_0(v) = 0$$

for all $u \in (0, \tau]$. If γ can be shown to be identically equal to 1, then the proof will be done. The supremum and infimum of γ on $[0, \tau]$ are either attained at a point or attained by evaluating a right-hand limit of γ at a point. For simplicity assume the former; the proof is similar if the latter holds.

Suppose that the maximum of γ on $[0, \tau]$ is attained at t_0 and $\gamma(t_0) \geq 1$. Setting $u = t_0$ into the equation for γ results in $\gamma(t_0) = 1$. Suppose that the minimum of γ is attained at t_0 and that $\gamma(t_0) \leq 1$. Once again setting $u = t_0$

into the equation for γ results in $\gamma(t_0) = 1$. These two results imply that γ is identically equal to 1. Therefore $\gamma = 1$ a.e. dA_0 .

So if $(\hat{\theta}, \hat{A})$ is a convergent subsequence, then it must converge to (θ_0, A_0) . Helly's selection theorem then implies that the entire sequence of $(\hat{\theta}, \hat{A})$ must converge to (θ_0, A_0) . The proof can be carried out for each $\omega \in \Delta$. Since Δ is the intersection of a countable number of sets of probability 1, $P(\Delta) = 1$ and $(\hat{\theta}, \hat{A}) \rightarrow (\theta_0, A_0)$ a.s. The continuity of A_0 then gives the a.s. convergence in supremum norm. \square

APPENDIX

PROOF OF THEOREM 1. An outline of the proof goes as follows:

1. Observe that \hat{A} must be discrete with positive jump sizes at the jumps of $\sum_{i=1}^n N_i$.

2. The log partial likelihood L_n is continuous function of θ and the jump sizes of A , that is, it is a continuous function on the set $[0, M] \times [0, U]^N$, where U is finite and $N = \sum_{i=1}^n N_i$.

3. Show that there exists a U such that for each possible value of $(\theta, A) \in \{[0, M] \times [0, \infty)^N\} \setminus \{[0, M] \times [0, U]^N\}$ there is a value of $(\theta, A) \in [0, M] \times [0, U]^N$ which has a higher value of L_n . This can be easily done by using a proof by contradiction, that is, assume the existence of $\theta_U, A_U \in \{[0, M] \times [0, \infty)^N\} \setminus \{[0, M] \times [0, U]^N\}$, which maximizes L_n for each U . Then show that $L_n(\theta_U, A_U)$ can be made as small as desired by increasing U . This is the desired contradiction.

The proofs of the above steps are relatively straightforward. However, care must be taken in step 3 because of the possibility that θ_U is equal to or arbitrarily close to 0. \square

LEMMA 1. Assume (b) of the consistency theorem. Then

$$\sup_{t \in [0, \tau]} |\bar{A}(t) - A_0(t)| \rightarrow 0 \quad a.s.,$$

and for each $\omega \in \Delta$ and any subsequence of $(\hat{\theta}, \hat{A})$, converging to some (θ, A) (\hat{A} converging to A at all continuity points of A),

$$\sup_{t \in [0, \tau]} \left| \frac{d\hat{A}}{dA}(t) - \gamma(t) \right| \rightarrow 0, \quad \sup_{t \in [0, \tau]} \left| \hat{A}(t) - \int_0^t \gamma dA_0 \right| \rightarrow 0,$$

where γ is defined by

$$\gamma(t) = E \left[\frac{1 + \theta_0 N(t-)}{1 + \theta_0 \int_0^{t-} Y dA_0} Y(t) \right] \bigg/ E \left[\frac{1 + \theta N(\tau)}{1 + \theta \int_0^\tau Y dA} Y(t) \right]$$

and A is continuous.

PROOF. The \bar{A} is given by

$$\bar{A}(\cdot) = \int_0^\cdot \left(n^{-1} \sum_i \frac{1 + \theta_0 N_i(u-)}{1 + \theta_0 \int_0^{u-} Y_i dA_0} Y_i(u) \right)^{-1} d\bar{N}(u).$$

Rao's (1963) strong law of large numbers implies that both the integrand and the integrator converge in supremum norm to their expectations. So

$$\begin{aligned} |\bar{A}(t) - A_0(t)| \leq & \int_0^t \left| \left(n^{-1} \sum_{i=1}^n \frac{1 + \theta_0 N_i(u-)}{1 + \theta_0 \int_0^{u-} Y_i dA_0} Y_i(u) \right)^{-1} \right. \\ & \left. - \left(E \left(\frac{1 + \theta_0 N(u-)}{1 + \theta_0 \int_0^{u-} Y dA_0} Y(u) \right) \right)^{-1} \right| d\bar{N}(u) \\ & + \left| \int_0^t \left(E \left(\frac{1 + \theta_0 N(u-)}{1 + \theta_0 \int_0^{u-} Y dA_0} Y(u) \right) \right)^{-1} d(\bar{N}(u) - E(N(u))) \right|. \end{aligned}$$

The first term goes to zero in supremum norm since $N(\tau)$ is bounded. To show that the last term goes to zero, note that the integrand is left-continuous with right-hand limits; hence a Helly-Bray argument can be used, that is, the integrand can be approximated in supremum norm by a function of bounded variation. This plus an integration-by-parts argument suffices to prove convergence to zero in supremum norm.

In order to prove the second result, it is helpful first to characterize the possible limit of a convergent subsequence of \hat{A}, A . The following says that A must be continuous. Let f be any nonnegative, bounded, continuous function. Then

$$\begin{aligned} \int f dA &= \int f d(A - \hat{A}) + \int f(u) \left[n^{-1} \sum_{i=1}^n \frac{1 + \hat{\theta} N_i(\tau)}{1 + \hat{\theta} \int_0^\tau Y_i d\hat{A}} Y_i(u) \right]^{-1} d\bar{N}(u) \\ &\leq \int f d(A - \hat{A}) + \int f(u) \left[n^{-1} \sum_{i=1}^n Y_i(u) \right]^{-1} d\bar{N}(u) (1 + MKa), \end{aligned}$$

where M is the bound on $\hat{\theta}$; K is the bound on Y ; and a is larger than $\limsup \hat{A}(\tau)$. The law of large numbers implies that $n^{-1} \sum_{i=1}^n Y_i$ and \bar{N} converge in supremum norm to $E(Y)$ and

$$\int_0^\cdot E \left(\frac{1 + \theta_0 N(u-)}{1 + \int_0^{u-} Y dA_0} Y(u) \right) dA_0(u),$$

respectively. Also, $E(Y)$ is bounded away from zero by assumption (b) of the consistency theorem. So

$$\int f dA \leq \int f(u) (E(Y(u)))^{-1} E \left(\frac{1 + \theta_0 N(u-)}{1 + \int_0^{u-} Y dA_0} Y(u) \right) dA_0(u) (1 + MKa).$$

Choosing f appropriately implies that A must be continuous at the continuity points of A_0 .

Consider

$$(A.1) \quad \frac{d\hat{A}}{dA}(t) = \left[n^{-1} \sum_{i=1}^n \frac{1 + \theta_0 N_i(t-)}{1 + \theta_0 \int_0^t Y_i dA_0} Y_i(t) \right] \Big/ \left[n^{-1} \sum_{i=1}^n \frac{1 + \hat{\theta} N_i(\tau)}{1 + \hat{\theta} \int_0^\tau Y_i d\hat{A}} Y_i(t) \right].$$

The numerator converges in supremum norm to its expectation by the strong law of large numbers. One expects that the denominator will converge to

$$E \left[\frac{1 + \theta N(\tau)}{1 + \theta \int_0^\tau Y dA} Y(t) \right].$$

This requires a more careful argument. Consider,

$$(A.2) \quad \begin{aligned} & \left| n^{-1} \sum_{i=1}^n \frac{1 + \hat{\theta} N_i(\tau)}{1 + \hat{\theta} \int_0^\tau Y_i d\hat{A}} Y_i(t) - E \left[\frac{1 + \theta N(\tau)}{1 + \theta \int_0^\tau Y dA} Y(t) \right] \right| \\ & \leq \left| n^{-1} \sum_{i=1}^n \frac{(\hat{\theta} - \theta) N_i(\tau)}{1 + \hat{\theta} \int_0^\tau Y_i d\hat{A}} Y_i(t) \right| \\ & \quad + \left| n^{-1} \sum_{i=1}^n \frac{(1 + \theta N_i(\tau)) Y_i(t) \int_0^\tau Y_i d(\hat{\theta} \hat{A} - \theta A)}{(1 + \hat{\theta} \int_0^\tau Y_i d\hat{A})(1 + \theta \int_0^\tau Y_i dA)} \right| \\ & \quad \times \left| n^{-1} \sum_{i=1}^n \frac{1 + \theta N_i(\tau)}{1 + \theta \int_0^\tau Y_i dA} Y_i(t) - E \left[\frac{1 + \theta N(\tau)}{1 + \theta \int_0^\tau Y dA} Y(t) \right] \right|. \end{aligned}$$

Note that the first term above is bounded by the difference between $\hat{\theta}$ and θ times a constant, and therefore goes to zero. The second term also goes to zero since it is bounded above by a constant times the supremum norm of $|\hat{\theta} \hat{A} - \theta A|$. Since A_0 is continuous, A is also continuous, and therefore $\sup_{t \in [0, \tau]} |\hat{A}(t) - A(t)|$ converges to zero. At first thought it appears that the strong law of large numbers is sufficient to prove that the last term in (A.2) converges to zero. Recall that Δ in Theorem 2 can be at most the intersection of a countable number of sets of probability 1. However, the set of probability 1 may change with A , and each ω can have more than one limit point for the sequence \hat{A} . Since there are an uncountable number of ω 's and therefore possibly an uncountable number of A 's, the intersection of the corresponding sets of probability 1 may have probability 0. Here is one way around this problem. The space of continuous distribution functions is separable under the supremum norm. Let $\{G_l\}_{l \geq 1}$ be a countable, dense set. Include in the intersection of sets forming Δ sets for which

$$\sup_{t \in [0, \tau]} \left| n^{-1} \sum_{i=1}^n \frac{1 + \eta N_i(\tau)}{1 + \eta \eta' \int_0^\tau Y_i dG_l} Y_i(t) - E \left[\frac{1 + \eta N(\tau)}{1 + \eta \eta' \int_0^\tau Y dG_l} Y(t) \right] \right|$$

converges to zero for each rational pair (η, η') and $l \geq 1$. Continuing with the

proof that the last term of (A.2) goes to zero,

$$\begin{aligned} & \left| n^{-1} \sum_{i=1}^n \frac{1 + \theta N_i(\tau)}{1 + \theta \int_0^\tau Y_i dA} Y_i(t) - E \left[\frac{1 + \theta N(\tau)}{1 + \theta \int_0^\tau Y dA} Y(t) \right] \right| \\ & \leq \left| n^{-1} \sum_{i=1}^n \frac{(\theta - \eta) N_i(\tau)}{1 + \theta \int_0^\tau Y_i dA} Y_i(t) \right| \\ & \quad + \left| n^{-1} \sum_{i=1}^n \frac{(1 + \eta N_i(\tau)) Y_i(t) \int_0^\tau Y_i d(\theta A - \eta \eta' G_l)}{(1 + \theta \int_0^\tau Y_i dA)(1 + \eta \eta' \int_0^\tau Y_i dG_l)} \right| \\ & \quad + \left| n^{-1} \sum_{i=1}^n \frac{1 + \eta N_i(\tau)}{1 + \eta \eta' \int_0^\tau Y_i dG_l} Y_i(t) - E \left[\frac{1 + \eta N(\tau)}{1 + \eta \eta' \int_0^\tau Y dG_l} Y(t) \right] \right| \\ & \quad + \left| E \left[\frac{(\theta - \eta) N(\tau)}{1 + \theta \int_0^\tau Y dA} Y(t) \right] \right| + \left| E \left[\frac{(1 + \eta N(\tau)) Y(t) \int_0^\tau Y d(\theta A - \eta \eta' G_l)}{(1 + \theta \int_0^\tau Y dA)(1 + \eta \eta' \int_0^\tau Y dG_l)} \right] \right|. \end{aligned}$$

As n increases, the above terms can be made as small as desired by proper choice of η, η' and l . The denominator of (A.1) converges as expected and the second result is proved.

The third result follows by one more application of the Helly-Bray argument. □

LEMMA 2. Assume (b) of the consistency theorem, for each $\omega \in \Delta$. Then the following hold:

$$\begin{aligned} & n^{-1} \sum_{i=1}^n \int_0^\tau \ln \left[\frac{([1 + \hat{\theta} N_i(u-)]/[1 + \hat{\theta} \int_0^{u-} Y_i d\hat{A}]) Y_i(u) d\hat{A}(u)}{([1 + \theta_0 N_i(u-)]/[1 + \theta_0 \int_0^{u-} Y_i dA_0]) Y_i(u) d\bar{A}(u)} \right] \\ & \quad \times \left[dN_i(u) - \frac{1 + \theta_0 N_i(u-)}{1 + \theta_0 \int_0^{u-} Y_i dA_0} Y_i(u) d\bar{A}(u) \right] \end{aligned}$$

converges to zero;

$$\begin{aligned} & n^{-1} \sum_{i=1}^n \int_0^\tau \left(\ln \left[\frac{([1 + \hat{\theta} N_i(u-)]/[1 + \hat{\theta} \int_0^{u-} Y_i d\hat{A}]) Y_i(u) d\hat{A}(u)}{([1 + \theta_0 N_i(u-)]/[1 + \theta_0 \int_0^{u-} Y_i dA_0]) Y_i(u) d\bar{A}(u)} \right] \right. \\ & \quad \left. - \left[\frac{([1 + \hat{\theta} N_i(u-)]/[1 + \hat{\theta} \int_0^{u-} Y_i d\hat{A}]) Y_i(u) d\hat{A}(u)}{([1 + \theta_0 N_i(u-)]/[1 + \theta_0 \int_0^{u-} Y_i dA_0]) Y_i(u) d\bar{A}(u)} - 1 \right] \right) \\ & \quad \times \frac{1 + \theta_0 N_i(u-)}{1 + \theta_0 \int_0^{u-} Y_i dA_0} Y_i(u) d\bar{A}(u) \end{aligned}$$

converges to

$$E \int_0^\tau \left(\ln \left[\frac{([1 + \theta N(u-)]/[1 + \theta \int_0^{u-} Y dA]) Y(u)}{([1 + \theta_0 N(u-)]/[1 + \theta_0 \int_0^{u-} Y dA_0]) Y(u)} \right] \gamma(u) \right. \\ \left. - \left[\frac{([1 + \theta N(u-)]/[1 + \theta \int_0^{u-} Y dA]) Y(u)}{([1 + \theta_0 N(u-)]/[1 + \theta_0 \int_0^{u-} Y dA_0]) Y(u)} \gamma(u) - 1 \right] \right) \\ \times \frac{1 + \theta_0 N(u-)}{1 + \theta_0 \int_0^{u-} Y dA_0} Y(u) dA_0(u);$$

and

$$L_n(\hat{\theta}, \hat{A}) - L_n(\theta_0, \bar{A}) - L'_n(\hat{\theta}, \hat{A}) + L'_n(\theta_0, \bar{A})$$

converges to zero.

PROOF. In the first two equations above, $d\hat{A}/d\bar{A}$ can be replaced by its limit, γ . This is justified by the second result of Lemma 1. Arguments similar to those used in proving the second result of Lemma 1 can also be used to justify the substitution of A for \hat{A} in the equations in Lemma 2. Finally, both results can be proved by employing the Helly-Bray argument outlined in Lemma 1, the $\{G_l\}_{l \geq 1}$ and the first result of Lemma 1.

Recall that

$$L'_n(\theta, A) - L_n(\theta, A) = n^{-1} \sum_{i=1}^n \int_0^\tau (1 + \theta N_i(u)) d \left[\ln \left(1 + \theta \int_0^u Y_i dA \right)^{1/\theta} \right] \\ - n^{-1} \sum_{i=1}^n \int_0^\tau \frac{1 + \theta N_i(u-)}{1 + \hat{\theta} \int_0^{u-} Y_i dA} Y_i(u) dA(u).$$

Since both \hat{A} and \bar{A} converge to continuous limits, it is easily shown that the above difference evaluated either at $(\theta, A) = (\hat{\theta}, \hat{A})$ or $(\theta, A) = (\theta_0, \bar{A})$ converges to zero. \square

Acknowledgment. The author is greatly indebted to Richard Gill for suggesting the topic of this investigation.

REFERENCES

- ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1988). Censoring, truncation and filtering in statistical models based on counting processes. *Contemp. Math.* **80** 19–59.
 ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

- BREMAUD, P. (1981). *Point Processes and Queues*. Springer, New York.
- CLAYTON, D. and CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *J. Roy. Statist. Soc. Ser. A* **148** 82–117.
- GREENWOOD, P. E. and WEFELMEYER, W. (1990). Efficiency of estimators for partially specified filtered models. *Stochastic Process. Appl.* **36** 353–370.
- GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York.
- KARR, A. F. (1987). Maximum likelihood estimation in the multiplicative intensity model via sieves. *Ann. Statist.* **15** 473–490.
- KEIDING, N. and GILL, R. D. (1990). Random truncation models and Markov processes. *Ann. Statist.* **18** 582–602.
- NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. and SØRENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19** 25–44.
- RAO, R. R. (1963). The law of large numbers for $D[0,1]$ -valued random variables. *Theory Probab. Appl.* **8** 70–74.
- SELF, S. G. and PRENTICE, R. L. (1986). Incorporating random effects into multivariate relative risk regression models. In *Modern Statistical Methods in Chronic Disease Epidemiology* (S. H. Moolgavkar and R. L. Prentice, eds.) 167–177. Wiley, New York.

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802