The Annals of Statistics
1978, Vol. 6, No. 4, 701-726

NONPARAMETRIC INFERENCE FOR A FAMILY OF COUNTING PROCESSES[1]

By Odd Aalen

University of California, Berkeley and University of Copenhagen

Let $\mathbf{N} = (N_1, \cdots, N_k)$ be a multivariate counting process and let $\mathscr{F}_t$ be the collection of all events observed on the time interval $[0, t]$. The intensity process is given by

$$\Lambda_i(t) = \lim_{h \downarrow 0} \frac{1}{h} E(N_i(t+h) - N_i(t) \mid \mathscr{F}_t) \qquad i = 1, \cdots, k\,.$$

We give an application of the recently developed martingale-based approach to the study of $\mathbf{N}$ via Λ. A statistical model is defined by letting $\Lambda_i(t) = \alpha_i(t) Y_i(t)$, $i = 1, \cdots, k$, where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_k)$ is an unknown nonnegative function while $\mathbf{Y} = (Y_1, \cdots, Y_k)$, together with $\mathbf{N}$, is a process observable over a certain time interval. Special cases are time-continuous Markov chains on finite state spaces, birth and death processes and models for survival analysis with censored data.

The model is termed nonparametric when $\boldsymbol{\alpha}$ is allowed to vary arbitrarily except for regularity conditions. The existence of complete and sufficient statistics for this model is studied. An empirical process estimating $\beta_i(t) = \int_0^t \alpha_i(s)\,ds$ is given and studied by means of the theory of stochastic integrals. This empirical process is intended for plotting purposes and it generalizes the empirical cumulative hazard rate from survival analysis and is related to the product limit estimator. Consistency and weak convergence results are given. Tests for comparison of two counting processes, generalizing the two sample rank tests, are defined and studied. Finally, an application to a set of biological data is given.

1. Introduction. A point process is, roughly speaking, a countable random collection of points on the real line. Throughout this paper we will, if not otherwise stated, restrict all processes to the time interval $[0, 1]$. Let $N(t)$ be the number of points in $[0, t]$. The process N can be said to count the events of the point process, and hence it is called a (univariate) counting process. A multivariate counting process is a collection of k univariate counting processes which may of course be dependent on each other.

Let $\mathscr{F}_t$ be the collection of all events observed on $[0, t]$. Sometimes the following limit exists:

$$\Lambda_i(t) = \lim_{h \downarrow 0} \frac{1}{h} E(N_i(t+h) - N_i(t) \mid \mathscr{F}_t) \qquad i = 1, \cdots, k\,.$$

Received February 1977; revised August 1977.

[1] This paper was prepared with the support of the Norwegian Research Council for Science and the Humanities.

AMS 1970 *subject classifications.* Primary 62G05, 62G10, 62M99, 62N05; Secondary 60G45, 60H05, 62M05.

Key words and phrases. Point process, counting process, intensity process, inference for stochastic processes, nonparametric theory, empirical process, survival analysis, martingales, stochastic integrals.

The stochastic process $\mathbf{\Lambda} = (\Lambda_1, \cdots, \Lambda_k)$ is called the intensity process of the counting process $\mathbf{N} = (N_1, \cdots, N_k)$.

Recently several papers concerned with the study of counting processes via the intensity process have appeared. The development seems to have started with the papers of McFadden (1965), Cox and Lewis (1972), Papangelou (1972), Rubin (1972), Snyder (1972), and Bremaud (1972). Entirely different approaches are taken by the different authors. We are particularly interested in the one taken by Bremaud. He uses the modern theory of square integrable martingales and stochastic integrals. Bremaud's approach has been developed further in Bremaud (1974), Dolivo (1974), Boel, Varayia and Wong (1975a, b), Segall and Kailath (1975a, b) and Davis (1976). Jacod (1973, 1975) has developed the theory from a somewhat different point of view. His papers constitute an important supplement to those of the other authors.

So far the martingale-based counting process theory has been developed and used mainly in the context of communication engineering. Martins-Neto (1974) considers applications to queuing theory and studies some estimation in that context, but otherwise no statistical applications seem to have appeared so far. The object of the present author is firstly to present the elements of the theory to a statistical audience, and secondly to demonstrate its usefulness for statistical inference. The results are revised versions of parts of the author's Ph.D. dissertation (Aalen, 1975).

Section 2 is a short introduction to stochastic integrals. Section 3 gives a short review of a part of the martingale-based counting process theory. In Section 4 we present a specific nonparametric statistical model for counting processes with several motivating examples. Section 5 considers the question of existence of complete and sufficient statistics while Sections 6 and 7 contain estimation and testing. Finally, in Section 8 we give an application to a set of real data. One of the points we want to emphasize is how the theory of martingales and stochastic integrals is a very useful tool not only in the probabilistic part of the theory, but also in the inference part.

2. Square integrable martingales and stochastic integrals. Expositions of the theory of square integrable martingales and stochastic integrals may be found in Kunita and Watanabe (1967), Meyer (1967, 1971), Doléans-Dadé and Meyer (1970) and Courrège (1963).

Let $(\Omega, \mathscr{F}, P)$ be a probability space and let $\{\mathscr{F}_t\}_{t\in[0\,1]}$ be a family of sub-σ-algebras of $\mathscr{F}$. Assume $\{\mathscr{F}_t\}$ is *increasing*, that is $\mathscr{F}_s \subset \mathscr{F}_t$ whenever $s < t$. Assume also that $\{\mathscr{F}_t\}$ is right-continuous, that is $\bigcap_{h>0} \mathscr{F}_{t+h} = \mathscr{F}_t$.

Let $\{X(t)\}_{t\in[0,1]}$ be a real-valued stochastic process. We will usually denote $\{X(t)\}$ simply by X. We say that X is *adapted to* $\{\mathscr{F}_t\}$ if $X(t)$ is $\mathscr{F}_t$-measurable for all $t \in [0, 1]$.

A martingale M satisfying $M(0) = 0$ and having sample functions that are right-continuous with left-hand limits is called *square integrable* if $\sup_t E(M(t)^2) < \infty$.

The space of these processes is denoted by $\mathscr{M}^2$. Let M, M_1, and M_2 be square integrable martingales. The *variance process* $\langle M, M\rangle$ of M and its extension $\langle M_1, M_2\rangle$ are defined in Kunita and Watanabe (1967). M_1 and M_2 are said to be *orthogonal* if $\langle M_1, M_2\rangle = 0$, and this is equivalent to $M_1 M_2$ being a martingale.

The concept of a *predictable* process plays an important role in the theory of stochastic integrals. For our purpose it is enough to note that *any left-continuous adapted process with right-hand limits is predictable.* Let M be a square integrable martingale. The class of predictable processes H satisfying

$$E[\int_0^1 H(s)^2 \, d\langle M, M\rangle(s)] < \infty$$

is denoted by $L^2(M)$. The integral above is a Lebesgue–Stieltjes integral as are all integrals below unless they are taken with respect to square integrable martingales.

The stochastic integral of H with respect to M is a mapping from $L^2(M)$ into $\mathscr{M}^2$. Its value at time t is denoted by $\int_0^t H(s)\, dM(s)$. The process itself will occasionally be written as $\int H\, dM$. The stochastic integral is usually defined in an abstract way by means of Hilbert space theory. In the present paper we will mainly be interested in the situation where the stochastic integral coincides with the corresponding Lebesgue–Stieltjes integral. By Proposition 3 of Doléans-Dadé and Meyer (1970) a sufficient condition for this to take place is the following:

$$E \int_0^1 |H(s)| \, d|M(s)| < \infty \, .$$

Let M_1 and M_2 be members of $\mathscr{M}^2$ and let $H_i \in L^2(M_i)$, $i = 1, 2$. An important property of the stochastic integral is the following:

$$\langle \int H_1\, dM_1, \int H_2\, dM_2\rangle = \int H_1 H_2 \, d\langle M_1, M_2\rangle \, . \tag{2.1}$$

Hence, $\int H_1\, dM_1$ and $\int H_2\, dM_2$ will be orthogonal whenever M_1 and M_2 are orthogonal.

Recently the theory we have reviewed in this section has been extended to the so-called *local martingales* which include martingales as special cases. This theory is given in Doléans-Dadé and Meyer (1970) and it is used in several of the counting process papers mentioned in the introduction, see e.g., Boel et al. (1975a, b). The advantage of using the theory of local martingales is that one requires slightly weaker conditions on the counting processes. From our point of view this is not very important since we are primarily concerned with statistical applications.

3. Basic theory of counting processes.

3.1. *Preliminaries.* A multivariate stochastic process $\mathbf{N}(t) = (N_1(t), \cdots, N_k(t))$ defined on the time interval $[0, 1]$ will be called a (multivariate) counting process whenever the following conditions are fulfilled: (i) The sample functions of each component process N_i are right-continuous step functions with value 0 at $t = 0$ and with a finite number of jumps, each positive and of size 1. (ii) Two component processes N_i and N_j $(i \neq j)$ cannot jump at the same time.

In the usual way $\mathbf{N}$ can be thought of as being defined on an underlying abstract space Ω. Let the σ-algebra $\mathscr{N}_t$ on Ω be defined for each $t \in [0, 1]$ in the following way: $\mathscr{N}_t$ is generated by all sets of the form $\{\omega \in \Omega \mid N(s) \in B\}$, $s \leqq t$, where the set B runs through all subsets of $\{0, 1, 2, \cdots\}$. For some purposes we will need a somewhat wider family of σ-algebras $\{\mathscr{F}_t\}$ where for each t $\mathscr{F}_t \supset \mathscr{N}_t$. We write $\mathscr{F} = \mathscr{F}_1$ and $\mathscr{N} = \mathscr{N}_1$. Clearly $\{\mathscr{N}_t\}$ is an increasing family of σ-algebras and by Corollary 2.5 of Boel et al. (1975a) it is also right-continuous. We will require the same to hold for $\{\mathscr{F}_t\}$. By Corollary 2.2 of Boel et al. (1975a) the jump times of $\mathbf{N}$ are stopping times with respect to $\{\mathscr{N}_t\}$ and hence also with respect to $\{\mathscr{F}_t\}$.

Throughout the paper, if not otherwise mentioned, all probabilistic statements will be made with respect to a fixed family $\{\mathscr{F}_t\}$ satisfying the general requirements above. In some cases we will assume that $\{\mathscr{F}_t\}$ coincides with the special family $(\mathscr{N}_t)$, but this will be explicitly stated each time.

Let P be a fixed measure on $\mathscr{F}$. All $\mathscr{F}_t$ and $\mathscr{N}_t$ are to be completed with respect to P (see Section III of Boel et al. (1975a)).

A stopping time T is a nonnegative random variable satisfying $\{T \leqq t\} \in \mathscr{F}_t$.

For a stopping time T we define $\mathscr{F}_T$ as consisting of those sets $A \in \mathscr{F}$ for which $A \cap \{T \leqq t\} \in \mathscr{F}_t$.

3.2. *The intensity process.* We will make the following basic assumption:

ASSUMPTION 3.1.

(i) $EN_i(1) < \infty \quad i = 1, \cdots, k$,

(ii) The jump times $T_1 < T_2 < \cdots$ of $\sum N_i$ are totally inaccessible.

REMARK. Condition (i) is needed in order to work within the square integrable martingale framework and avoid the local martingales. For discussion of the concept of *total inaccessibility* of jump times, see Meyer (1966) and Boel et al. (1975a, Theorem 2.1). To give it some intuitive content, consider the situation where $\mathscr{F}_t = \mathscr{N}_t$ for all t. By Theorem 2.1 of Boel et al. (1975a), total inaccessibility of the jump times is then implied if all (the possibly degenerate) distribution functions $P(T_{n+1} - T_n \leqq t \mid \mathscr{F}_{T_n})$, $n = 0, 1, \cdots$, are continuous in t.

THEOREM 3.2. *Under Assumption* 3.1 *there exists a unique vector of continuous increasing processes* $\mathbf{A} = (A_1, \cdots, A_k)$ *with* $A_i(0) = 0$ *adapted to* $\{\mathscr{F}_t\}$ *such that*

$$M_i = N_i - A_i \in \mathscr{M}^2 \qquad i = 1, \cdots, k .$$

The martingales M_i *also satisfy*:

$$\text{(i)} \qquad \langle M_i, M_i \rangle = A_i \qquad i = 1, \cdots, k ,$$

$$\text{(ii)} \qquad \langle M_i, M_j \rangle = 0 \qquad \textit{whenever } i \neq j .$$

REMARK. This theorem is stated within the local martingale framework (see Section 2) in Boel et al. (1975a, Proposition 3.2 and Lemma 3.1). The extension to the square integrable martingale framework is made in the Appendix of the present paper.

Occasionally the sample functions of the processes A_i will be absolutely continuous with respect to t. Sufficient conditions for this to take place are given in the remark after Proposition 3.2 of Boel et al. (1975a). We will assume that there exists a nonnegative process $\Lambda = (\Lambda_1, \cdots, \Lambda_k)$ with sample functions that are left-continuous and have right-hand limits at each point and such that

$$A_i(t) = \int_0^t \Lambda_i(s)\, ds \qquad i = 1, \cdots, k\,.$$

Λ is adapted to $\{\mathscr{F}_t\}$ since that family of σ-fields is right-continuous.

The process Λ is called the *intensity process* of $\mathbf{N}$ with respect to $\{\mathscr{F}_t\}$ and the measure P and we will show that it satisfies the requirements that one would usually place on an intensity process.

LEMMA 3.3. *Assume that the* Λ_i, $i = 1, \cdots, k$, *are bounded by an integrable random variable. Put* $\bar{N}(t) = \sum_{i=1}^k N_i(t)$. *Then the following statements hold*:

(i) $$\lim_{h\downarrow 0} \frac{1}{h} E(N_i(t+h) - N_i(t) \mid \mathscr{F}_t) = \Lambda_i(t+) \qquad i = 1, \cdots, k\,;$$

(ii) $$\lim_{h\downarrow 0} \frac{1}{h} P(N_i(t+h) - N_i(t) = 1 \mid \mathscr{F}_t)$$
$$= \lim_{h\downarrow 0} \frac{1}{h} [1 - P(N_i(t+h) - N_i(t) = 0 \mid \mathscr{F}_t)] = \Lambda_i(t+)\,;$$

(iii) $$\lim_{h\downarrow 0} \frac{1}{h} P(\bar{N}(t+h) - \bar{N}(t) > 1 \mid \mathscr{F}_t) = 0\,;$$

(iv) $$\lim_{h\downarrow 0} \frac{1}{h} E[(\bar{N}(t+h) - \bar{N}(t)) I(\bar{N}(t+h) - \bar{N}(t) > 1) \mid \mathscr{F}_t] = 0\,.$$

REMARK. Part (i) can be found in Bremaud (1972) and Dolivo (1974). The remaining part of the lemma does not seem to have been explicitly stated and proved in the martingale-based counting process literature. They are however quite straightforward consequences of Theorem 3.2.

PROOF. By Theorem 3.2 and the dominated convergence theorem we have

$$\lim_{h\downarrow 0} E\left[\frac{1}{h}(N_i(t+h) - N_i(t)) \,\middle|\, \mathscr{F}_t\right] = \lim_{h\downarrow 0} E\left[\frac{1}{h}\int_t^{t+h} \Lambda_i(s+)\, ds \,\middle|\, \mathscr{F}_t\right]$$
$$= E\left[\lim_{h\downarrow 0} \frac{1}{h}\int_t^{t+h} \Lambda_i(s+)\, ds \,\middle|\, \mathscr{F}_t\right]$$
$$= E(\Lambda_i(t+) \mid \mathscr{F}_t) = \Lambda_i(t+)\,.$$

This proves that (i) holds. To prove (ii) fix i and t and let $S > t$ be the time of the first jump after t. Since S is a stopping time, it follows that the process $I_{\langle t,S]}(u)$ is adapted to $\{\mathscr{F}_u\}$, hence it is predictable since it is obviously left-continuous. Hence we can, for $h > 0$, define the stochastic integral

$$\int_t^{t+h} I_{\langle t,S]}(u)\, dM_i(u)\,.$$

This process, with h as time variable, is a martingale, and hence we have:

$$\begin{aligned}\lim_{h\downarrow 0}\frac{1}{h}\, &P(N_i(t+h)-N_i(t)\geqq 1\,|\,\mathscr{F}_t)\\ &= \lim_{h\downarrow 0}\frac{1}{h}\, E[\textstyle\int_t^{t+h} I_{\langle t,S]}(u)\,dN_i(u)\,|\,\mathscr{F}_t]\\ &= \lim_{h\downarrow 0}\frac{1}{h}\, E[\textstyle\int_t^{t+h} I_{\langle t,S]}(u)\Lambda_i(u+)\,du\,|\,\mathscr{F}_t]\\ &= E\left[\lim_{h\downarrow 0}\frac{1}{h}\textstyle\int_t^{t+h} I_{\langle t,S]}(u)\Lambda_i(u+)\,du\,\middle|\,\mathscr{F}_t\right]\\ &= E[\Lambda_i(t+)\,|\,\mathscr{F}_t] = \Lambda_i(t+)\,.\end{aligned}$$

The second-to-last equality follows since $S > t$. Hence, the second equality of (ii) is proved. Let now S' be the time of the second jump of N_i after t. Similarly to above we have:

$$\begin{aligned}\lim_{h\downarrow 0}\frac{1}{h}\, &P(N_i(t+h)-N_i(t)\geqq 2\,|\,\mathscr{F}_t)\\ &= E\left[\lim_{h\downarrow 0}\frac{1}{h}\textstyle\int_t^{t+h} I_{\langle S,S']}(u)\Lambda_i(u+)\,du\,\middle|\,\mathscr{F}_t\right] = 0\,.\end{aligned}$$

Hence, the first part of (ii) follows.

Parts (iii) and (iv) follow by observing that (i) and (ii) are valid for the counting process $\bar{N}$ with intensity process $\bar{\Lambda} = \sum_{i=1}^k \Lambda_i$. □

Assume that $H \in L^2(M_i)$ for one of the basic martingales M_i defined above. Then the stochastic integral $\int_0^t H\,dM_i$ is well defined. One can, however, also think of this integral as a Lebesgue–Stieltjes integral and one may ask when the two interpretations coincide. This is important when we actually want to compute the integral as we will do later on. By applying Proposition 3 of Doléans-Dadé and Meyer (1970) we get the following sufficient condition:

$$E \int_0^1 |H(s)|\,dN_i(s) < \infty\,. \tag{3.1}$$

We will end this section by stating the innovation theorem. The theorem is a trivial multivariate generalization of the univariate case which may be found for instance in Segall and Kailath (1975a, Theorem 3).

THEOREM 3.4. *Let $\{\mathscr{G}_t\}$ and $\{\mathscr{F}_t\}$ with $\mathscr{N}_t \subset \mathscr{G}_t \subset \mathscr{F}_t$ for all t be families of σ-algebras satisfying the general assumptions of Section* 3.1. *Assume that a counting process* $\mathbf{N}$ *has intensity process* $\mathbf{\Lambda}$ *with respect to $\{\mathscr{F}_t\}$. Then the intensity process with respect to $\{\mathscr{G}_t\}$ is $E(\mathbf{\Lambda}(t)\,|\,\mathscr{G}_t)$.*

3.3. *The likelihood of a counting process.* Assume in this section that $\mathscr{F}_t = \mathscr{F}_0 \vee \mathscr{N}_t$ for all $t \in [0, 1]$, i.e., $\mathscr{F}_t$ is generated by all sets $A \cup B$ where $A \in \mathscr{F}_0$ and $B \in \mathscr{N}_t$. Let P_1 and P_2 be two measures on $\mathscr{F}$, both giving $\mathbf{N}$ the same intensity process. According to Jacod (1975, Theorem 3.4) P_1 and P_2 coincide on $\mathscr{F}$ if they coincide on $\mathscr{F}_0$.

Let P_0 be a probability measure on $(\Omega, \mathscr{F})$ that makes $N_1, \cdots, N_k$ independent Poisson processes, each with intensity 1. Let $\mathbf{\Lambda}$ be the intensity process corresponding to the fixed measure P. Assume that P is absolutely continuous with respect to P_0 on $\mathscr{F}_0$. Then under very weak conditions (see Jacod and Memin (1977)) P is absolutely continuous with respect to P_0 on $\mathscr{F}$ and $L_t = E(dP/dP_0 \mid \mathscr{F}_t)$ is given by

$$L_t = L_0 \prod_{n=1}^{\bar{N}(t)} \Lambda_{J_n}(T_n) \exp(-\int_0^t \bar{\Lambda}(s)\, ds + kt) \tag{3.2}$$

where $\bar{N} = \sum_{i=1}^k N_i$, $\bar{\Lambda} = \sum_{i=1}^k \Lambda_i$, where $J_n = i$ if jump number n occurs in N_i, and where $T_1 < T_2 < \cdots$ are the jump times of $\bar{N}$.

4. The multiplicative intensity model. In the remaining part of the paper we will study the following statistical model, called *the multiplicative intensity model*. Assume that $\mathbf{\Lambda}$ can be written in the form

$$\Lambda_i(t) = \alpha_i(t) Y_i(t) \qquad i = 1, \cdots, k,\ t \in [0, 1]$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_k)$ is an unknown function while $\mathbf{Y} = (Y_1, \cdots, Y_k)$ is a stochastic process which together with $\mathbf{N}$ can be observed over the time interval $[0, 1]$. For each i we require that α_i and the sample functions of Y_i be nonnegative, left-continuous functions with right-hand limits. $\mathbf{Y}$ has to be adapted to $\{\mathscr{F}_t\}$ since $\mathbf{\Lambda}$ is adapted. We will also assume that $\int_0^1 \alpha_i(s)\, ds < \infty$, $i = 1, \cdots, k$. Let $\mathscr{A}$ denote the set of all functions $\boldsymbol{\alpha}$ satisfying the assumptions given here.

Next we will proceed to give several examples of the multiplicative intensity model to indicate its broad scope.

After the examples we will continue with a study of the model that arises by letting $\boldsymbol{\alpha}$ vary freely in $\mathscr{A}$. We will term this situation *nonparametric*. We will show that inference procedures can be developed and studied without further specifications of the probabilistic structure of $(\mathbf{N}, \mathbf{Y})$. For instance, it is not necessary to assume that $(\mathbf{N}, \mathbf{Y})$ is Markovian, an assumption which is usually made in the literature when one studies processes of this kind.

EXAMPLE 1. *A simple life testing model.* Let $X_1, \cdots, X_n$ be independent and identically distributed nonnegative random variables. Let F denote the cumulative distribution function of X_i and assume $F(1) < 1$. Assume that the density $f(t)$ exists for $t \in [0, 1]$. In life testing one is interested in the hazard rate defined by $\alpha(t) = f(t)/(1 - F(t))$ on $[0, 1]$. The requirement $F(1) < 1$ is equivalent to the finiteness of $\int_0^1 \alpha(s)\, ds$. Assume also that α is left-continuous with right-hand limits.

Consider the multivariate counting process $\mathbf{N}$ with components $N_i(t) = I(X_i \leq t)$, $i = 1, \cdots, n$, and the family of σ-fields $\{\mathscr{N}_t\}$. F defines uniquely a measure P on $\mathscr{N}$. In the following lemma we state that *N_i has intensity process $\alpha_i(t)I(X_i \geq t)$ with respect to $\{\mathscr{N}_t\}$ and P.*

LEMMA 4.1.

$$M_i(t) = I(X_i \leqq t) - \int_0^t \alpha_i(u) I(X_i \geqq u)\, du \qquad i = 1, \cdots, n$$

are orthogonal square integrable martingales with respect to $\{\mathscr{N}_t\}$.

PROOF. The square integrability and orthogonality follow immediately from the assumptions. For each i we have to prove $E(M_i(t) \mid \mathscr{N}_s) = M_i(s)$ whenever $0 \leqq s < t \leqq 1$. Since X_j for $j \neq i$ is independent of X_i it is enough to consider the following elements of $\mathscr{N}_s$: $\{X_i > s\}$ and $\{X_i = r\}$ for $0 \leqq r \leqq s$. We have:

$$\begin{aligned} E(M_i(t) \mid X_i > s) &= P(X_i \leqq t \mid X_i > s) - \int_0^t \alpha(u) P(X_i \geqq u \mid X_i > s)\, du \\ &= (F(t) - F(s))(1 - F(s))^{-1} - \int_0^s \alpha(u)\, du \\ &\quad - \int_s^t \alpha(u)(1 - F(u))(1 - F(s))^{-1}\, du = -\int_0^s \alpha(u)\, du \end{aligned}$$

where the last expression is the value of $M_i(s)$ on $\{X_i > s\}$. Further, we have for $r \leqq s$:

$$\begin{aligned} E(M_i(t) \mid X_i = r) &= P(X_i \leqq t \mid X_i = r) - \int_0^t \alpha(u) P(X_i \geqq u \mid X_i = r)\, du \\ &= 1 - \int_0^r \alpha(u)\, du \end{aligned}$$

where the last expression is the value of $M_i(s)$ on $\{X_i = r\}$. □

Define

$$\begin{aligned} \bar{N}(t) &= \textstyle\sum_{i=1}^n N_i(t)\,, \\ \bar{\Lambda}(t) &= \textstyle\sum_{i=1}^n \Lambda_i(t) = \alpha(t)(n - \bar{N}(t-))\,. \end{aligned}$$

Clearly $\bar{N}$ is a counting process with intensity process $\bar{\Lambda}$ relative to P and $\{\mathscr{N}_t\}$. $\bar{N}$ is a one-to-one function of the order statistics corresponding to $\mathbf{X}$. Hence $\bar{N}$ is a sufficient statistic for the nonparametric family of the kind defined above. Let $\{\bar{\mathscr{N}}_t\}$ be the family of σ-algebras generated by $\bar{N}$. From the expression for $\bar{\Lambda}$ above, we see that $\bar{\Lambda}$ is adapted to $\{\bar{\mathscr{N}}_t\}$. Hence, by the innovation theorem (Theorem 3.4), $\bar{\Lambda}$ is also the intensity process of $\bar{N}$ relative to $\{\bar{\mathscr{N}}_t\}$.

Clearly we have a multiplicative intensity model with $\bar{N}$ being the counting process, α the unknown function, $Y(t) = n - \bar{N}(t-)$ and $\mathscr{F}_t = \bar{\mathscr{N}}_t$.

EXAMPLE 2. *A general life testing model.* Using common life testing terminology one can say that $Y(t)$ in Example 1 is the number of items at risk, the *risk set*, at time t. The process $\bar{N}$ counts the failures that occur. Of course, there is no reason in general why Y should have the specific relationship to $\bar{N}$ given in the previous example. For instance, in medical survival studies one usually has lots of censoring and also patients may come into the study during the trial. This means that the size of the risk set, measured by Y, may have several changes during the trial which are unrelated to the failure process $\bar{N}$. One of the advantages of the multiplicative intensity model defined at the beginning of this section, is that it allows for the process Y to change quite arbitrarily during the observation period. If the changes are due to outside random influences this may be modelled by increasing the σ-algebras $\{\mathscr{F}_t\}$ appropriately.

Note that the most important requirement made to the process Y in the multiplicative intensity model is that it be *adapted to* $\{\mathscr{F}_t\}$. This means that the changes in the risk set can depend quite arbitrarily on the past, but *not* on the future, e.g., future failures. This has also been commented upon by Efron (1975, Section 5F) on a more intuitive basis.

When several groups of items are under study, or when the same group is studied with respect to several causes of failure, one needs the full multivariate version of the multiplicative intensity model. Note that the processes $Y_1, \cdots, Y_k$ may depend on each other quite arbitrarily.

In this paper we only mention this application to life testing as an illustration. One should of course study in greater detail different kinds of censoring schemes and relate them to the present model. We are planning to do that in a later paper. It should however be clear that our model is considerably more general than the censoring models commonly considered in the literature (see e.g., Breslow (1970) and Kaplan and Meier (1958)).

EXAMPLE 3. *A Markov chain model with censoring.* Let $\{1, \cdots, m\}$ be the finite state space of a time-continuous Markov chain. Let $\alpha_{ij}(t)$ be the infinitesimal transition probability (or force of transition) between the states i and j. Let all functions α_{ij} satisfy the same requirements as the α_i in the multiplicative intensity model.

We will now assume that several "particles" are moving around on the state space independently of each other and according to the probabilistic structure given above. Each particle starts out in a given state at time 0. This state does not have to be the same for all particles.

Let $N_{ij}(t)$ for $i \neq j$ be a process counting the number of direct transitions from i *to* j. Let $Y_i(t)$ be the number of particles in state i at time t, and let it have left-continuous sample paths. Denote by $\mathbf{N}$ the multivariate counting process consisting of all processes N_{ij} and let $\{\mathscr{N}_t\}$ be the σ-algebras generated by $\mathbf{N}$. It is intuitively reasonable that each component process N_{ij} of $\mathbf{N}$ has an intensity process given by $Y_i(t)\alpha_{ij}(t)$ with respect to $\{\mathscr{N}_t\}$. One can prove that this is so (see Aalen, 1975, Section 5D). Hence we have again the multiplicative intensity model.

As in the previous example one can let the Y_i-processes vary quite arbitrarily and allow for general kinds of censoring. The importance of this is clear since Markov chain models are much used in demography, actuarial science and medical statistics (see, e.g., Hoem 1971).

It should be clear that birth and death processes and branching processes can similarly be described within our framework.

EXAMPLE 4. *The birth and death process.* Consider a birth and death process with population size $X(t)$ at time t. Let $\lambda(t)$ and $\mu(t)$ denote the time-dependent intensities of an individual giving birth to a new individual or dying. Let $(B(t), D(t))$ be the bivariate process counting the number of births and deaths in the population. Let $\mathscr{F}_t$ be generated by $\{X(s), s \leqq t\}$. It is intuitively clear that $(B(t), D(t))$ is a counting process with intensity process $(\lambda(t)X(t), \mu(t)X(t))$ with respect to $\{\mathscr{F}_t\}$. A further example of the multiplicative intensity model is studied in Section 8.

5. Completeness in the nonparametric model. We assume that $\boldsymbol{\alpha}$ varies freely in $\mathscr{A}$, that $\{\mathscr{F}_t\} = \{\mathscr{N}_t\}$ and that the likelihood representation (3.2) is valid. We will denote by $P(\boldsymbol{\alpha})$ the corresponding measure P for a given $\boldsymbol{\alpha}$. The likelihood is given by

$$L_1 = \exp\{\sum_{i=1}^k [\int_0^1 \log(Y_i(s))\, dN_i(s) + \int_0^1 \log \alpha_i(s)\, dN_i(s) - \int_0^1 \alpha_i(s) Y_i(s)\, ds] + k\}\,.$$

Since we have assumed that $\mathbf{Y}$ is adapted to $\{\mathscr{N}_t\}$, it follows that the process $\mathbf{N}$ is a sufficient statistic. One may ask whether it will also be complete (for the concept of completeness, see Lehmann (1959)).

THEOREM 5.1. *Assume that we can write*

$$Y_i(t) = \sum_{j=1}^k \int_0^t f_{ji}(s)\, dN_j(s) + g_i(t)$$

where the f_{ij} and g_i are known, left-continuous, nonnegative bounded functions on $[0, 1]$. *Assume also that the f_{ij} are step functions and that the g_i have right-hand limits. Then* $\mathbf{N}$ *is a complete statistic.*

REMARK. Theorem 5.1 covers many situations of interest. One example is the life testing model described in Example 1 of Section 4. In this case Theorem 5.1 reduces to the well-known statement that the order statistics of a random sample is complete in the nonparametric case. Another example is the Markov chain case in Example 3 of Section 4. When there is no censoring, then we have $Y_i(t) = Y_i(0) + \sum (N_{ji}(t) - N_{ij}(t))$ where the sum is taken over all j different from i, hence the conditions of Theorem 5.1 are satisfied. When there is censoring present, the situation is more complicated and the theorem will usually not be applicable. It is also not applicable in the example of Section 8. A special case of the Theorem 5.1 for a competing risks model is given in Aalen (1976).

PROOF. Let $\boldsymbol{\gamma} = \phi(\boldsymbol{\alpha})$ be the transformation defined by

$$\gamma_i(t) = \log \alpha_i(t) - \sum_{j=1}^k (f_{ji}(t) \int_t^1 \alpha_j(s)\, ds) \qquad i = 1, \cdots, k\,,$$
$$\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_k)\,.$$

By some algebraic manipulations L_1 reduces to

$$L_1 = \exp\{\sum_{i=1}^k [\int_0^1 \log(Y_i(s))\, dN_i(s) + \int_0^1 \gamma_i(s)\, dN_i(s) - \int_0^1 \alpha_i(s) g_i(s)\, ds] + k\}\,.$$

One sees that the only part of $\log L_1$ that depends on both the unknown parameter function $\boldsymbol{\alpha}$ and the process $\mathbf{N}$ is

$$\sum_{i=1}^k \int_0^1 \gamma_i(s)\, dN_i(s)\,.$$

Hence we have some sort of an analogy to the exponential class of distributions in the parametric case. The idea of the proof is to use a similar method to that used in the proof of completeness of a regular exponential family (see Lehmann (1959), page 133). But first we need to know something about the set of functions $\boldsymbol{\gamma}$ that are generated when $\boldsymbol{\alpha}$ varies in $\mathscr{A}$. More precisely, we will prove that all step functions $\boldsymbol{\gamma}$ satisfying a certain condition are generated in this way.

LEMMA. *Let $M > 0$ be a constant such that*

$$\text{(i)} \qquad \sup_{i,t} \textstyle\sum_{j=1}^{k} f_{ij}(t) < M .$$

Let γ_i, $i = 1, \cdots, k$, be left-continuous step functions with right-hand limits such that

$$\text{(ii)} \qquad \sup_{i,t} \gamma_i(t) \leqq -1 - \log M .$$

Then there exists an $\boldsymbol{\alpha} \in \mathscr{A}$ such that $\boldsymbol{\gamma} = \phi(\boldsymbol{\alpha})$.

PROOF. Assume that $\boldsymbol{\gamma} = \phi(\boldsymbol{\alpha})$ for some $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. Then $\boldsymbol{\gamma}' = \phi(\boldsymbol{\alpha}')$ for $\boldsymbol{\gamma}'$ and $\boldsymbol{\alpha}'$ given by

$$\alpha_i' = M\alpha_i , \qquad \gamma_i' = \gamma_i + \log M$$
$$f'_{ji} = \frac{1}{M} f_{ji} \qquad\qquad i = 1, \cdots, k$$

and clearly f'_{ji} satisfies (i) and γ_i' (ii) with $M = 1$. Hence we can without loss of generality assume $M = 1$.

Assume then that $\boldsymbol{\gamma}$ is a step function satisfying the conditions given in the lemma with $M = 1$. We will construct an increasing sequence of step functions converging to an $\boldsymbol{\alpha} \in \mathscr{A}$, satisfying $\boldsymbol{\gamma} = \phi(\boldsymbol{\alpha})$.

Let $t_0 = 0 < t_1 < t_2 < \cdots < t_m = 1$ be a partition, denoted Π, of $[0, 1]$ such that $\boldsymbol{\gamma}$ and the f_{ij} are constant on each subinterval. We define a step function $\boldsymbol{\alpha}^\Pi$ in the following way: $\boldsymbol{\alpha}^\Pi$ is constant on each subinterval of Π. We denote the value of α_i^Π on $\langle t_{\nu-1}, t_\nu]$ by $\alpha_{i\nu}$. We define recursively for each i:

$$\alpha_{im} = \exp \gamma_i(1)$$
$$\alpha_{i\nu} = \exp[\gamma_i(t_\nu) + \textstyle\sum_{j=1}^{k} (f_{ji}(t_\nu) \sum_{l=\nu+1}^{m} \alpha_{jl}(t_l - t_{l-1}))] .$$

(When $l > m$ the summation is empty.)

Let Π' be a finer partition than Π. Let t' be the largest partition point for Π' that is not also a partition point for Π, and assume that t' is in the subinterval $\langle t_{\nu-1}, t_\nu]$. Then

$$\begin{aligned} \alpha_i^{\Pi'}(t') &= \exp[\gamma_i(t') + \textstyle\sum_{j=1}^{k} (f_{ji}(t')\{\sum_{l=\nu+1}^{m} \alpha_{jl}(t_l - t_{l-1}) + \alpha_{j\nu}(t_\nu - t')\})] \\ &\geqq \exp[\gamma_i(t_\nu) + \textstyle\sum_{j=1}^{k} (f_{ji}(t_\nu) \sum_{l=\nu+1}^{m} \alpha_{jl}(t_l - t_{l-1}))] \\ &= \alpha_{i\nu} = \alpha_i^\Pi(t') . \end{aligned}$$

By continuing this procedure through all smaller partition points of Π', we see that for all t

$$\text{(5.1)} \qquad \alpha_i^\Pi(t) \leqq \alpha_i^{\Pi'}(t) \qquad i = 1, \cdots, k .$$

Since (i) and (ii) are supposed to hold with $M = 1$, we also have that

(5.2) $$\alpha_i^{\Pi'}(t) \leqq 1 \qquad i = 1, \cdots, k$$

for all Π' finer than Π. Let Π_r, $r = 1, 2, \cdots$, be a sequence of increasingly finer partitions, each of them finer than Π, such that the maximum length of all subintervals goes to 0. By (5.1) and (5.2) it follows that the sequence of functions $\boldsymbol{\alpha}^{\Pi_r}$ has a finite pointwise limit. Denote this by $\boldsymbol{\alpha}$. We will show that $\boldsymbol{\gamma} = \phi(\boldsymbol{\alpha})$.

Let t be any point in $[0, 1]$ and let t_r be the smallest partition point of Π_r satisfying $t_r \geqq t$. We have:

$$\begin{aligned}\gamma_i(t) &= \gamma_i(t_r)\\ &= \log \alpha_i^{\Pi_r}(t_r) - \sum_{j=1}^k f_{ji}(t_r) \int_{t_r}^1 \alpha_j^{\Pi_r}(s)\, ds\\ &= \log \alpha_i^{\Pi_r}(t) - \sum_{j=1}^k f_{ji}(t_r) \int_t^1 \alpha_j^{\Pi_r}(s)\, ds + \sum_{j=1}^k f_{ji} \int_t^{t_r} \alpha_i^{\Pi_r}(s)\, ds\\ &\to_{r\to\infty} \log \alpha_i(t) - \sum_{j=1}^k f_{ji}(t) \int_t^1 \alpha_j(s)\, ds\,,\end{aligned}$$

that is, $\boldsymbol{\gamma} = \phi(\boldsymbol{\alpha})$. By the definition of ϕ we have:

$$\log \alpha_i(t) = \gamma_i(t) + \sum_{j=1}^k \left(f_{ji}(t) \int_t^1 \alpha_j(s)\, ds\right).$$

The right-hand side is left-continuous with right-hand limits, and so the same holds for $\log \alpha_i(t)$, and hence for $\boldsymbol{\alpha}$.

We have now proved the lemma and can continue with the proof of the theorem. Let g be a real valued function on Ω, integrable with respect to $P(\boldsymbol{\alpha})$ for all $\boldsymbol{\alpha} \in \mathscr{A}$ and assume that

$$\int g\, dP = 0 \qquad \text{for all} \quad \boldsymbol{\alpha} \in \mathscr{A}.$$

This implies

$$\int g L_1\, dP_0 = 0 \qquad \text{for all} \quad \boldsymbol{\alpha} \in \mathscr{A}.$$

Thus by the lemma

(5.3) $$\int g Z(\boldsymbol{\gamma})\, dP_0 = 0 \qquad \text{for all} \quad \boldsymbol{\gamma} \in \Gamma$$

where Γ is the set of all step functions $\boldsymbol{\gamma}$ satisfying the assumptions of the lemma with $M = 1$, and

$$Z(\boldsymbol{\gamma}) = \exp\left\{\sum_{i=1}^k \left[\int_0^1 \log (Y_i(s))\, dN_i(s) + \int_0^1 \gamma_i(s)\, dN_i(s)\right]\right\}.$$

Let g^- and g^+ be the positive and negative parts of g. Put

$$K = \int g^+ Z(-\mathbf{1})\, dP_0$$

where $\mathbf{1}$ denotes the vector $(1, \cdots, 1)$ with k elements and define the probability measures Q_1 and Q_2 by

$$\begin{aligned} dQ_1 &= K^{-1} g^+ Z(-\mathbf{1})\, dP_0\,,\\ dQ_2 &= K^{-1} g^- Z(-\mathbf{1})\, dP_0\,. \end{aligned}$$

Put $\gamma_i' = \gamma_i + 1$, $i = 1, \cdots, k$. Then (5.3) translates into

$$\int \exp\left[\sum_{i=1}^k \int_0^1 \gamma_i'(t)\, dN_i(t)\right] dQ_1 = \int \exp\left[\sum_{i=1}^k \int_0^1 \gamma_i'(t)\, dN_i(t)\right] dQ_2$$

for all nonpositive step functions γ_i', $i = 1, \cdots, k$ which are left-continuous and have right-hand limits. By the uniqueness of the Laplace transform of R^m-valued random variables it follows that Q_1 and Q_2 give the same joint distribution for the numbers of jumps in any set of disjoint intervals. Since the Poisson measure P_0 is uniquely determined by all such distributions, and since Q_1 and Q_2 are absolutely continuous with respect to P_0, it follows that $Q_1 = Q_2$ on $\mathscr{N}$. □

6. Nonparametric estimation.

6.1. *The estimator.* Let $\boldsymbol{\alpha}$ be an arbitrary element of $\mathscr{A}$. We will treat the problem of estimating the functions

$$\beta_i(t) = \int_0^t \alpha_i(s)\, ds \qquad i = 1, \cdots, k$$

or, rather, closely related quantities. We put $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_k)$.

Our estimator can be regarded as generalizing the empirical cumulative hazard rate studied by Nelson (1969), Altshuler (1970) and Aalen (1976). There is also a close relationship to the empirical distribution function for censored data studied by Kaplan and Meier (1958) and Meier (1975).

The *empirical $\boldsymbol{\beta}$-function*, as we will denote our estimator, is mainly intended to be used for plotting purposes. An application to a set of data is given in Section 8. An application to a birth and death process is given by Keiding (1976).

A basic requirement in order to be able to estimate $\boldsymbol{\beta}$ on the whole of $[0, 1]$ in a meaningful way is that the processes Y_i be strictly positive on the whole of $[0, 1]$. Often this will not be the case. One example is the Markov chain case treated in Example 3 of Section 4. In that case it is clearly possible that some state may become "empty," i.e., that $Y_i(t) = 0$ on subintervals of $[0, 1]$. Sometimes it may be possible to control this by "putting new particles" into a state when it becomes empty. At other times this may not be possible. In general, we will have to restrict ourselves to estimating the following quantities:

$$\beta_i^*(t) = \int_0^t \alpha_i(s) J_i(s)\, ds \qquad i = 1, \cdots, k$$

where

$$J_i(s) = \lim_{h \downarrow 0} I(Y_i(s-h) > 0)$$

and I is an indicator function. Our way of defining J_i has the purpose of making it left-continuous. We put $\boldsymbol{\beta}^* = (\beta_1^*, \cdots, \beta_k^*)$ and $\mathbf{J} = (J_1, \cdots, J_k)$.

Of course $\boldsymbol{\beta}^*$ is in general not a function but a stochastic process. However, in accordance with the terminology of the engineering literature, it is quite proper to talk about estimating a stochastic process. In fact, it seems that it will often be the case in a stochastic process context that one cannot know at the outset which parameters one will be able to estimate in a meaningful way, because it depends on how the process develops.

The expression $J_i(t)(Y_i(t))^{-1}$ is interpreted as 0 whenever $Y_i(t) = 0$. We need the following assumption:

ASSUMPTION 6.1. There exists a finite number c such that

$$\sup_{0<t<1} J_i(t)(Y_i(t))^{-1} < c \quad \text{a.s.} \qquad i = 1, \cdots, k .$$

Define:

$$\hat{\beta}_i(t) = \int_0^t J_i(s)(Y_i(s))^{-1}\, dN_i(s) \qquad i = 1, \cdots, k .$$

The integral is to be taken as a Stieltjes integral. Put $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \cdots, \hat{\beta}_k)$.

THEOREM 6.2. *Suppose Assumption* 6.1 *holds. Then the processes*

$$\hat{\beta}_i - \beta_i^* \qquad i = 1, \cdots, k$$

are orthogonal square integrable martingales with

(i) $$\langle \hat{\beta}_i - \beta_i^*, \hat{\beta}_i - \beta_i^* \rangle(t) = \int_0^t \alpha_i(s) J_i(s)(Y_i(s))^{-1}\, ds \qquad i = 1, \cdots, k .$$

REMARK. *We suggest* $\hat{\boldsymbol{\beta}}$ *as an estimator of* $\hat{\boldsymbol{\beta}}^*$. We will translate a part of the content of the theorem into more common language. The martingale property of $\hat{\beta}_i - \beta_i^*$ implies that

$$E\hat{\beta}_i(T) = E\beta_i^*(T) \qquad i = 1, \cdots, k$$

for all $\boldsymbol{\alpha}$ and for any bounded stopping time T. Hence, $\hat{\boldsymbol{\beta}}$ is in this sense an *unbiased* estimator of $\boldsymbol{\beta}^*$. The martingale property also implies that processes $\hat{\beta}_i - \beta_i^*$ have *uncorrelated increments*.

The orthogonality implies that for any s, t and $i \neq j$ $\hat{\beta}_i(t) - \beta_i^*(t)$ is uncorrelated with $\hat{\beta}_j(s) - \beta_j^*(s)$.

(i) implies that for any bounded stopping time T and for all $\boldsymbol{\alpha}$

$$E[\hat{\beta}_i(T) - \beta_i^*(T)]^2 = E[\int_0^T \alpha_i(s) J_i(s)(Y_i(s))^{-1}\, ds] \qquad i = 1, \cdots, k .$$

If T equals a fixed time t we get

$$\tau_i(t) = E[\hat{\beta}_i(t) - \beta_i^*(t)]^2 = \int_0^t \alpha_i(s) E[J_i(s)(Y_i(s))^{-1}]\, ds .$$

This last function we will call the *mean squared error function* of the estimator $\hat{\beta}_i$.

When Theorem 5.1 applies any random variable measurable with respect to $\mathscr{N}$ and with expectation equal to $E\beta^*(t_0)$ for all $\boldsymbol{\alpha}$ and some t_0, will be almost surely equal to $\hat{\beta}(t_0)$ for all $\boldsymbol{\alpha}$.

PROOF. The following holds:

$$\hat{\beta}_i(t) - \beta_i^*(t) = \int_0^t J_i(s)(Y_i(s))^{-1}\, dM_i(s) \qquad i = 1, \cdots, k ,$$

where M_i, $i = 1, \cdots, k$, are the martingales from Theorem 3.2 and the integral is a Stieltjes integral. The condition (3.1) for equivalence between Stieltjes and stochastic integrals is fulfilled in this case by Assumptions 3.1 and 6.1. Hence, the theory of stochastic integrals may be applied, see Section 2. The theorem follows immediately. □

Clearly, a theorem like the one above could be proved without applying the theory of stochastic integrals. However, the proof would be a long and tedious one, and one might have to impose stronger conditions. See Aalen (1976) for such a proof applied to a special case.

6.2. *Estimation of the mean squared error function.* Assumption 6.1 will be supposed to hold. We want to estimate the functions $\tau_i(t)$, $i = 1, \cdots, k$.

Define

$$\hat{\tau}_i(t) = \int_0^t J_i(s) Y_i(s)^{-2}\, dN_i(s) \qquad i = 1, \cdots, k .$$

Note that $\tau_i(t)$ is the expectation of

$$\hat{\tau}_i(t) - \int_0^t J_i(s)(Y_i(s))^{-2}\, dM_i(s) .$$

The last integral is a Stieltjes integral, but by (3.1) and Assumption 6.1 it coincides with the corresponding stochastic integral. Hence, for each t and i $\hat{\tau}_i(t)$ is an unbiased estimator of $\tau_i(t)$. When Theorem 5.1 applies it is the unique such estimator.

6.3. *Consistency.* Assume that we have a sequence of counting processes $\mathbf{N}_n = (N_{1,n}, \cdots, N_{k,n})$, each one satisfying Assumption 3.1 and having an intensity process $\mathbf{\Lambda}_n$. Assume that each $\mathbf{\Lambda}_n$ can be factorized as described in Section 4 with the function $\boldsymbol{\alpha}$ being the same for all n. Let $\mathbf{J}_n$, $\mathbf{Y}_n$, $\hat{\boldsymbol{\beta}}_n$, and $\boldsymbol{\beta}_n{}^*$ have the obvious definitions. The sequence of processes may come about in any number of ways. However, in order to get asymptotic results we will put conditions on the sequence which insure that the intensity process increases with n over the whole time interval [0, 1], i.e., the number of jumps in the counting process becomes large on each subinterval of [0, 1]. Assumption 6.1 will be supposed to hold for each n, however the constant c is allowed to vary with n.

PROPOSITION 6.3. *Assume*

$$\int_0^1 \alpha_i(s) E[J_{i,n}(s)(Y_{i,n}(s))^{-1}]\, ds \to 0 \qquad i = 1, \cdots, k .$$

Then

$$E[\sup_{t \in [0,1]} (\hat{\beta}_{i,n}(t) - \beta_{i,n}^*(t))^2] \to 0 \qquad i = 1, \cdots, k .$$

The limits are taken with respect to n.

PROOF. We apply the semi-martingale inequality in Theorem 3.4 of Doob (1953) (Doob states on page 354 that it is valid in the continuous case too):

$$E[\sup_{t \in [0,1]} (\hat{\beta}_{i,n}(t) - \beta_{i,n}^*(t))^2] \leqq 4E(\hat{\beta}_{i,n}(1) - \beta_{i,n}^*(1))^2 .$$

The proposition follows immediately from formula (i) of Theorem 6.2. □

6.4. *Asymptotic normality.* There is a large literature proving that martingales under certain conditions converge to normal processes with independent increments. In Aalen (1977) we have used the theory of McLeish (1974) to prove weak convergence of stochastic integrals with respect to the martingales M_i. We will apply that result to the processes $\hat{\beta}_i - \beta_i{}^*$. We know that $\hat{\beta}_i - \beta_i{}^*$ coincides with the stochastic integral $\int H_i\, dM_i$ where $H_i = J_i Y_i{}^{-1}$. Before we consider the sequence of processes of the previous section we will first look at one single process. We have to ensure that Requirement B of Aalen (1977) is fulfilled. To be specific, we will give explicitly a set of simpler assumptions which ensures this to be the case. Those assumptions seem to cover most practical applications, but they are of course in no way necessary for the weak convergence to take place.

We will assume that α_i, $i = 1, \cdots, k$, are of bounded variation on [0, 1]. Let $\phi(t)$ be a nonnegative nondecreasing function satisfying

$$|\alpha_i(t) - \alpha_i(s)| \leqq \phi(t) - \phi(s)$$

for $0 \leqq s < t \leqq 1$ and $i = 1, \cdots, k$. We will also assume that each sample function of the processes $J_i Y_i^{-1}$, $i = 1, \cdots, k$, is piecewise constant and has a finite number of jumps on [0, 1]. Let us denote by $R(t)$ the total number of jumps on $[0, t]$ of the processes $J_i Y_i^{-1}$, $i = 1, \cdots, k$.

Assume ($\bar{N}$ and $\bar{\Lambda}$ are defined in Section 3.3):

$$E(R(1)\bar{N}(1)) < \infty ,$$
$$E(R(1) \textstyle\int_0^1 \bar{\Lambda}(s)\, ds) < \infty .$$

Define

$$Z(t) = cR(t) \sup_{i,t} \alpha_i(t) + c\phi(t) .$$

The supremum in the last expression is finite by the assumptions made above. The variation in the process H_i is either due to the jumps in the process $J_i Y_i^{-1}$ or it is due to the variation in α_i. Hence, part (i) of Requirement B of Aalen (1977) is fulfilled with the process Z we have defined here. Parts (ii), (iii), and (iv) also follow immediately from the assumptions we have made.

We now reintroduce the sequence of counting processes of Section 6.3. We will assume that the conditions stated in the first part of this section are fulfilled for each n.

Let D be the space of real functions on [0, 1] which are right-continuous and have left-hand limits. Let D be equipped with the Skorohod topology (see Billingsley (1968), Chapter 3). Denote by D^k the Cartesian product of D with itself k times and let D^k be equipped with the product topology. By $\Rightarrow$ we will denote weak convergence of random elements of D^k with respect to the given topology. See Billingsley (1968) for the theory of weak convergence.

The next theorem is a consequence of Theorem 2.1 of Aalen (1976b). $\rightarrow_p$ denotes convergence in probability. $\bar{\Lambda}_n$ and $\bar{N}_n$ are defined as in Section 3.3. Let $S_{i,n}^{(m)}$, $m = 1, 2, \cdots, N_{i,n}(1)$, be the successive jump times of the process $N_{i,n}$. Define

$$Z_{i,n}^{(m)} = J_{i,n}(S_{i,n}^{(m)})(Y_{i,n}(S_{i,n}^{(m)}))^{-1} .$$

The limits in the following theorem are taken with respect to n.

THEOREM 6.4. *Assume that there exists a sequence of positive constants $\{a_n\}$ increasing to infinity such that the following holds.*

(a) *There exist nonnegative functions $g_i \in L^2(0, 1)$, $i = 1, \cdots, k$, such that*

$$a_n^2 \int_0^t \alpha_i(s) J_{i,n}(s) Y_{i,n}(s)^{-1}\, ds \rightarrow_p \int_0^t g_i^2(s)\, ds$$

for each $t \in [0, 1]$ and $i = 1, \cdots, k$.

(b) *For every* $\varepsilon > 0$:

$$E \sum_{i=1}^{k} \sum_{m=1}^{N_{i,n}(1)} [(a_n Z_{i,n}^{(m)})^2 I(a_n Z_{i,n}^{(m)} > \varepsilon)] \to 0 .$$

Let $W_1, \cdots, W_k$ *be independent Wiener processes and put* $X_i(t) = \int_0^t g_i(s)\, dW_i(s)$ *and* $\mathbf{X} = (X_1, \cdots, X_k)$. *Then*

$$a_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n^*) \Rightarrow \mathbf{X} .$$

REMARK. Condition (a) guarantees the stabilization of the variance which is needed to get a normal process in the limit. Condition (b) is a sort of Lindeberg condition. It guarantees that the jumps of the process disappear in the limit. Note that this condition implies that the Y-processes become large, i.e., the number of jumps in the counting process becomes large.

A simple Markov chain example with explicit verification of the conditions in the theorem is given in Aalen (1977).

Finally, let $\hat{\tau}_{i,n}(t)$ be defined relatively to $(\mathbf{N}_n, \mathbf{Y}_n)$ as in Section 6.2.

PROPOSITION 6.5. *Suppose that condition* (a) *of Theorem* 6.4 *holds and assume in addition*

(i) $$a_n^4 \int_0^1 E(J_{i,n}(s) Y_{i,n}(s)^{-3}) \alpha_i(s)\, ds \to 0$$

for $i = 1, \cdots, k$. *Then*

$$a_n^2 \hat{\tau}_{i,n}(t) \to_p \int_0^t g_i^2(s)\, ds$$

for each $t \in [0, 1]$ *and* $i = 1, \cdots, k$.

PROOF. Consider the expression:

$$U_{i,n}(t) = \hat{\tau}_{i,n}(t) - \int_0^t J_{i,n}(s) Y_{i,n}(s)^{-1} \alpha_i(s)\, ds .$$

Applying condition (a) of Theorem 6.4 we see that it is enough to prove that $a_n^2 U_{i,n}(t) \to_p 0$ for all i and t. $U_{i,n}(t)$ is equal to the Stieltjes integral

$$\int_0^t J_{i,n}(s) Y_{i,n}(s)^{-2}\, dM_{i,n}(s) .$$

By (3.1) and Assumption 6.1 this coincides with the corresponding stochastic integral. Hence the processes $U_{i,n}$ are square integrable martingales and

$$\langle U_{i,n}, U_{i,n} \rangle (t) = \int_0^t J_{i,n}(s) Y_{i,n}(s)^{-3} \alpha_i(s)\, ds .$$

We now use the same argument as in the proof of Proposition 6.3. It follows by condition (i) that $a_n^2 U_{i,n}(t) \to_p 0$ for all i and t. □

7. Nonparametric comparison of two counting processes.

7.1. *A general test statistic.* We consider again the multiplicative intensity model of Section 4. We want to test whether $\alpha_1(t) = \alpha_2(t)$ over a specified random subinterval of the time interval [0, 1]. Denote this subinterval by R^* and let $R(t)$ be its indicator function. $R(t)$ is a stochastic process and we require that it be adapted to $\{\mathscr{F}_t\}$ and predictable. The hypothesis that α_1 and α_2 coincide on R^* is denoted by H_0.

One should note that the comparison of two random samples with or without censoring is a special case of our general setup (see Examples 1 and 2 of Section 4).

Assumption 6.1 is supposed to be in force throughout this section. The processes $\hat{\beta}_1$ and $\hat{\beta}_2$ from Section 6.1 are then well defined. If H_0 is true, then $\hat{\beta}_1 - \hat{\beta}_2$ is a square integrable martingale. Let $K(t)$ be a process which is a member of $L^2(\hat{\beta}_1 - \hat{\beta}_2)$ under H_0 and which also satisfies the following conditions whether H_0 is true or not (the integral is a Stieltjes integral):

$$E(\int_0^1 K(s)R(s)(d\hat{\beta}_1(s) + d\hat{\beta}_2(s))) < \infty . \tag{7.1}$$

The following Stieltjes integral is well defined

$$Z = \int_0^1 K(s)R(s)(d\hat{\beta}_1(s) - d\hat{\beta}_2(s)) . \tag{7.2}$$

When H_0 is true, then Z is also well defined as a stochastic integral. By (7.1) and by Proposition 3 of Doléans-Dadé and Meyer (1970) the Stieltjes and stochastic integral interpretation of (7.2) coincide.

Under H_0 we have $EZ = 0$. When H_0 is not true Z may be used as a measure of how much α_1 differs from α_2 on R^*. According to how K is chosen Z will be more or less sensitive (or completely insensitive) to different kinds of deviations from H_0. By choosing different processes K and using Z as a test statistic one generates a large class of different tests. Later on we will have a closer look at this class in relation to the life testing example and show that it includes most two sample rank tests for censored and uncensored random samples.

In general one will have to rely on large sample theory in order to actually perform the testing. Since Z can be expressed as a stochastic integral under H_0 we can, as in Section 6.4, use the theory of Aalen (1977) to prove asymptotic normality of Z when H_0 holds. Since one can make similar assumptions and a similar statement to that of Section 6.4, we will not go further into the details here.

The weak convergence theory of stochastic integrals can also be used to study the asymptotic power under contiguous alternatives. This is done in a special case in Aalen (1975) where for $R^* = [0, 1]$ we have proved that the test based on Z with $K = (Y_1 + Y_2)^{-1} Y_1 Y_2$ is asymptotically most powerful similar against the alternatives $\alpha_1(t) = \theta\alpha_2(t)$ for $\theta > 1$. The proof of that result will be published elsewhere.

7.2. *An estimator of the variance under H_0.* We can get an unbiased estimator of Var Z under H_0 in the following way. Assume that H_0 holds. Denote the common value of α_1 and α_2 by α. Write:

$$\begin{aligned} Z(t) &= \int_0^t K(s)R(s)(d\hat{\beta}_1(s) - d\hat{\beta}_2(s)) \\ &= \int_0^t K(s)R(s)(Y_1(s)^{-1}\, dM_1(s) - Y_2(s)^{-1}\, dM_2(s)) . \end{aligned}$$

By Theorem 3.2 and (2.1) we have:

$$\langle Z, Z\rangle(t) = \int_0^t K^2(s)R(s)(Y_1(s)^{-1} + Y_2(s)^{-1})\alpha(s)\, ds .$$

We have Var $Z = E\langle Z, Z\rangle(1)$. Consider the following process:

$$V(t) = \int_0^t K^2(s)Y_1(s)^{-1}Y_2(s)^{-1}(dN_1(s) + dN_2(s))$$

and assume $EV(1) < \infty$. We have:

$$V(t) - \langle Z, Z\rangle(t) = \int_0^t K^2(s)R(s)Y_1(s)^{-1}Y_2(s)^{-1}(dM_1(s) + dM_2(s))$$

where the integral is a Stieltjes integral. It is well defined as a stochastic integral if

$$E\int_0^t K^4(s)R(s)Y_1(s)^{-1}Y_2(s)^{-1}(Y_1(s) + Y_2(s))\alpha(s)\,ds < \infty .$$

The assumptions we have made ensure by Proposition 3 of Doléans-Dadé and Meyer (1970) that the stochastic and Stieltjes integral coincide. Hence $V(1)$ *is an unbiased estimator of* Var Z.

One should note that all assumptions made about the process K in this section will hold automatically if we assume that K is bounded by a constant with probability 1. It seems that this will usually hold in applications.

When H_0 is true then (N_1, Y_1) and (N_2, Y_2) "collapse" by a sufficiency reduction into $(N_1 + N_2, Y_1 + Y_2)$. It is therefore clear that in general the variance estimator $V(1)$ will not be based on statistics that are minimal sufficient under H_0. This is a drawback but preliminary investigations indicate that it means little in practice. Also, since we usually have to base the testing on asymptotic theory it is only the consistency of the variance estimator that really matters. A consistency result can be proved along the lines of Section 6.4.

7.3. *Application to the life testing model.* Consider Example 1 of Section 4. Assume that we have two independent random samples of the kind described there. Let n_1 and n_2 denote the sizes of the two samples and let F_1 and F_2 correspond to F in Example 1. All assumptions of Example 1 will be in force except that we will allow for $F_1(1)$ and $F_2(1)$ to be equal to 1. We will however require $F_i(x) < 1$, $i = 1, 2$, whenever $x < 1$. It is easily seen that Lemma 4.1 is still valid on the whole of $[0, 1]$ for each of the samples. Let α_1 and α_2 be the hazard rates for the two distributions and let (N_1, Y_1) and (N_2, Y_2) be defined as $(\bar{N}, Y)$ of Example 1. We have:

$$Y_i(t) = n_i - N_i(t-) \qquad i = 1, 2 . \tag{7.3}$$

If $F_i(1) = 1$ then α_i is not integrable on $[0, 1]$ and so the assumptions of the multiplicative intensity model are not formally satisfied. However, in that case all jumps in N_i have to take place before $t = 1$. Hence, considering the quantities defined in Section 6.1, we see that even if $\beta_i(1)$ may be infinite, $\beta_i{}^*(1)$ will always be finite with probability 1. It is easily checked that the estimation performed in Section 6.1 is still valid.

Consider the situation of Section 7.1 with $R^* = [0, 1]$ and $\{\mathscr{F}_t\}$ generated by (N_1, N_2). If $F_i(1) = 1$, $i = 1, 2$, a general *linear rank statistic* for comparison of two samples can be written

$$U = \int_0^1 L^*(Y_1(s) + Y_2(s))\, dN_1(s) \tag{7.4}$$

where L^* is the score function.

PROPOSITION 7.1. *Assume $F_i(1) = 1$, $i = 1, 2$. Let L be a function such that the process*

$$K = L(Y_1 + Y_2)Y_1 Y_2 \tag{i}$$

satisfies the assumptions made in Section 7.1. *Substitute* (i) *and $R \equiv 1$ in* (7.2). *Then Z is a linear rank statistic as defined by* (7.4). *In particular, one gets the Wilcoxon test by choosing*

$$K = Y_1 Y_2 \tag{ii}$$

and the Savage test by choosing

$$K = (Y_1 + Y_2)^{-1} Y_1 Y_2 . \tag{iii}$$

REMARK. See, for instance, Hájek and Šidák (1968) for the theory of rank tests. Note that for linear rank tests the distribution of the test statistic under H_0 is independent of α_1 and α_2. Hence the variance can be "estimated" without error. Using the estimator of Section 7.2 we get an estimator *with* random variation. See the remarks at the end of Section 7.2 for the explanation of this.

PROOF. Dropping the integration variable from the notation we have with K as in (i):

$$\begin{aligned} Z &= \int_0^1 L(Y_1 + Y_2)(Y_2\, dN_1 - Y_1\, dN_2) \\ &= \int_0^1 L(Y_1 + Y_2)(Y_1 + Y_2)\, dN_1 - \int_0^1 L(Y_1 + Y_2) Y_1\, d(N_1 + N_2) \\ &= \int_0^1 L(Y_1 + Y_2)(Y_1 + Y_2)\, dN_1 + \int_0^1 \int_0^t L(Y_1 + Y_2)\, d(N_1 + N_2)\, dY_1 . \end{aligned}$$

The last integral is arrived at by partial integration. Applying (7.3) we can substitute dY_1 by $-dN_1$. Finally, we get:

$$Z = \int_0^1 L'(Y_1 + Y_2)\, dN_1(s)$$

where

$$L'(\nu) = L(\nu)\nu - \textstyle\sum_{\nu < j \leq n_1 + n_2} L(j) .$$

It is straightforward to show that (ii) and (iii) produces the Wilcoxon and Savage test respectively. □

Consider now the general life testing model in Example 2 of Section 4. It turns out that most rank tests suggested for comparison of two censored random samples are also special cases of the statistic Z. We then interpret $Y_i(t)$, $i = 1, 2$, as the observed numbers at risk at time t in the two samples while N_1 and N_2 count the observed failures.

Putting $R = 1$ and using the choice (ii) of Proposition 7.1 produces the Wilcoxon test generalizations of Halperin (1960), Gehan (1965), and Gastwirth (1965).

Using the choice (iii) produces the Savage test generalizations of Rao et al. (1960), Basu (1968), and Thomas (1969, 1971). This is also related to the paper of Cox (1972). The Savage test generalization of Gastwirth (1965) occurs by putting

$$K = -Y_1 Y_2 \log(1 - (Y_1 + Y_2)^{-1})$$

which is very close to (iii). The test proposed by Efron (1967) occurs by putting $K = \hat{P}_1 \hat{P}_2$ where $\hat{P}_i(t)$, $i = 1, 2$, are the product limit estimates (Kaplan and Meier, 1958) of $\exp(-\int_0^t \alpha_i(s)\, ds)$, $i = 1, 2$. This test is another generalization of the Wilcoxon test. Finally, we will mention the statistic given by Crowley (1974). It is a further generalization of the Savage statistic specifically designed for a heart transplant study. Our counting process formulation will apply in that situation too and we get once more a special case of Z with K given as in (iii).

The claims made in the preceding paragraph can be verified by straightforward computations of the kind made in the proof of Proposition 7.1. The advantage of our approach is, of course, the fact that it gives a unifying theory. The disadvantage is that we do not get exact distribution results of the kind one may get for simple censoring schemes.

8. An application. The sexual behavior of Drosophila has been the object of some biological studies (Christiansen 1971). A mathematical model for such studies has been proposed by Barndorff-Nielsen (1968) and Bartlett and Jennifer (1971). We will give a short description of it.

A number of male and virgin female Drosophila are introduced at time 0 into an observation chamber called a "pornoscope." The flies are observed continuously over a time interval, and one records all times at which copulations are initiated or terminated.

Let $M(t)$ and $F(t)$ be the numbers of male and female flies which have at time t not participated in any mating. Let $N(t)$ be the number of matings initiated in the time interval $[0, t]$. N is a counting process and we assume that it has an intensity process $\alpha(t)M(t)F(t)$ where α satisfies the same assumptions as in Example 1 and the sample functions of M and F are supposed to be left-continuous.

We have again a special case of the multiplicative intensity model. The process (N, M, F) is also Markovian; however, that is not the case with the following modification which has in some cases turned out to be a more realistic description: We assume that $M(t)$ consists not only of those males who have not initiated a copulation at time t, but also of those who may already have participated in one or more copulations but who are free at time t. Clearly the process (N, M, F) is not any longer Markovian since at each time t it is of importance how long the ongoing copulations have lasted.

To illustrate the use of the empirical β-function from Section 6.1 we will apply it to a set of data. The data are unpublished and come from an experiment carried out by Freddy Christiansen at the University of Aarhus, Denmark.

We will use the first model mentioned above where each fly only mates once.

We use data from two experiments, one with so-called ebony flies and the other with oregon flies. In the experiment with ebony flies 40 males and 30 virgin females are introduced into the pornoscope at time 0. Let the processes N_1, M_1, F_1 and the function α_1 be defined as above. We put $Y_1 = M_1F_1$. In the oregon fly experiment we start out with 39 males and 29 females. N_2, Y_2, and α_2 are defined analogously to above. In both experiments the flies are observed continuously for 45 minutes, and then observation is continued until no mating is observed to take place or until the experiment has lasted for one hour.

The processes $\hat{\beta}_1$ and $\hat{\beta}_2$ can be constructed relative to the functions α_1 and α_2 as done in Section 6.1. In this example it is natural to call α_1 and α_2 the mating intensities and $\hat{\beta}_1$ and $\hat{\beta}_2$ the empirical cumulative mating intensities.

The times at initations of matings for the two experiments are given in Table 1. The values of $\hat{\beta}_1$ and $\hat{\beta}_2$ *at the jump times* are plotted in Figure 1. One sees that the plot gives a picture of the development of the mating in the two populations.

When statistical inference is to be made from the plot, one needs estimates of the "error variation" of the curves. As for the usual cumulative distribution functions it is possible to construct confidence bands for β_1 and β_2. That subject

TABLE 1
Times in seconds at initiations of mating

Ebony flies	143, 180, 184, 303, 380, 431, 455, 475, 500, 514, 521, 552, 558, 606, 650, 667, 683, 782, 799, 849, 901, 995, 1131, 1216, 1591, 1702, 2212.
Oregon flies	555, 742, 746, 795, 934, 967, 982, 1043, 1055, 1067, 1081, 1296, 1353, 1361, 1462, 1731, 1985, 2051, 2292, 2335, 2514, 2570, 2970.

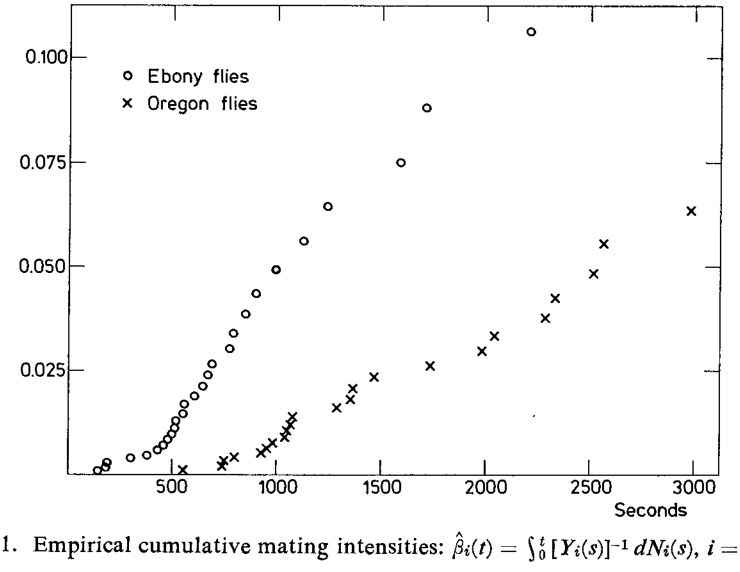

FIG. 1. Empirical cumulative mating intensities: $\hat{\beta}_i(t) = \int_0^t [Y_i(s)]^{-1}\, dN_i(s)$, $i = 1, 2$.

will be pursued in a later paper. Here, we will be satisfied with the pointwise error estimates $e_i(t) = [\hat{\tau}_i(t)]^{\frac{1}{2}}$, $i = 1, 2$, (see Section 6.2). The following values are sufficient to indicate the sizes of these estimates (the times are given in seconds):

$$e_1(1000) = .0124\,, \qquad e_1(1500) = .0165\,,$$
$$e_2(1000) = .0029\,, \qquad e_2(1500) = .0058\,.$$

Since confidence bands for the β_i will be considerably wider than pointwise confidence intervals, it should be clear from inspection of Figure 1 that a hypothesis of constant mating intensities after the first 500 seconds cannot be rejected.

One also sees that the mating intensity for the ebony flies seems to be considerably higher than that of the oregon flies for most of the observation period. The theory of Section 7 can be used to test whether the difference is significant. We let the period of 3000 seconds correspond to the time interval [0, 1]. We use the statistic Z of (7.1) with $R \equiv 1$ and the simple choice $K = Y_1 Y_2$. Using the variance estimator $V(1)$ of Section 7.2, we can compute $ZV(1)^{-\frac{1}{2}} = 5.44$. Assuming that the normal approximation is reasonable we conclude that the observed difference is strongly significant.

Acknowledgment. The main results in this paper are from the author's Ph.D. dissertation written under the supervision of Professor Lucien Le Cam. I am also grateful to Jan M. Hoem, Søren Johansen, Niels Keiding and Mats Rudemo for helpful discussions. Freddy Christiansen has kindly permitted the use of data from his experiments.

APPENDIX

We will prove Theorem 3.2 as an extension of results in Boel et al. (1975a, Proposition 3.2 and Lemma 3.1). Our proof is in part taken from Dolivo (1974, Theorem 2.4.8).

By Assumption 3.1, part (i), N_i, $i = 1, \cdots, k$, are positive submartingales. Hence the existence of unique natural increasing processes A_i, $i = 1, \cdots, k$, such that $M_i = N_i - A_i$, $i = 1, \cdots, k$ are martingales follows from the Doob–Meyer decomposition theorem (Meyer (1966)).

Let i be fixed. We have to prove that the martingale M_i is square integrable. Let T_n, $n = 1, 2, \cdots$, be the jump times of $\mathbf{N}$. Define $t \wedge T_n = \min(t, T_n)$. By Boel et al. (1975a, Proposition 3.2) the martingale $M_i(t \wedge T_n)$ is square integrable with

$$EM_i^2(t \wedge T_n) = EN_i(t \wedge T_n)\,.$$

Since $M_i(t \wedge T_n)$ converges to $M_i(t)$ when $n \to \infty$ we have by Fatou's lemma:

$$EM_i^2(t) \leq \lim_{n\to\infty} EN_i(t \wedge T_n) = EN_i(t)\,.$$

M_i^2 is then a positive submartingale and so by the Doob–Meyer decomposition theorem there is a unique natural increasing process B_i such that $M_i^2 - B_i$ is a martingale. Hence $B_i(t \wedge T_n)$ is the unique natural increasing process such that $M_i^2(t \wedge T_n) - B_i(t \wedge T_n)$ is a martingale. By Boel et al. (1975a, Lemma 3.1) $B_i(t \wedge T_n)$ coincides with $A_i(t \wedge T_n)$ for each n. Hence

$$B_i(t) = \lim_{n\to\infty} B_i(t \wedge T_n) = \lim_{n\to\infty} A_i(t \wedge T_n) = A_i(t) .$$

It only remains to prove formula (ii) in Theorem 3.2. Let $i \neq j$ be fixed. Then $N^* = N_i + N_j$ is a counting process and clearly $A^* = A_i + A_j$ is the unique natural increasing process making $M^* = N^* - A^*$ a martingale. By what we have already proved $(M^*)^2 - A^*$ is a martingale. But

$$(M^*)^2 - A^* = (M_i + M_j)^2 - A_i - A_j = 2M_iM_j + M_i^2 - A_i + M_j^2 - A_j .$$

Hence M_iM_j is a martingale, i.e., M_i and M_j are orthogonal.

REFERENCES

[1] AALEN, O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation, Univ. of California, Berkeley.

[2] AALEN, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian J. Statist.* **3** 15–27.

[3] AALEN, O. (1977). Weak convergence of stochastic integrals related to counting process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **38** 261–277.

[4] ALTSHULER, B. (1970). Theory for the measurement for competing risks in animal experiments. *Math. Biosciences* **6** 1–11.

[5] BARNDORFF-NIELSEN, O. (1968). Stochastic models for pornoscopes. Technical report, Univ. of Aarhus.

[6] BARTLETT, M. S. and JENNIFER, M. BRENNAN (1971). Stochastic analysis of some experiments on the mating of blowflies. *Biometrics* **27** 725–730.

[7] BASU, A. P. (1968). On a generalized Savage statistic with applications to life testing. *Ann. Math. Statist.* **38** 905–915.

[8] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[9] BOEL, R., VARAYIA, P. and WONG, E. (1975a). Martingales on jump processes I: Representation results. *SIAM J. Control* **13** 999–1021.

[10] BOEL, R., VARAYIA, P. and WONG, E. (1975b). Martingales on jump processes II: Applications. *SIAM J. Control* **13** 1022–1061.

[11] BREMAUD, P. (1972). A martingale approach to point processes. Memorandum ERL-M345, Univ. California, Berkeley.

[12] BREMAUD, P. (1974). The martingale theory of point processes over the real half line admitting an intensity. *Proc. of the IRIA Coll. on Control Theory.* Lect. Notes (grey) **107** Springer-Verlag. 519–542.

[13] BRESLOW, N. (1970). A generalized Kruskal-Wallis test for comparing samples subject to unequal patterns of censoring. *Biometrika* **57** 579–594.

[14] CHRISTIANSEN, F. B. (1971). On the measurement of the sexual behavior of Drosophila Melanogaster. Technical report, Univ. of Aarhus.

[15] COURRÈGE, P. (1963). Intégrales stochastiques et martingales de carré intégrable. *Séminaire Brelot-Choquet-Deny: Théorie du Potentiel.* Paris, Université, Faculté des Sciences. 7e année.

[16] COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

[17] COX, D. R. and LEWIS, P. A. W. (1972). Multivariate point processes. *Proc. Sixth Berkeley Symp. Math. Statist. Probability* **3** 401–448.

[18] Crowley, J. (1974). Asymptotic normality of a new nonparametric statistic for use in organ transplant studies. *J. Amer. Statist. Assoc.* **69** 1006–1011.

[19] Davis, M. H. A. (1976). The representation of martingales of jump processes. *SIAM J. Control.* **14** 623–638.

[20] Doléans-Dadé, C. and Meyer, P. A. (1970). Intégrales stochastiques par rapport aux martingales locales. *Séminaire de probabilités IV. Lecture Notes in Mathematics* **124** 77–107. Springer-Verlag, Berlin.

[21] Dolivo, F.-B. (1974). Counting processes and integrated conditional rates: A martingale approach with application to detection. Technical Report, College of Engineering, Univ. Michigan.

[22] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[23] Efron, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probability* **4** 831–853. Univ. of California Press.

[24] Efron, B. (1975). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565.

[25] Gastwirth, J. L. (1965). Asymptotically most powerful rànk tests for the two-sample problem with censored data. *Ann. Math. Statist.* **36** 1243–1247.

[26] Gehan, E. (1965). A generalized Wilcoxon test for comparing arbitrarily single censored samples. *Biometrika* **52** 203–223.

[27] Hájek, J. and Šidák, Z. (1968). *Theory of Rank Tests.* Academic Press, New York.

[28] Halperin, M. (1960). Extension of the Wilcoxon-Mann-Whitney test to samples censored at the same fixed point. *J. Amer. Statist. Assoc.* **55** 125–138.

[29] Hoem, J. M. (1971). Point estimation of forces of transition in demographic models. *J. Roy. Statist. Soc. Ser. B.* **33** 275–289.

[30] Jacod, J. (1973). On the stochastic intensity of a random point process over the half-line. Technical Report 51, Department of Statistics, Princeton Univ.

[31] Jacod, J. (1975). Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31** 235–253.

[32] Jacod, J. and Memin, J. (1976). Caractéristiques locales et conditions de continuité pour les semi-martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35** 1–37.

[33] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

[34] Keiding, N. (1977). Statistical comments to a paper by J. E. Cohen on the application of a simple stochastic population model to natural primate troops. To appear in *Amer. Naturalist.*

[35] Kunita, H. and Watanabe, S. (1967). On square integrable martingales. *Nagoya Math. J.* **30** 209–245.

[36] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[37] Martins-Neto, A. F. (1974). Martingale approach to waiting line problems. Memorandum no. ERL-M475, Univ. California, Berkeley.

[38] McFadden, R. (1965). The entropy of a point process. *SIAM J. Appl. Math.* **13** 988–994.

[39] McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probability* **2** 620–628.

[40] Meier, P. (1975). Estimation of a distribution function from incomplete observations. In *Perspectives in Probability and Statistics: Papers in Honour of M. S. Bartlett* (J. Gani, ed.). Academic Press, New York.

[41] Meyer, P. A. (1966). *Probability and Potentials.* Blaisdell, Waltham, Mass.

[42] Meyer, P. A. (1967). Intégrales Stochastiques, I, II, III et IV. *Lectures Notes in Mathematics* **39** 77–162. Springer-Verlag, Berlin.

[43] Meyer, P. A. (1971). Square integrable martingales, a survey. *Lecture Notes in Mathematics* **190** 32–37. Springer-Verlag, Berlin.

[44] Nelson, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Tech.* **1** 27–52.

[45] PAPANGELOU, F. (1972). Integrability of expected increments of point processes and a related change of scale. *Trans. Amer. Math. Soc.* **165** 483–506.

[46] RAO, U. V. R., SAVAGE, I. R. and SOBEL, M. (1960). Contributions to the theory of rank order statistics: Two sample censored case. *Ann. Math. Statist.* **31** 415–426.

[47] RUBIN, I. (1972). Regular point processes and their detection. *IEEE Trans. Inform. Theory.* **IT-18** (5) 547–557.

[48] SEGALL, A. and KAILATH, T. (1975a). The modeling of randomly modulated jump processes. *IEEE Trans. Inform. Theory* **IT-21** (2) 135–143.

[49] SEGALL, A. and KAILATH, T. (1975b). Radon-Nikodym derivatives with respect to measures induced by discontinuous independent-increment processes. *Ann. Probability* **3** 449–464.

[50] SNYDER, D. L. (1972). Filtering and detection for doubly stochastic Poisson processes. *IEEE Trans. Inform. Theory* **IT-18** 91–102.

[51] THOMAS, D. R. (1969). Conditionally locally most powerful rank tests for the two-sample problem with arbitrarily censored data. Technical report no. 7, Department of Statistics, Oregon State Univ.

[52] THOMAS, D. R. (1971). On the asymptotic normality of a generalized Savage statistic for comparing two arbitrarily censored samples. Technical report, Department of Statistics, Oregon State Univ.

INSTITUTE OF MATHEMATICAL
AND PHYSICAL SCIENCES
UNIVERSITY OF TROMSØ
P.O. BOX 953
N-9001 TROMSØ
NORWAY