NONPARAMETRIC INference FOR A CLASS 
OF SEMI-MARKOV PROCESSES 
WITH CENSORED OBSERVATIONS

Joseph G Voelkel and John J Crowley
NONPARAMETRIC INference FOR A CLASS OF SEMI-MARKOV PROCESSES
WITH CENSORED OBSERVATIONS\(^1\)

By Joseph G Voelkel and John J Crowley

University of Wisconsin-Madison and

Fred Hutchinson Cancer Research Center, Seattle

1. INTRODUCTION

Counting process techniques (Aalen (1975, 1978)) have proven
to be valuable tools in the nonparametric analysis of right-
censored data. The large-sample theory for the Kaplan-Meier
estimator and its associated hazard estimator (Breslow and
Crowley (1974)), for the log-rank test (Mantel (1966)), and for a
censored-data generalization of the Wilcoxon test (Gehan (1965))
have all been realized without counting process techniques;
nevertheless, such techniques provide an unifying and elegant
approach for, and lead to a deeper understanding of, such results.

---

\(^1\) Research supported in part by the Wisconsin Clinical Cancer Center
under NIH grant P30-CA-14520, R01-CA-18332 and a University of
Wisconsin WARF Fellowship.

AMS 1970 subject classifications. Primary 62G05; secondary 60K15.

Key words and phrases. Non-parametric inference, semi-Markov models,
censored observations, clinical trials.
preliminary results in Section 2. In Section 3 we define the class of semi-Markov models considered here: the class consists of finite-state semi-Markov models that (a) are extended to have (possibly stochastic) proportional hazards and (b) are restricted to be forward-going, in the sense that if state $j$ can be reached from state $i$, then state $i$ cannot be reached from state $j$. In Section 4 we introduce random time changes and show how these can be used to transform our original counting processes to fit into the multiplicative-intensity model. Finally, in Section 5 we consider a four-state semi-Markov model that has proven useful in certain clinical trials. This model, whose Markov analog was studied by Temkin (1978), assumes that from an initial state either a progression state or a response state may be entered, and from the response state a relapse state may be entered (see Figure 1). Using this model, we establish the large-sample properties for an estimator of a useful measure of a treatment's efficacy, the probability-of-being-in-response function (PBRF). All proofs are relegated to the appendix.

Aalen (1975) presented an example of a simple semi-Markov model. Although his example contains some minor errors -- his $L_1$ process is in fact not a counting process -- the idea of random time changes is inherent in his paper. Our paper formalizes this idea, extends it to a larger class of models, and uses an example to show how asymptotic theory may be employed in this larger class.
2. NOTATION AND PRELIMINARY RESULTS

We first introduce some notation and results for general multivariate counting processes and then establish the framework for the semi-Markov model.

Let \((\Omega, \mathcal{F}, P)\) be a complete probability space and let \(\{F_t\}\) for \(t \in [0, 1]\) be a history, i.e. an increasing, right-continuous family of sub-sigma-fields of \(F = F_1\). All the processes below are defined on this space and have either real-valued or vector-valued outcomes. To this end, let \(D[0, 1]\) and \(C[0, 1]\) be defined in the usual way, and let \(S[0, 1]\) be the left-handed partner of \(D[0, 1]\); i.e., \(S[0, 1]\) is the set of real-valued functions on \([0, 1]\) that are left-continuous and have right-hand limits. A stochastic process \(X(\cdot)\) with outcomes in one of these spaces is said to be a random element of that space, and, if \(X(t)\) is \(F_t\)-measurable, it is said to be adapted (to \(\{F_t\}\)). Finally, let \(\mathcal{L}\) be Lebesgue measure on \([0, 1]\).

The following definition and theorems may be found in Aalen ((1975), (1978)).

Definition 1. A stochastic process \(N = (N_1, N_2, \ldots, N_p)\)
is a multivariate counting process if

I) The sample paths of each \(N_i\) are right-continuous step functions with a finite
Frequently, \( L_i(t) \) can be written in the form \( L_i(t) = a_i(t)Y_i(t) \), all \( i \), where each \( Y_i \) is an adapted random element of \( S[0, 1] \) (and is often, from a statistician's point of view, observable), while the \( a_i \)'s, deterministic functions of \( S[0, 1] \) (usually unknown to the statistician), are underlying hazard functions associated with the counting processes. When \( L \) has such a form, the multiplicative intensity model is said to hold. As we shall see below, such a model allows us to make inferences on \( A_i(t) = \int_0^t a_i \, d\xi \) and related functions.

We examine certain integrals below, and to ensure that they are all well-defined we will assume in the sequel that positive \( Y_i(t) \) values are bounded away from 0, for all \( i \) and \( t \). Note that when \( Y_i \) has the interpretation of "number at risk" this assumption is automatically satisfied. Define \( K_i(t) = I[Y_i(t) > 0] \), where \( I \) is the standard indicator function.

**Theorem 2.** Define \( N_i \), \( L_i \), and \( M_i \) as in Theorem 1 and assume that the multiplicative intensity model holds with \( L_i = a_iY_i \) for all \( i \). Let \( H_i \), all \( i \), be adapted random elements of \( S[0, 1] \) that satisfy \( \sup_t |H_i(t)K_i(t)(Y_i(t))^{-1}| < d \), for some finite \( d \). Then

\[
\hat{B}_i(t) - B_i(t) = \int_0^t K_i \, dN_i - \int_0^t a_i \, d\xi
\]

\[
= \int_0^t K_i \, dM_i
\]
and also define the censored counting processes corresponding to the above, \( N_i(t) = \int_0^t J_i dN_i^* \). Since \( M_i^*(t) = N_i^*(t) - \int_0^t L_i^* d\lambda \) are orthogonal square-integrable martingales with respect to \( \{ F_t \} \), so are \( M_i(t) = \int_0^t J_i dM_i^* = N_i(t) - \int_0^t J_i L_i^* d\lambda \), by Theorem 2; i.e., \( N_i \) has intensity \( L_i = J_i L_i^* \). Note that in this setup \( \zeta \) is \( F_0 \)-measurable, and this can be interpreted to mean that the latent censoring times are known at time zero. Usually this is not reasonable; however, if we let

\[
\zeta_t = \sigma(N(s), Y_i^*(s) J_i(s), s \leq t, i = 1, \ldots, p),
\]

then \( N_i(t) \) has intensity \( L_i(t) \) with respect to \( \{ \zeta_t \} \), and this is the history one would want to use because it is the observable one. This argument is merely a precise way of saying that the martingale property, and hence inferences on the \( a_i(t) \), are not affected by whether or not we know the latent censoring times at time zero.

If we let \( \bar{N} = \sum_i N_i, \bar{Y} = \sum_i Y_i^* J_i, \) and \( \bar{L} = \sum_i L_i \), then, since a sum of martingale is again a martingale, \( \bar{N} \) has intensity \( \bar{L} = a \bar{Y} \). By Theorem 2,

\[
\hat{A}(t) - A^+(t) = \int_0^t \bar{Y}^{-1} I(\bar{Y} > 0) d\bar{N} - \int_0^t I(\bar{Y} > 0) ad\lambda
\]

is a martingale, so a reasonable estimator of \( A(t) = \int_0^t ad\lambda \) is \( \hat{A}(t) \). This estimator has been suggested on intuitive grounds by
3. A CLASS OF SEMI-MARKOV MODELS

We restrict the class of semi-Markov models that we will examine in three ways. First, we assume that the number of states in the model is finite. Second, we assume that the underlying distributions of the model are absolutely continuous. This allows us to view the model purely in terms of hazard functions $a_{ij}$: if $S$ and $S'$ are the (first, say) entry and exit times from state $i$ in the customary semi-Markov process, $V$ is the next state visited, and $\{G_t\}$ is the history of the process, then, in obvious notation, $a_{ij}(t) = \lim_{h \to 0} P(t < S' - S < t + h, V = j \mid S' - S > t, G_S)$.

Third, we assume that if state $j$ can be reached from state $i$, then state $i$ cannot be reached from state $j$. (This assumption, apparently needed for the martingale theory to function properly, was also made by Aalen (1975) in his attempt to model a specific semi-Markov process.) We use this assumption to simplify notation by numbering the states in the model such that an $i \to j$ transition is possible only if $i < j$. We also generalize the model, in the customary proportional-hazards manner: in a sense to be made more precise below, the $i \to j$ hazard for a particular observation in state $i$ at time $t$ is $a_{ij}(t - T_i)Z_i(t)$, where $T_i$ is the time state $i$ was entered and $Z_i$ is a stochastic process.

To put the above ideas on a more concrete footing, we now develop notation for the $k^{th}$ of $n$ observations.
Some non-negative process such that $X^o_{ik}$ is adapted and is a random element of $S[0, 1]$. In the simplest case, the one-sample semi-Markov model without censoring, $X^o_{ik}(t) = 1$ for all $i$, $k$ and $t$. More generally, it can be constructed to include a censoring scheme (as in Section 5) and it also can attach weights to the observations' hazard functions as done in the Cox model (of course, the martingale theory will hold directly only if we use the true parameters in the Cox model, not their estimators). Note that $X^o_{ik}$ may also be a function of transition times up to $T_{ik}$, e.g.,

$$X^o_{ik} = e^{-\beta T_{ik}}.$$ That the corresponding $I_{ik}X^o_{ik}$ is adapted follows along the lines of an analogous result for $I_{ijk}$.

$S_{mk}$: The time of the $m^{th}$ transition. This is also a stopping time for each $m$. If the $m^{th}$ transition has not been reached by time 1, set $S_{mk} = 1^+$. The $m^{th}$ state visited, where $V_{0k}$ is the initial state. Note that if $J(t)$ were to be defined as the state the observation is in at time $t$, then $J$ is an adapted random element of $D[0, 1]$ and $V_{mk} = J(S_{mk})$; hence (Neveu, 1965),
4. THE RANDOM TIME CHANGE AND THE TRANSFORMED COUNTING PROCESSES

Define the random time change function $\psi_k(t)$ by

$$
\psi_k(t) = t - s_{mk} + v_{mk} - 1, \quad t \in [s_{mk}, s_{m+1,k}).
$$

(3.1)

Thus, for a fixed outcome $\omega$, $\psi_k(\omega, \cdot) : [0, 1] \to [0, m']$, where $m'$ is the number of states in the model; in addition, note that $\psi_k(s_{mk}) = v_{mk} - 1$. Also define $\psi_k^{-1}(u) = \inf\{t : \psi_k(t) \geq u\}$.

This is almost a bona fide inverse, since $\psi_k^{-1} \psi_k(t) = t$, and $\psi_k \psi_k^{-1}(u) \geq u$, with equality if and only if the right-hand derivative of $\psi_k^{-1}(u)$ is 1. It also follows from the definition of $\psi_k^{-1}$ that

$$
\psi_k^{-1}(u) \leq t \iff \psi_k(t) \geq u.
$$

(3.2)

and that
1 \to 2 and 2 \to 4 transitions occurred at times .1 and .5, respectively. Then the transition number index m assumes the values 0, 1, and 2, the respective values of S_{mk} are 0(T_{1k}), .1(T_{2k}), and .5(T_{4k}), and the respective values of V_{mk} are 1, 2, and 4. The corresponding values of \psi_k and \psi_k^{-1} are graphed in Figure 2.

We now apply the random time change to the counting processes and then derive the corresponding intensity processes.

These new intensity processes are in the form of a multiplicative-intensity model and also conform to one's intuition.

Define N_{ijk}^o(t) = N_{ijk}^o(t) - \int_0^t L_{ijk}^o \, dt, all (i,j); these are square-integrable orthogonal martingales with respect to \{F_{kt}\}.

Theorem 3. For u \in [0, m'],

\[ M_{ijk}^*(u) = M_{ijk}^o(\psi_k^{-1}(u)), \quad \text{all (i, j)}, \]

are square-integrable orthogonal martingales with respect to \{F^*_{ku}\}, where \[ F^*_{ku} = F_{ku}^{\psi_k^{-1}}(u) \]. Also,

\[ M_{ijk}^*(u) = M_{ijk}^o(u) - \int_0^u a_{ij}^* Y_{ik} \, dt, \]
where

\[ N_{ijk}^*(u) = I\{u \geq i - 1 + T_{ik}' - T_{ik}\} I\{N_{ijk}^0(1) = 1\}, \]

\[ a_{i1}^*(s) = a_{i1}'(s - (i - 1)), \]

and

\[ Y_{ik}^*(s) = X_{ik}^0(s + T_{ik}' - (i-1)) I\{i-1 < s \leq i-1 + T_{ik}' - T_{ik}\}. \]

In particular, \(N_{ijk}^*\) is a counting process with intensity process \(L_{ijk}^* = a_{i1}^* Y_{ik}^*\) with respect to \(\mathcal{F}_{ku}^*: \ u \in [0, m']\).

We now introduce a bit more notation, most of which we use repeatedly in the sequel. Let \(N_{ij}(t) = \sum_k N_{ijk}^*(t + i-1)\), the number of \(i + j\) transitions whose transition times from \(i\) to \(j\) (i.e., amount of time spent in \(i\) before going to \(j\)) is less than or equal to \(t\) -- here we begin to let "\(t\)" represent time other than real time. Also, let \(Y_{i}(t) = \sum_k Y_{ik}^*(t + i - 1)\). In the one-sample case with right-censoring, this is the number of observations at risk in state \(i\) just before time \(t\), where the time is "local" to state \(i\). Finally, define \(L_{ij} = a_{i1} Y_{i1}^*\), \(M_{ij}^* = \sum_k M_{ijk}^*\), \(F^* = \sum_k F_{kt}^*\), \(\mathcal{F}^* = \mathcal{V}\), and \(N_{ij}(t) = M_{ij}^*(t + i-1) = N_{ij}(t) - \int_0^t L_{ij} d\mathcal{F}^*\). (To fix ideas, we continue the example where \(1 \rightarrow 2\) transitions were made at \(.2, .3,\) and \(.6,\) and suppose the corresponding \(2 \rightarrow 4\) transitions were made at \(.9, .8,\) and \(1.0\). Then the amounts of time spent in state \(2\) are \(.7, .5,\) and \(.4,\) so \(N_{24}(.3) = 0, N_{24}(.6) = 2, Y_{2}(.3) = 3, Y_{2}(.6) = 1,\)
implies that observation 1 will make a $2 \rightarrow 4$ transition. Our modelling of these processes does allow us to let the observations' histories affect each other in the transformed time without losing the martingale structure of the $\hat{M}_{ij}^*$'s, but from a practical point this would be absurd.

Forcing the semi-Markov processes into the multiplicative-intensity model allows us to apply the corresponding asymptotic theory (e.g. Aalen and Johansen (1978)) after suitable modifications. The most substantial change we make to such asymptotic theory arises because our integrals may be of the form

$$\int_0^T H^n(t, s) dM^n(s)$$

where $\{M^n\}$ is a sequence of martingales corresponding to counting processes and $\{h^n\}$ is a sequence of random functions. The usual asymptotic theory assumes

$$H^n(t, s) = h^n(s)$$

is $F^s_n$-adapted, while in our situation it is not. Incidentally, it is this same reason that has precluded us from ascertaining any small-sample properties of the estimator we present in the next section.
5. THE PBRF AND ITS LARGE-SAMPLE PROPERTIES

A reasonable model in cancer clinical trials assumes that each patient may either remain in an initial state, or progress, or respond and then possible relapse — see Figure 1. Temkin (1978) realized the use of such a model. She first pointed out that when two treatments are compared, the two most commonly used measures of a treatment's efficacy, probability of responding and time to progression or relapse, may yield contradictory information. A more cohesive measure of a treatment's ability, she argued, is the PBRF, viewed as a function of time. Assuming a Markov model, she developed an estimator of the PBRF in the presence of right-censored data and estimated its large-sample variance in the manner of Lagakos, Sommer and Zelen, by formally assuming only a finite number of transition times. At the same time, Aalen and Johansen (1978) developed both small- and large-sample theory for the finite-state Markov model in a more rigorous framework; in particular, this theory can be applied to Temkin's estimator.

In this section, we present an estimator of the PBRF in the semi-Markov one-sample setting with censored data and derive its large-sample distributional properties. The semi-Markov PBRF at time \( t \) is \( R(t) = \int_0^t F_1(s)a_{12}(s)F_2(t-s)ds \), where \( F_i \), in non-standard notation, is the survival function corresponding to state \( i \). Note that this equals its Markov counterpart,
Theorem 4. Let \( 0 < c < 1 \) and assume for \( t \in [0, c] \)

I) \( Y_i(t)/n = p_i(t) + o_p(1) \) for \( i = 1, 2 \), where the \( p_i \)
are deterministic functions bounded away from 0.

II) \( n K_i(t) Y_i^{-1}(t) \) is uniformly integrable in \((n, t, i)\).

III) \( n^{1/2} \int_0^1 (1 - K_i) a_{ij} d\xi = o_p(1) \) for all \((i, j)\).

Then \( n^{1/2}(R - R) \overset{d}{\longrightarrow} Z \) on \([0, c]\), where

\[ Z(t) = \int_0^t g_{12}(t, \cdot) dW_{12} \]

\[ - \int_0^t [\int_0^{t-s} F_1 F_2; a_{12} d\xi] h_{24}(s) dW_{24}(s) \]

\[ - \int_0^t [\int_0^{t-s} F_1 F_2; a_{12} d\xi] (h_{12}(s) dW_{12}(s) + \]

\[ h_{13}(s) dW_{13}(s)), \]

the \( W_{ij} \) are independent Wiener processes, and

\[ g_{12}(t, s) = F_1(s) F_2(t - s) (a_{12}(s)/p_1(s))^{1/2} \]

\[ h_{12}(s) = (a_{12}(s)/p_1(s))^{1/2} \]

\[ h_{13}(s) = (a_{13}(s)/p_1(s))^{1/2} \]

\[ h_{24}(s) = (a_{24}(s)/p_2(s))^{1/2}. \]

Note that we restrict our attention to \([0, c]\). In our setup
we only observe \( 2 \to 4 \) transitions in real time \([0, 1]\), so

\( Y_2(1) = 0 \) a.e. \([P]\), for each \( n \), assuming that each observation is
in the initial state at time zero. Thus we must choose \( c < 1 \) so
that assumption I of the theorem holds.
Corollary 2. Let $w: [0, 1] \rightarrow \mathbb{R}$ be bounded and measurable. Under the conditions of Theorem 4, $n^{1/2} \int_0^t \langle R - R \rangle w d \xi \rightarrow^d Z_1(t)$ as a process in $t$, where the Gaussian process $Z_1$ is $Z_1(t) = \int_0^t Z w d \xi$.

The covariance of $Z_1(u)$ and $Z_1(v)$, for $0 \leq u \leq v \leq c$, is

$$
\begin{align*}
\int_0^u \left[ \int_0^u (F_1(s) - F_2(t-s) - I(s,t,t))w(t)dt \right] \times \\
\left[ \int_s^v (F_1(s) - F_2(t-s) - I(s,t,t))w(t)dt \right] (a_{12}(s)/p_1(s))ds \\
+ \int_0^u \left[ \int_0^u I(0,t-s,t)dt \right] \left[ \int_s^v I(0,t-s,t)dt \right] (a_{24}(s)/p_2(s))ds \\
+ \int_0^u \left[ \int_s^u I(s,t,t)dt \right] \left[ \int_s^v I(s,t,t)dt \right] (a_{13}(s)/p_1(s))ds,
\end{align*}
$$

where $I(x,y,z) = \int_x^y F_1 F_2 z^{a_{12}} d \xi$.

A suitable estimate of the above covariance can be determined along the lines of Corollary 1.

A graph of $t \rightarrow \int_0^t \langle R \rangle d \xi$ provides another way to examine the relative efficacies of drugs; however, for large $t$ the graph may be misleading, for the same reason that the expected value of a random variable may be a poor measure of center for its corresponding distribution.

Under what situations will the conditions of Corollary 1 hold, so that we can set confidence intervals on $R(t)$, perform two-sample tests, etc? We restrict ourselves to the one-sample random censorship model, and whether the above conditions are met simply depends on the type of censoring in effect.
APPENDIX

Proof of Theorem 3. That the $M^*_\text{ijk}$ have the stated martingale properties follows from Doob (1953). To write $M^*_\text{ijk}$ in terms of $N^*_\text{ijk}$, $a^*_\text{ij}$, and $Y^*_\text{ik}$, we first write

$$M^*_\text{ijk}(u) = N^*_\text{ijk}(\psi^{-1}_k(u))$$

$$= N^*_\text{ijk}(\psi^{-1}_k(u)) - \int_0^{\psi^{-1}_k(u)} a^*_\text{ij}(s - T^*_{\text{ik}}) Y^*_\text{ik}(s) ds$$

$$= N^*_\text{ijk}(\psi^{-1}_k(u)) - \int_{\psi^{-1}_k(0)}^{\psi^{-1}_k(u)} a^*_\text{ij}(\psi^{-1}_k(s) - T^*_{\text{ik}}) \times$$

$$Y^*_\text{ik}(\psi^{-1}_k(s)) d\psi^{-1}_k(s).$$

Since $\psi^{-1}_k$ has absolutely continuous sample paths, then $\psi^{-1}_k'$ exists a.e. for all sample paths. In fact, let us mean by $\psi^{-1}_k'$ the left-hand derivative of $\psi^{-1}_k$. This has sample paths in $S[0, m']$, so we transform the integral above to get

$$M^*_\text{ijk}(u) = N^*_\text{ijk}(\psi^{-1}_k(u)) - \int_{\psi^{-1}_k(0)}^{u} a^*_\text{ij}(\psi^{-1}_k(s) - T^*_{\text{ik}}) Y^*_\text{ik}(\psi^{-1}_k(s)) \psi^{-1}_k'(s) ds.$$

Note that the integrand is 0 for all values between $u$ and $\psi^{-1}_k(u)$, so the upper limit of integration is correct even though $u$ and $\psi^{-1}_k(u)$ may differ.

Define $N^*_\text{ijk}(u) = N^*_\text{ijk}(\psi^{-1}_k(u))$. Since $N^*_\text{ijk}$ has at most one jump, which can only occur at $T^*_{\text{ik}}$, and since the smallest $u$ for which
Proof of Theorem 4. For real numbers and random variables, define \( \sigma, \sigma_p, \) and \( \sigma_p \) in the usual way; for functions and stochastic processes define them to hold uniformly in \( t \). To avoid writing dummy arguments in the integrals that follow, let \( f_+ (s) = f(t-s) \) and \( f_- (s) = f(s^-) \) for suitable functions or stochastic processes \( f \).

Define

\[
\hat{R}_2(t) = \int_0^t \hat{F}_1 \cdot \hat{F}_2; t \cdot K_1 \cdot Y_1^{-1} dN_{12}
\]

\[
\hat{R}_{12}(t) = \int_0^t \hat{F}_1 \cdot \hat{F}_2; t \cdot K_1 \cdot Y_1^{-1} dN_{12}
\]

\[
R^+(t) = \int_0^t \hat{F}_1 \cdot \hat{F}_2; t \cdot K_1 a_{12} d\ell
\]

\[
\hat{R}_2^+(t) = \int_0^t \hat{F}_1 \cdot \hat{F}_2; t \cdot K_1 a_{12} d\ell
\]

\[
\hat{R}_{12}^+(t) = \int_0^t \hat{F}_1 \cdot \hat{F}_2; t \cdot K_1 a_{12} d\ell
\]

and view \( \hat{R} - R \) as a sum of five terms: \( \hat{R} - R^+ - \hat{R}_2 + R_2^+ \); \( \hat{R}_2 - R_2^+ \); \( R^+ - R_2^+ \); \( \hat{R}_2^+ - R_{12}^+ \); and \( R_{12}^+ - R \). The first and last of these terms are \( o_p(n^{-1/2}) \) under reasonable conditions.

Lemma 1. Assume that for \( t \in [0, c] \), where \( c < 1 \),

I) \( E[K_1(t)Y_1(t)^{-1}] + E[1 - K_1(t)] = o(1) \) for each \( (i, t) \).

II) \( Y_1(t)/n = p_1(t) + o_p(1) \) for each \( t \), and \( p_1 \) is bounded away from 0.

III) \( nK_1(t)Y_1(t)^{-1} \) is uniformly integrable in \( (n, t) \).

IV) \( n^{1/2} \int_0^1 (1 - K_1)a_{12} d\ell = o_p(1) \).

Then \( \hat{R} - R^+ - \hat{R}_2 + R_2^+ \) and \( R_{12}^+ - R \) are both \( o_p(n^{-1/2}) \).
This suggests examining the weak convergence of

\[ Z_n(t) = (\int_0^t G_{12} \, dM_{12}, \int_0^t H_{12} \, dM_{12}, \int_0^t H_{13} \, dM_{13}, \int_0^t H_{24} \, dM_{24}), \quad (A.1) \]

where the \( G \) and \( H \)'s are certain stochastic processes (depending on \( n \), of course), and then, by applying the delta method and making some linear transformations, arriving at the weak-convergence result we seek.

First we state the following lemmas, the first of which is proven in Chung (1968, p 90), and the second of which is a Fubini theorem for stochastic integrals (proof omitted).

**Lemma 2.** Let \( X_n \) be a sequence of uniformly integrable random variables which converges in probability to a constant \( d \).

Then \( E[X_n] = d + o(1) \).

**Lemma 3.** For each \( t \in [0, 1] \), let \( h(\cdot, \cdot, t):[0,1]^2 \to (-\infty, \infty) \) be square integrable with respect to Lebesgue measure on \([0, 1]^2\).

If \( W \) is a Wiener process on \([0, 1]\), then

\[
\int_0^t \int_0^t h(u, s, t) \, dW(u) \, ds = \int_0^t \int_0^t h(u, s, t) \, ds \, dW(u).
\]

The processes we associate with (A.1) are

\[
G_{12}(t, s) = n^{1/2} \mathcal{F}_1(s-) \mathcal{F}_2(t-s) K_1(s) Y_1^{-1}(s),
\]

\[
H_{12} = n^{1/2} K_1 Y_1^{-1}, \quad H_{13} = n^{1/2} K_1 Y_1^{-1}, \quad \text{and} \quad H_{24} = n^{1/2} K_2 Y_2^{-1}.
\]
\[ Z_n' (t) \xrightarrow{d} (\Sigma a_i \int_0^t g_{12}(u_i, \cdot) \, dW_{12} + \sum_{b=1}^v \int_0^t h_{12} \, dW_{12} + \int_0^t h_{13} \, dW_{13} + \int_0^t h_{24} \, dW_{24}). \]

By the Cramer-Wold argument, the finite-dimensional distributions of \( Z_n \) converge weakly to the finite-dimensional distributions asserted by the theorem. But \( Z_n \) is also tight. Its last three coordinates clearly are because, by the above, they converge weakly to \( (\int h_{12} \, dW_{12}, \int h_{13} \, dW_{13}, \int h_{24} \, dW_{24}) \). Now let \( G_{12}'(s) = G_{12}(t, s) / F_{21}(s) \),

\[ M_{12}'(t) = \int_0^t G_{12}' \, dM_{12}. \]

By the change of variables formula, the first coordinate is \( F_2(0)M_{12}'(t) - F_2(t)M_{12}'(0) - \int_0^t M_{12}' \, dF_2; t \). The conditions above insure that \( M_{12}' \) converges weakly, so the continuous mapping theorem implies that the first coordinate also converges weakly, and is hence tight. \( \square \)

**Lemma 5.** If the conditions of Lemma 4 hold and if

\[ n^{1/2}\int_0^1 (1 - K_i) a_{ij} \, d\mathbb{L} = o_p(1) \text{ for each } (i, j), \]

then, as processes in \( t \in [0, c] \),

\[ n^{1/2}(R_2(t) - R_2^+(t), F_1(t) - F_1(t), F_2(t) - F_2(t)) \xrightarrow{d} (\int_0^t g_{12}(r, \cdot) \, dW_{12}, - F_1(t) \int_0^t h_{12} \, dW_{12} + h_{13} \, dW_{13},
\]

\[ - F_2(t) \int_0^t h_{24} \, dW_{24}). \]
The Kaplan-Meier estimator of \( F \), say \( \hat{F}^{KM} \), can replace the \( \hat{F} \) because \( \hat{F}^{KM} - \hat{F} = o_p(n^{-1/2}) \), which in turn can be proven from the relation \( 0 < \hat{F}(t) - \hat{F}^{KM} \leq \int_o^t Y^{-2} dN \), which holds as long as the non-zero values of \( Y(t) \) are at least unity.

**Proof of Corollary 1.** The first part of the corollary follows directly from Theorem 4, so we only need to show the convergence of these terms to their counterparts in Corollary 8. By using the submartingale inequality and Lemma 2 of the Appendix, conditions (I) and (II) insure that the terms in the brackets converge uniformly to the correct terms. It thus suffices to show that

\[
\sup_t \left| \int_o^t K_i Y^{-2} n_dN_{ij} - \int_o^t \left( a_{ij}/p_i \right) d\ell \right| = o_p(1).
\]

This difference can be written as

\[
\int_o^t K_i Y^{-2} n_dM_{ij} + \int_o^t \left( K_i Y^{-1} n - p_i^{-1} \right) a_{ij} d\ell.
\]

By Theorem 2, the variance of the first term is less than \( \int_o^1 E[K_i Y^{-3} n^2] a_{ij} d\ell \). By using (I) and (II) with Lemma 2 and then applying the Dominated Convergence Theorem, this bound on the variance converges to 0, so the first term is \( o_p(1) \) uniformly in \( t \). That the second term is also \( o_p(1) \) uniformly in \( t \) follows from (I), (II), and Lemma 2. \( \square \)
REFERENCES


Proof of Corollary 3. If suffices to show that (I) of Theorem 4 and the added condition of corollary 1 are satisfied, since these imply (II) and (III) of Theorem 4. Let \( C \) be the survival function associated with each \( C_k \). Then the distribution of \( Y_1(t) \) is binomial with parameters \( n \) and \( F_1(t)G(t) \), so \( Y_1(t)/n = F_1(t)G(t) + o_p(1) \) and, by assumption, \( F_1(t)G(t) \) is bounded away from 0 for \( 0 \leq t \leq 1 \); thus, \( p_1(t) \) is identified.

The distribution of \( Y_2(t) \) is also binomial, with parameters \( n \) and \( \int_0^1 F_1(s) \alpha_{12}(s) F_2(t) G(t+s) \, ds \). By the condition made on \( \alpha_{12} \), this integral is bounded away from 0 for all \( t \) in \([0, c]\), so \( Y_2(t)/n = p_2(t) + o_p(1) \). Thus (I) of Theorem 4 is satisfied.

Now let the family of random variables \( \{X_n(t)\} \), \( t \in [0, c] \), \( n = 1, 2, \ldots \) have distributions

\[ X_n(t) \sim \text{Binomial}(n, p(t)) \] where \( \inf_{t \in [0, c]} p(t) = p_0 > 0 \).

To prove the added condition of Corollary 1 holds, it is enough to show that \( E[n^3 I\{X_n(t) > 0\} X_n^{-3}(t)] \) is bounded uniformly in \( (n, t) \) (eg, Neveu (1965), p 54). That this expected value is indeed uniformly bounded follows by an extension of an idea, used by Aalen (1976), which shows that a uniform bound is

\[ p_0^{-3} + 115p_0^{-4}. \]
Proof: The condition added in this corollary insures that we may replace $A_{1j}^+$ in $Z_n$ by $A_{1j}$. Two applications of the delta method yield the desired result after the second and third processes of $Z_n$ have been summed. □

Lemma 6. If the conditions of Lemma 5 hold, then

$$n^{1/2}(R_2(t) - R_2^+(t), R_2^+(t) - R_2^+(t), R_2^+(t) - R_{12}^+(t)) \xrightarrow{d} \\
(j^{t}_{\delta} g_{12}(t, \cdot) d\tilde{W}_{12}, - j^{t}_{\delta} F_1(s) F_2(t - s) [\int_{0}^{t - s} h_{24} d\tilde{W}_{24}] \times \\
a_{12}(s) ds, - \int_{0}^{t} F_1(s) X(s) S_2(t - s) a_{12}(s) ds)$$

as processes in $t$, where $X(s) = \int_{0}^{s} h_{12} d\tilde{W}_{12} + \int_{0}^{s} h_{13} d\tilde{W}_{13}$.

Proof: The conclusion follows by noting that the $K_1$ terms become 1 in the limit, by summing the second and third processes of $Z_n$, by applying the delta method several times, (eg, $\hat{A}_{24} - A_{24}$ to $F_2 - F_2$), by noting that $\hat{F}_1 - F_1 + o_p(1)$, and by employing some continuous maps (eg, $\hat{F}_2 - F_2$ to $R_2^+ - R_{12}^+$). □

The main assertion of Theorem 4 now follows: the sum of the three terms of Lemma 6 is, by Lemma 1, asymptotically equivalent to $n^{1/2}(\hat{F} - F)$, noting that (I) of Lemma 1 is implied by (I) and (II) of this theorem via Lemma 2. That the order of integration may be reversed follows from Lemma 3. □
Then $Z_n = n^{1/2}(\hat{R}_2 - R_2^+, \hat{A}_{12} - A_{12}^+, \hat{A}_{13} - A_{13}^+, \hat{A}_{24} - A_{24}^+)$,

where $A_{ij}^+(t) = \int_0^t K_{1j} a_{ij} d\ell$.

In what follows, let $p_1(t)$ and $p_2(t)$ be two functions that are bounded away from 0 on $[0, c]$, where $c < 1$.

**Lemma 4.** Assume that for $t \in [0, c]$

I. $Y_i(t)/n = p_i(t) + o_p(1)$ for $i = 1, 2$.

II. $nK_i(t)Y_i(t)^{-1}$ is uniformly integrable in $(n, t, i)$.

Then on $[0, c]$

$$Z_n(t) \xrightarrow{d} (\int_0^t g_{12}(t, \cdot) dW_{12}, \int_0^t h_{12} dW_{12}, \int_0^t h_{13} dW_{13}, \int_0^t h_{24} dW_{24}),$$

as a process in $t$, where the $W_{ij}, g_{12}, h_{12}, h_{13}$ and $h_{24}$ are defined in Theorem 4.

**Proof:** The result would be an application of Aalen and Johansen's (1978) Theorem 4.1 except that $G_{12}(t, s)$ is a function of $t$ as well as $s$ and that two of the integrals in $Z_n$ are taken with respect to the same martingale. To circumvent these problems, consider instead the weak convergence of $Z'_n(t) = (\int_0^t H(a, u, b, v, s) dM_{12}(s), \int_0^t H_{13} dM_{13}, \int_0^t H_{24} dM_{24}),$

where $H(a, u, b, v, s) = \sum_i I\{s < u_i\} G_{12}(u_i, s) + \sum_j I\{s < v_j\} H_{12}(s),$

the $a_i$ and $b_i$ are arbitrary real numbers, the $u_i$ and $v_j$ are in $[0, 1]$, and the sums are finite. Then the conditions of Aalen and Johansen's theorem are satisfied for $Z'_n$, and, as a sequence of processes,
Proof: The first assertion follows because
\[ n^{1/2} \left( \hat{R}_2(t) - R_2^+(t) - \hat{R}_2(t) + R_2^+(t) \right) \leq \]
\[ \left( \sup_{0 \leq u \leq 1} |\hat{F}_2(u) - F_2(u)| \right) n^{1/2} \int_0^t \int_0^{F_1^{-1}} \int_0^{\hat{Y}_1^{-1}} dM_{12}. \]

Since the submartingale inequality (e.g., Doob (1953), p 317) and (I) insure the uniform consistency of \( \hat{F}_2 \) to \( F_2 \), the first term on the right-hand side is \( o_p(1) \); by Theorem 2, the variance of the second term is bounded by \( n^{1/2} \int_0^t E[K_1^{-1}]a_{12} d\lambda \), which in turn is uniformly bounded in \( (n, t) \), by (III), so the entire term is \( o_p(1) \) uniformly in \( t \). The left-hand side of this inequality can likewise be bounded from below by a \( o_p(1) \) term.

The second assertion clearly holds. \( \square \)

What remains is to find the joint asymptotic distribution of
\[ n^{1/2} (\hat{R}_2 - R_2^+, \hat{R}_2^+, R_2^+, \hat{R}_2^+, R_2^+ - R_2^+). \]
Note that
\[ \hat{R}_2(t) - R_2^+(t) = \int_0^t \int_0^{\hat{F}_2(t)} K_1^{-1} dM_{12}, \]
the integrand of \( \hat{R}_2^+(t) - R_2^+(t) \)
contains the term \( \hat{F}_2(t) - F_2(t) = \exp(- \hat{A}_{24}(t-s)) - \exp(- A_{24}(t-s)) \),
and the integrand of \( R_2^+(t) - R_2^+(t) \) contains the term
\[ \hat{F}_1^{-}(s) - F_1^{-}(s) = \exp(- \hat{A}_{12}(s^-) - \hat{A}_{13}(s^-)) - \exp(- A_{12}(s) - A_{13}(s)). \]
\( \psi_{ik}^{-1}(u) = T'_{ik} \) is \( u = i - 1 + T'_{ik} - T_{ik} \), then

\[
N_{ijk}^*(u) = I(u > i - 1 + T'_{ik} - T_{ik}) I(N_{ijk}^O(1) = 1).
\]

Let \( Y_{ik}^*(s) = Y_{ik}^O(\psi_{ik}^{-1}(s)) \psi_{k}^{-1}(s) \). Now \( Y_{ik}^O(\psi_{k}^{-1}(s)) \) is non-zero only when \( \psi_{k}^{-1}(s) \in (T_{ik}, T'_{ik}] \), or when \( s \) is an element of

\[
(i-1, i-1 + T'_{ik} - T_{ik}] \cup (i-1+T'_{ik} - T_{ik}, i] = (i-1, i].
\]

Also \( \psi_{k}^{-1}(s) = 1 \) if and only if \( s \in \bigcup_i (i-1, i-1 + T'_{ik} - T_{ik}] \), where we interpret \( (i - 1, i - 1 + T'_{ik} - T_{ik}] \) as \( (i - 1, i - T_{ik}] \) if state \( i \) is entered but no \( i \rightarrow j \) transition occurs, and as the null set if state \( i \) is never entered. Thus, \( Y_{ik}^*(s) \) is non-zero only when \( s \in (i - 1, i - 1 + T'_{ik} - T_{ik}] \) and on this interval equals \( X_{ik}^O(s + T_{ik} - i + 1) \).

We now simplify \( a_{ij}(\psi_{k}^{-1}(s) - T_{ik}) \), first noting that the term need be examined only for \( s \) such that \( Y_{ik}^*(s) \) is non-zero. For such values of \( s \), those in \( (i - 1, i - 1 + T'_{ik} - T_{ik}] \), we have

\[
a_{ij}(\psi_{k}^{-1}(s) - T_{ik}) = a_{ij}(s + T_{ik} - i + 1 - T_{ik})
= a_{ij}(s - (i - 1)).
\]

With this in mind, we define \( a_{ij}^*(s) = a_{ij}(s - i + 1) \).

That the \( N_{ijk}^* \) are counting processes with intensities \( L_{ijk}^* \) is immediate. □
\[ \int_0^t F_1(s) \tilde{a}_{12}(s)(F_2(t)/F_2(s))ds \], whenever \( a_{24}(t) \) is a constant function. (The case in which all the hazard functions are constants, the homogeneous Markov case, was investigated by Begg and Larson (1980).)

Now define \( \hat{F}_1 = \exp(-\hat{A}_{12} - \hat{A}_{13}) \) and \( \hat{F}_2 = \exp(-\hat{A}_{24}) \), where \( \hat{A}_{ij}(t) = \int_0^t K_{ij}Y_i dN_{ij} \), and \( K_{i}(t) = I\{Y_i(t) > 0\} \). We use the \( \hat{F}_1 \)'s for the large-sample derivations because they exhibit the correct large-sample properties in more general settings than the one on which we concentrate here; however, in the actual one-sample setting, we recommend that the corresponding Kaplan-Meier (1958) estimators be used instead. (Such an interchange has no effect on the large-sample results.)

Also define our estimator of \( R(t) \),

\[ \hat{R}(t) = \int_0^t \hat{F}_1(s^-) \hat{F}_2(t - s) K_1(s)(Y_1(s))^{-1} dN_{12}(s). \]

Note that although all these estimators are functions of \( n \), the number of observations, we leave this relation implicit for readability.
Figure 3. Sample paths for two observations.

Observation 1
Real Time: ————
Randomly Changed Time: ————

Observation 2
Real Time: ————
Randomly Changed Time: ————

Note: Observation 2 is censored at time 0.6
Figure 2. Values of $\psi_k$ and $\psi_k^{-1}$
\[
\psi_k^{-1}(u) = \begin{cases} 
0, & u \in [0, V_{0k} - 1] \\
u + S_{mk} - V_{mk} + 1, & u \in (V_{mk} - 1, V_{mk} - 1 + S_{m+1,k} - S_{mk}] \quad (3.3) \\
S_{m+1,k} & u \in (V_{mk} - 1 + S_{m+1,k} - S_{mk}, V_{m+1,k} - 1], \\
\end{cases}
\]

The relationship (3.2) implies

\[
\{\psi_k^{-1}(u) \leq t\} = \bigcup_{m} \{t - S_{mk} + V_{mk} - 1 \geq u\} \cap \\
\{S_{m+1,k} > t\} \cap \{S_{mk} < t\},
\]

\[
= \bigcup_{m} \bigcup_{i} S_{mk} \leq t - u + i - 1 \cap \{V_{mk} = i\} \cap \{S_{m+1,k} > t\} \cap \\
\{S_{mk} < t\}.
\]

Now \(S_{mk}\) is a stopping time for each \(m\), and \(V_{mk}\) is \(F_{kS_{mk}}\) -measurable, so \(\psi_k^{-1}(u)\) is a stopping time for each \(u\). Also, using (3.2) again, \(\psi_k(t)\) is \(F_{kt}\) -measurable, and it is also a random element of \(D[0, 1]\) because of its sample-path properties.

Before proceeding further we give a simple example. Using the state space in Figure 1, assume for a particular outcome that
Define the following:

\[ N_{ijk}^O \]: A counting process which counts the actual number (i.e., in the presence of censoring) of transitions from state \( i \) to state \( j \). We assume that the histories generated by observations, denoted by \( \{ \mathcal{F}_t : t \in [0, 1] \} \), are independent so, e.g., information on one observation cannot be used to censor another observation. We show below that this restriction is needed to preserve the martingale property in the randomly changed time.

\[ T_{ik} \]: The time at which state \( i \) is entered. This is a stopping time for each \( i \).

\[ T'_{ik} \]: The time at which state \( i \) is left, also a stopping time for each \( i \).

\[ L_{ijk}^O \]: The intensity process of \( N_{ijk}^O \). Assume the \( L_{ijk}^O \) have the form \( L_{ijk}^O(t) = a_{ij}(t-T_{ik})y_{ik}^O(t) \), where \( a_{ij}(t) \) is the underlying hazard function of the \( i \to j \) transition (and hence is in \( S[0, 1] \)), and \( y_{ik}^O(t) = I_{ik}(t)x_{ik}^O(t) \) is defined via

\[ I_{ik} \]: The indicator function for state \( i \), defined so that its sample paths are in \( S[0, 1] \). Note that \( I_{ik} \) is adapted.
Nelson (1969). In the case where the latent censoring variables $C_i$ are i.i.d. with survival function $G$, it follows that
\[ E(\hat{A}(t)) = A(t) - \int_0^t a(s)(1 - F(s)G(s))^p ds. \]
That the expected value of the estimator converges to $A(t)$ at this exponential rate (when $G(t) > 0$) was pointed out by Aalen (1976) in the context of no censoring.
are orthogonal square-integrable martingales, and
\[
(B_i(t) - B_i^+(t))^2 - \int_0^t a_i H_i^2 K_i Y_i^{-1} d\ell
\]
are martingales.

The conditions placed on the $H_i$ allow the stochastic integrals
\[
\int_0^t K_i H_i^{-1} dM_i
\]
to be interpreted as Lebesgue-Stieltjes integrals
(Doléans - Dade and Meyer (1970)).

**An Example of the Multiplicative-Intensity Model**

Let $U_1, U_2, \ldots, U_p$ be non-negative i.i.d. random variables on $(\Omega, F, P)$ and have survival function $F$, where $P(t) > 0$ for $t \leq 1$ and the hazard function $a(t) = -(dF(t)/d\ell)F(t)^{-1}$ exists and is in $S[0, 1]$. Let $N_i^*(t) = I\{U_i \leq t\}$ and $Y_i^*(t) = I\{U_i > t\}$ be adapted. It is easy to show that each $N_i^*$ is a counting process and that each $L_i^* = aY_i^*$ satisfies the conditions of Theorem 1; hence, $L_i^*(t)$ is the intensity process of $N_i^*(t)$ with respect to $\{F_t\}$.

This choice of $L_i^*$ is intuitively appealing: if $U_i$ is the death time of the $i^{th}$ patient in a clinical trial, the above says that the individual is subjected to the hazard $a(t)$ of dying until death itself occurs, at which point the hazard ceases. To include a simple form of censoring in the above scheme, let $C = (C_1, \ldots, C_p)$ be $F_0$-measurable random variables such that $C$ is independent of $U_1, \ldots, U_p$. The $C_i$ are to be considered latent censoring times: to this end, define the censoring processes $J_i$ by $J_i(t) = I\{C_i \geq t\}$,
number of jumps, each positive
and of size 1, and \( N_1(0) = 0 \).

II) Two component processes, \( N_i \) and \( N_j \),
i \( \neq j \), cannot jump at the same time.

III) \( \sim \) is adapted to \( \{ F_t \} \).

Let \( S_1 < S_2 < \ldots \) be the jump times of \( \sum N_i \) -- note that the
\( S_m \) are stopping times -- and let \( V_m = i \) if \( N_i \) jumps at \( S_m \). In
the sequel it will always hold that \( \mathbb{E}(N_i(\cdot)) < \infty \) for each \( i \),
and that \( P(S_{m+1} - S_m \leq t, V_{m+1} = i | F_{S(m)}) \) is absolutely continuous
in \( t \) for all \( m \), and has derivatives in \( S[0, 1] \). Under these
conditions, one can prove the following theorem.

Theorem 1. There exists a unique (up to equivalence) non-
negative adapted process \( L = (L_1, L_2, \ldots, L_p) \) such that each \( L_i \)
is a random element of \( S[0, 1] \) and \( M_i(t) = N_i(t) - \int_0^t L_i \, d\xi \) are
square-integrable martingales. In addition, \( M_i \) and \( M_j \) for \( i \neq j \)
are orthogonal (i.e., their product is a martingale) and
\( M_i^2(t) - \int_0^t L_i \, d\xi \) are martingales.

The process \( L_i \), called the intensity process of \( N \) with respect
to \( \{ F_t \} \), can be viewed as a conditional hazard function, since one
can show that \( L_i(t^+) = \lim_{h \downarrow 0} h^{-1} P(N_1(t + h) - N_1(t) | F_t) \), whenever
\( L_i \) is bounded by an integrable random variable.
Figure 1. State Space of a Model Used in Clinical Trials
In addition, these techniques have also been used recently in their own right to analyze censored data. These include Fleming's, et al., modification of the Kolmogorov-Smirnov test (1980), as well as a family of two-sample tests (1979); Aalen's (1978) development of a regression model that complements the proportional-hazard one proposed by Cox (1972); and Aalen and Johanssen's (1978) study of the (nonhomogeneous) finite-state Markov model. An excellent review of counting process techniques appears in Andersen, Borgan, Gill, and Keiding (1981).

All the above uses of counting-process theory rely on the multiplicative-intensity model for making inferences: roughly speaking, if $N(\cdot)$ is a counting process with intensity $L(\cdot)$, the multiplicative-intensity model is said to hold if $L(\cdot) = Y(\cdot) a(\cdot)$, where $Y(\cdot)$ is a stochastic process and $a(\cdot)$ is a hazard function for which inferences are sought. For example, in the one-sample case $N(t)$ is the number of events up to time $t$, $Y(t)$ is the number at risk at $t$, and $a(t)$ is the hazard associated with the distribution in question. In the Markov model mentioned above, $N = N_{ij}$ counts the $i \rightarrow j$ (state $i$ to state $j$) transitions, $Y = Y_i$ counts the number in state $i$, and $a = a_{ij}$ is the $i \rightarrow j$ transition intensity.

The semi-Markov model does not readily fit into the multiplicative-intensity framework, precisely because of its renewal nature. In order to circumvent this, we first establish some notation and
SUMMARY

Nonparametric Inference for a Class of Semi-Markov Processes

A class of semi-Markov models, those which have proportional hazards and which are forward-going (if state j can be reached from i, then i cannot be reached from j), are shown to fit into the multiplicative intensity model of counting processes after suitable random time changes. Standard large-sample results for counting processes following this multiplicative model can therefore be used to make inferences on the above class of semi-Markov models (assuming the proportionality functions are known), including the case where observations may be censored. Large-sample results for a four-state model used in clinical trials are presented.