SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853

TECHNICAL REPORT NO. 676

December 1985

ESTIMATING THE COMPENSATOR FROM POISSON TYPE COUNTING PROCESSES

By

Michael Phelan

ESTIMATING THE COMPENSATOR FROM POISSON TYPE COUNTING PROCESSES

By
Michael J. Phelan
Cornell University, Ithaca, NY

Summary

Let $N = (N_1, \ldots, N_m)$ be a multivariate counting process having predictable compensator $A = (A_1, \ldots, A_m)$ so that in particular $N-A$ is a martingale. A probability model is considered where we assume A has a relatively simple form given by

$$A\{(0,t]\} = \int_{(0,t]} Y_s B\{ds\}$$

where $Y = (Y_1, \ldots, Y_m)$ is a nonnegative process and B is an arbitrary Borel measure on $(R^+, \sigma(R^+))$. When $Y \equiv \lambda > 0$ (a constant) and $B(t) = t$ then $A(t) = \lambda t$ so that N is a Poisson process. Thus we call the family of counting processes having compensator as above Poisson type counting processes. Special cases are discrete and continuous time Markov chains, models for survival analysis and the multiplicative intensity model.

In this paper we introduce this family of counting processes and discuss its elementary properties. We then consider estimation of the compensator based on observations of N and Y over a period of time. Consistency and weak convergence results are given for the estimator, thus generalizing the results of Aalen (1978) to this wider class of counting processes. The theory is illustrated by estimating the cumulative hazard rate from censored survival times having arbitrary distribution function.

Title: Estimating the Compensator from Poisson Type Counting Processes

Abbreviated Title: Compensator Poisson Type Counting Process

Key words and phrases. Point process, counting process, compensator, inference from stochastic processes, survival analysis, martingales, stochastic integrals.

1. INTRODUCTION

We consider a counting process $N = \{N_t, t \geq 0\}$ where for each $t > 0$ N_t counts the frequency of occurrence of some event over the interval $(0,t]$. An important relative of N is its compensator function $A = \{A_t, t \geq 0\}$ where for each $t > 0$ A_t is, roughly speaking, the cumulative rate of occurrence of events over $(0,t]$. Usually we are given a filtration $F = \{\mathcal{F}_t, t \geq 0\}$ where for each $t > 0$ $\mathcal{F}_t$ is a collection of all events observed over the interval $(0,t]$ and the compensator A of N is defined relative to F. In the martingale approach to counting processes emphasis is placed on the fact that N-A forms a (local) martingale relative to F which describes the sense in which A compensates N.

Sometimes the compensator has a relatively simple form where for each $t > 0$ A_t may be written

$$A_t = \int_{(0,t]} Y_s B\{ds\} \tag{1.1}$$

where Y satisfies a measurability requirement of a technical nature (see section 2) and B is a right continuous, nondecreasing function on $[0,\infty)$ with $B(0) = 0$. In statistical applications the process Y is typically observable whereas the function B is unknown and to be estimated.

Numerous examples of (1.1) occur in various connections as we illustrate in section 2. A special case of a general sort occurs when the function B admits a density b relative to the Lebesgue measure so that the intensity of the point process N exists. In this case the process

$$\Lambda(t) = Y_t b(t), \quad t \geq 0 \tag{1.2}$$

is called the intensity process. The model (1.2) is called the multiplicative intensity model and has been studied in detail from a statistical point of view and has been shown to provide a general framework for the study of censored survival data by Aalen (1978) and inhomogeneous Markov chains by Aalen and Johansen (1978). In addition, Andersen and Gill (1982) used this model to study Cox's proportional hazards model extended to counting processes generated by a recurring event (see also Prentice and Self (1982) for a related extension along these lines).

The counting process approach to probability models arising in statistical applications has proved a powerful tool. For example, the methods of this approach have been used to derive asymptotic normality of maximum likelihood estimators in parametric counting process models by Borgan (1984), to study the product limit estimator of an arbitrary continuous distribution function F on $[0,\infty)$ by Gill (1983) and in a general treatment of two sample clinical trials involving arbitrary censorship of survival times by Slud (1984). Of course, many other examples of the power of these methods may be cited.

In the examples cited above in one form or another the compensator function A is assumed to be continuous or special methods have been introduced to cover the discrete case (see for example section 4). In our treatment of counting processes we make no such assumptions. In order to deal with this generality we have relied on the theory of multivariate counting processes in Jacod (1975) and the general functional convergence theorems for semimartingales found in Jacod, Klopotowski and Memin (1982) and Lipster and Shiryayev (1980). The estimation theory we develop may be used to extend Aalen's (1978) earlier work to a wider class of counting process models.

In this paper we assume a basic knowledge of the martingale approach to counting processes although a sufficient review may be found in Aalen (1978). Some of the other ideas used in this paper such as predictability and local martingale are discussed, for example, in an excellent expository paper on martingales by Shiryayev (1981).

Section 2 introduces the probability model for the compensator in the family of Poisson-type counting processes and gives a number of examples. In section 3 we consider the statistical problem of estimating the compensator function. Finally, in section 4 we apply the theory developed in section 3 to estimating the cumulative risk (hazard) function in a random censorship model in survival analysis when the distribution function of the survival times is arbitrary.

2. POISSON-TYPE COUNTING PROCESSES

2.1 Definition and Examples

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $F = \{(\mathcal{F}_t),\ t \geq 0\}$ a given family of sub-σ-algebras of $\mathcal{F}$ satisfying the usual conditions (i.e. F is nondecreasing, right continuous and complete relative to P). A stochastic process $X = \{(X_t),\ t \geq 0\}$ is called F-adapted if for each t, X_t is $\mathcal{F}_t$-measurable (written $X_t \in \mathcal{F}_t$) which we denote by $X = \{(X_t, \mathcal{F}_t),\ t \geq 0\}$. An F-adapted process is also F-predictable if it is measurable with respect to the smallest σ-algebra over $\Omega \times [0,\infty)$ generated by the left-continuous F-adapted processes. The families $\mathcal{M}(F,P)$, $\mathcal{M}^2(F,P)$, $\mathcal{M}_{loc}(F,P)$, and $\mathcal{M}^2_{loc}(F,P)$ denote the classes of F-adapted processes which are uniformly integrable martingales, square integrable martingales and their local counterparts, respectively.

Let E be a discrete space of m $(m < \infty)$ distinguished "events" and denote a generic event in E by i. We take as our starting point an m-variate counting process $N = (N(1),\ldots,N(m))$ having predictable compensator function $A = (A(1),\ldots,A(m))$. Thus for each $i \in E$ and $t > 0$ $N_t(i)$ counts the number of events of type i which have occurred over the time interval $(0,t]$ and $A_t(i)$ measures the cumulative rate of occurrence of event i over $(0,t]$. In this case if $M(i) = N(i) - A(i)$ then $M(i) = \{M_t(i), F_t), t \geq 0\}$ is of class $M^2_{loc}(F,P)$ for all $i \in E$.

In this sequel we restrict our attention to a family of m-variate counting processes where the compensator function has a particularly simple form defined as follows.

Definition 2.1. Let B denote a Borel measure on $(R^+ = [0,+\infty), \sigma(R^+))$. An m-variate counting process N is called a Poisson-type counting process if for each $i \in E$ its compensator has the form:

$$(2.1) \qquad A_i\{(0,t]\} = \int_{(0,t]} Y_t(i) B_i\{dt\}, \ \forall\ t \geq 0$$

where $Y = \{(Y_t, F_t), t \geq 0\}$ is an m-dimensional nonnegative F-predictable process. □

According to definition 1.1 we view the compensator function A as a random measure defined on the Borel sets. Accordingly, we also view $M = N-A$ as a random martingale measure which is in agreement with Jacod (1975) who also calls the kernel $A\{dt\}$ the dual predictable projection of N. When $m = 1$ with $Y_t \equiv \lambda > 0$ (a constant) and $B\{(0,t]\} \equiv t$ for all t we obtain from (2.1) $A\{(0,t]\} = \lambda t$, so that N is a simple Poisson process of rate λ and thus in the general case N is called a Poisson-type point process by Lipster and Shiryayev (1978). Also, when B has a density relative to the Lebesgue measure then (2.1) reduces to the multiplicative intensity model considered by Aalen (1978).

Consider the following examples.

Example 1. Life testing: random censorship model. For fixed positive integer n, suppose X_i and U_i, $i = 1,\ldots,n$ are $2n$ independent random variables with X_i or U_i almost surely finite for each i. X_i has distribution measure F and U_i has distribution measure L. The observable random variables $\tilde{X}_i$ and δ_i are given by $\tilde{X}_i = X_i \wedge U_i$ and $\delta_i = 1(X_i \le U_i)$, where $a \wedge b = \min(a,b)$ and $1(\cdot)$ is the characteristic or indicator function of $(\cdot)$.

To analyze problems arising in this model we define the following quantities for each $i = 1,\ldots,n$:

$$N_t^i = 1(\tilde{X}_i \le t,\ \delta_i = 1),\ t \ge 0 \tag{2.2}$$

$$Y_t^i = 1(\tilde{X}_i \ge t),\ t \ge 0 \tag{2.3}$$

$$B(t) = \int_0^t (1-F(s-))^{-1}F\{ds\},\ t \ge 0, \tag{2.4}$$

where (2.4) is an ordinary Lebesgue-Stieltjes integral and $F(s-) = \lim_{t\uparrow s} F(t)$. The process $N^i = \{N_t^i,\ t \ge 0\}$ is called a simple counting process and is equal to zero until the ith observation time elapses and has not been censored while $Y^i = \{Y_t^i,\ t \ge 0\}$ is called a risk process and is equal to one as long as the ith unit remains alive or under observation. The risk process, by virtue of its left continuity, is F-predictable where F is determined by (2.5).

Let $(\Omega, \mathcal{F}, P)$ be a probability space on which $\{X_i, U_i\}$ are defined and let the filtration $F = \{\mathcal{F}_t, t \geq 0\}$ be given by

$$\mathcal{F}_t = \mathcal{F}_0 \vee \sigma\text{-}(1(\tilde{X}_i \leq s),\ \delta_i 1(\tilde{X}_i \leq s),\ s \leq t,\ i = 1,\dots,n) \tag{2.5}$$

where $\mathcal{F}_0$ contains the P-null sets of $\mathcal{F}$ and their subsets. Next define the process M^i as follows:

$$M_t^i = N_t^i - \int_0^t Y_s^i B\{ds\},\ t \geq 0. \tag{2.6}$$

According to Gill (1980), $M^i = \{(M_t^i, \mathcal{F}_t),\ t \geq 0\} \in M^2(F,P)$ and therefore N^i has unique predictable compensator A_i given by

$$A_i\{(0,t]\} = \int_{(0,t]} Y_s^i B\{ds\}. \tag{2.7}$$

Thus as A_i satisfies definition 1.1 for each $i = 1,\dots,n$, N^i is a Poisson-type counting process having compensator A_i relative to F. □

<u>Example 2</u>. <u>Compound Poisson-type Processes</u>. Let $(\Omega, \mathcal{F}, P)$ be a probability space on which we define a sequence $X = \{X_n, n \geq 1\}$ of independent random vectors and a counting process $\pi = \{\pi_t, t \geq 0\}$. Let $\{F_n, n \geq 1\}$ be a sequence of distribution measures on R^d (d an integer) and let B be a Borel measure on $(R^+, \sigma(R^+))$. We assume that for each $n = 1,2,\dots$ X_n has distribution F_n and that π has compensator B relative to its internal history $F^\pi = \{\mathcal{F}_t^\pi, t \geq 0\}$, defined by

$$F_t^{\pi} = F_0 \vee \sigma\text{-}(\pi_s\colon 0 \le s \le t),\ t \ge 0) \tag{2.8}$$

where $F_0 \subset F$ contains the P-null sets of F and their subsets. We further assume that X and π are independent processes.

Next define a process $Y = \{Y_t,\ t \ge 0\}$ by the relation

$$Y_t = \sum_{n=1}^{\pi_t} X_n,\ \forall\ t \ge 0. \tag{2.9}$$

We interpret (2.9) so that $Y_t = 0$ whenever $\pi_t = 0$. Since B is the deterministic compensator for π, according to theorem 18.9, Lipster and Shiryayev (1978), π is a process of independent increments from which it is easy to derive that Y is a process of independent increments. Further if B is continuous and $B(R^+) = \infty$ then theorem 18.10, Lipster and Shiryayev (1978) implies that π can be transformed into a Poisson process by a change of time. In this case we may view Y as a compound Poisson process up to a change of time and therefore for the general case of arbitrary B we call Y a compound Poisson-type process. We use the process Y to construct an example of a Poisson-type counting process.

For fixed E, a Borel set in R^d-{0}, we define the following sequence of stopping times $T = \{\tau_n,\ n \ge 1\}$:

$$\tau_1 = \inf\{t\colon t > 0,\ Y_t \in E\}$$

(2.10)

$$\tau_{n+1} = \inf\{t\colon t > \tau_n,\ Y_t \in E,\ Y_{t-} \in E^c\},\ n \ge 1$$

where E^c denotes the complement of E. On $\{\omega\colon Y_t(\omega) \in E^c,\ \forall\ t \ge 0\}$ set

$\tau_1 = \infty$, and for $n \geq 1$ set $\tau_{n+1} = \tau_{n+2} = \tau_{n+3} = \ldots = \infty$ on $\{\omega: Y_t(\omega) \in E^c \ \forall\ t > \tau_n\}$. Clearly T forms a point process by recording the epochs at which the process Y visits the set E. We define the counting process $N = \{N_t, t \geq 0\}$ as follows:

$$(2.11) \qquad N_t = \sum_{n\geq 1} 1(\tau_n \leq t), \ \forall\ t \geq 0.$$

Here N_t counts the frequency of distinct visits to E by Y in $(0,t]$.

Next define the filtration $F = \{\mathcal{F}_t, t \geq 0\}$ given by

$$(2.12) \qquad \mathcal{F}_t = \mathcal{F}_0 \vee \sigma\text{-}(Y_s, \pi_s: 0 \leq s \leq t), \ \forall\ t \geq 0$$

where $\mathcal{F}_0$ is defined at (2.8). It is evident that N is F-adapted and that $N = \{(N_t, \mathcal{F}_t), t \geq 0\}$ is a local submartingale with localizing sequence $T = \{\tau_n, n \geq 1\}$ defined at (2.10). Thus there exists a unique predictable compensator $A = \{(A_t, \mathcal{F}_t), t \geq 0\}$ such that N-A is a local martingale.

Let $\{Q(k), k \geq 1\}$ be an R-sequence of partitions (see for example Brown (1978) or Helland (1982)) where $Q(k) = \{t_{0,k}, t_{1,k}, \ldots\}$ with $0 = t_{0,k} < t_{1,k} < \ldots < t_{j,k}$ and $t_{j,k} \to \infty$ as $j \to \infty$, $Q(k+1) \supset Q_{(k)}$ (i.e. $Q(k+1)$ is a refinement of $Q(k)$) and $\Delta_k(t) = \max_j\{t_{j+1,k} - t_{j,k}: t_{j,k} \leq t\} \to 0$ as $k \to \infty$ for each t. For $k \geq 1$ define the process A_k by

$$(2.13) \qquad A_k(t) = \sum_{j:t_{j,k}\leq t} E(N_{t_{j+1},k} - N_{t_j,k} | \mathcal{F}_{t_j,k}), \quad t \geq 0.$$

Next consider the process A defined as follows:

$$(2.14) \qquad A_t = \int_{(0,t]} F_{\pi_{s-}+1}\{E-Y_{s-}\}1(Y_{s-} \in E^c)B\{ds\}, \quad t \geq 0$$

where $F_n\{\cdot\}$ denotes the F_n-measure of $\{\cdot\}$ and $E-y = \{x: x+y \in E\}$ and $1(\cdot)$ is the indicator function of $(\cdot)$ (i.e. $1(\cdot) = 1$ or 0 as $(\cdot)$ holds or not).

By a series of reexpressions of (2.13) and the use of lemma 2.3 (see next section) it is possible to show that $A_k(t) \to A_t$ almost surely as $k \to \infty$ and that this result holds over arbitrary R-sequences. Put $\xi_t = F_{\pi_{t-}+1}\{E-Y_{t-}\}1(Y_{t-} \in E^c)$ for $t \geq 0$ so that A_t may be written

$$(2.15) \qquad A_t = \int_{(0,t]} \xi_s B\{ds\}, \quad t \geq 0.$$

Observe that since $\xi = \{(\xi_t, \mathcal{F}_t), t \geq 0\}$ is a $\mathcal{F}_t$-predictable process $A = \{A_t, \mathcal{F}_t), t \geq 0\}$ is predictable and satisfies definition 1.1 with Y identified with ξ. For N to be a Poisson-type counting process it remains to prove the A is its compensator.

Suppose $\tilde{A} = \{(\tilde{A}_t, \mathcal{F}_t), t \geq 0\}$ is the predictable compensator for N. Then according to Murali-Roa's (1969) proof of the Meyer decomposition theorem $A_k \to \tilde{A}$ in the weak L_1 topology (i.e. $EA_k(t)1_B \to E\tilde{A}_t 1_B$ for all $B \in \mathcal{F}$). Hence $\{A_k, k \geq 1\}$ is uniformly integrable. Combining this fact with the almost sure convergence of A_k to A allows one to show that $A_t = \tilde{A}_t$ almost surely and therefore N is a Poisson-type counting process. □

Had we known a-priori that the compensator of N relative to F is calculable (see Brown (1978) for a definition) we could have gone directly to the in probability limit of $A_k(t)$ as $k \to \infty$. As this example illustrates we may extend proposition 1, Brown (1978) on the calculability of A relative to F to the class of counting processes N generated as shown above from compound Poisson-type processes.

Example 3. First passage time for a Wiener process. This example may be found also in Lipster and Shiryayev (1978). Let $W = \{(W_t, \mathcal{F}_t),\ t \geq 0\}$ be a Wiener process and $\tau = \inf\{t \geq 0: W_t = 1\}$ where $\tau = \infty$ if $\sup_{t\geq 0} W_t < 1$. We consider the simple counting process $N = \{(N_t, \mathcal{F}_t),\ t \geq 0\}$ with $N_t = 1(\tau \leq t)$ for all $t \geq 0$. By virtue of the predictability of W it is easily shown that τ is a predictable markovian (stopping) time and therefore the unique predictable compensator of N relative to F is N itself which provides us with an example of a compensator not of the Poisson-type.

Alternatively, we consider the minimal representation of $N_t = 1(\tau \leq t),\ t \geq 0$ (i.e. for $\mathcal{F}_t^N = \sigma\{\omega: N_s,\ s \leq t\}$ and $\bar{N} = \{(N_t, \mathcal{F}_t^N),\ t \geq 0\}$ we find the compensator $\bar{A} = \{(\bar{A}_t, \mathcal{F}_t^N),\ t \geq 0\}$ of $\bar{N}$). First, define the function F on $[0,\infty)$ by:

$$F(t) = \sqrt{2/\pi} \int_{t^{-1/2}}^{\infty} e^{-y^2/2} dy, \quad t \geq 0. \tag{2.16}$$

Then it is easily shown that $\bar{A}$ is given by:

$$\begin{aligned} \bar{A}_t &= -\ell n(1-F(t \wedge \tau)) \\ &= \int_0^t 1(\tau \geq s) dF(s)/(1-F(s)), \quad t \geq 0. \end{aligned} \tag{2.17}$$

Thus $\bar{A}$ is a Poisson-type counting process with the process $\{Y_t\}$ identified with $\{1(\tau \geq t)\}$ and the Borel measure B identified with the measure generated by the function $-\ell n(1-F)$. We further recognize $\bar{A}$ to be in the family of multiplicative intensity models (see Aalen (1978)) where the intensity process $\{\lambda_t\}$ is identified with $\{1(\tau \geq t)(2\pi)^{-1/2}t^{-3/2}e^{-1/2t}/(1-F(t))\}$. □

Example 4. Discrete time Markov chain. Let $X = \{X_n, n \geq 0\}$ be a homogeneous Markov chain having states $\{1,2,\ldots,m\}$ $(m < \infty)$ and probability transition matrix $P = \|p_{ij}\|_{1\leq i,j\leq m}$. For each pair (i,j), $1 \leq i,j \leq m$, let $N_{ij} = \{N_{ij}(n), n \geq 0\}$ count the number of direct transitions from state i into j realized by the Markov chain X. Also, if $\mathcal{F}_n = \sigma\text{-}(X_0,X_1,\ldots,X_n)$, $n \geq 0$, then for $1 \leq i,j \leq m$ it is easily shown that $N_{ij} = \{(N_{ij}(n),\mathcal{F}_n), n \geq 0\}$ has unique predictable compensator $A_{ij} = \{(A_{ij}(n), \mathcal{F}_n, n \geq 0\}$ given by:

$$A_{ij}(n) = \sum_{k=1}^{n} 1(X_{k-1} = i)p_{ij} \tag{2.18}$$

$$= \int_{(0,n]} 1(X_{\mu(s)-1} = i)p_{ij}\mu\{ds\}, \quad n \geq 0$$

where $\mu(s) = \mathrm{card}\{k\colon k$ is a positive integer and $k \leq s\}$. Here if we identify $\{Y_t\}$ with $\{X_{\mu(t)-1}\}$ (which is readily shown to be predictable) and B with $p_{ij}\mu$ we obtain a forth example of a Poisson-type counting process. □

2.2 Basic Properties and Preliminaries

2.2.1

In many applications the counting process N is the primary object of observation. For each $i \in E$ there is an associated sequence of stopping times $T(i) = \{\tau_n(i), n \geq 1\}$ which is the point process over R^+ recording the epochs of events of type i. Thus for each $t \geq 0$

$$N_t(i) = \sum_{n\geq 1} 1(\tau_n(i) \leq t), \quad i \in E. \tag{2.19}$$

By ordering the sequences $\{T(1),\ldots,T(m)\}$ into a new sequence $T = \{\tau_n, n \geq 1\}$ we obtain a sequence of event epochs in a marked point process having mark space E. The epochs T satisfy the following relations:

(2.20)

(1) $\tau_1 > 0$ a.s.

(2) $\tau_n < \tau_{n+1}$ a.s. on $\{\tau_n < \infty\}$

(3) $\tau_n = \tau_{n+1}$ a.s. on $\{\tau_n = \infty\}$

(4) $\bar{N}_t = \sum_{i\in E} N_t(i) = \sum_{n\geq 1} 1(\tau_n \leq t), \ \forall\, t \geq 0$

where a.s. means almost surely with respect to P. We denote the random variable $\tau_\infty = \lim_{n\uparrow\infty} \tau_n$ for the limit of the sequence T. On $\{\omega: \tau_\infty(\omega) < \infty\}$ we obtain an infinite number of events in a finite time and we say there has been an explosion in the marked point process at τ_∞. Beyond this point the process is terminated.

2.2.2

The fundamental relation in this approach to counting processes is given by the decomposition

$$N = A + m \tag{2.21}$$

where m is a local martingale which defines the sense in which the compensator process A compensates N. The localizing sequence for m is given by the sequence of event epochs T and in accordance with Lipster and Shiryayev (1978) we use the terminology τ_∞-local martingale when referring to m.

According to definition 2.1 we may view the compensator A as a random Borel measure on $(R^+, \sigma(R^+))$ or alternatively we introduce the random function $A_s = A\{(0,s]\}$, $s \geq 0$. For each $\omega \in \Omega$ $A.(\omega)$ is clearly a monotone nondecreasing right continuous function on R^+ with $A_0 = 0$. We know that A may have at most a countable number of discontinuities and we use the notation $\Delta A_s = A_s - A_{s-}$, where $A_{s-} = \lim_{h \downarrow 0} A_{s-h}$, to denote the magnitude of the jump in A at $s \geq 0$. According to theorem 2.1, Jacod (1975), for each $i \in E$ the compensator $A(i)$ satisfies $\sup_{s \leq \tau_\infty} \Delta A_s(i) = \sup_{s < \tau_\infty} \Delta A_s(i) < 1$ and $\Delta A_{\tau_\infty}(i) = 0$ with probability one. This coincides with the fact that for any t the number of points occurring at t, namely ΔN_t, is identically zero or one and anticipates a latter result which allows us to interpret $\Delta A_t > 0$ as the conditional probability $\Delta N_t = 1$ given F_{t-}.

2.2.3

For each $i \in E$ the decomposition in (2.21) uniquely defines a local martingale $m(i) = \{(m_t(i), F_t),\ t \geq 0\}$ and the family of processes $\{m(i),\ i \in E\}$ induced by (2.21) is a family of τ_∞-locally square integrable local martingales. The (local) square integrability of $m(i)$ and $m(j)$,

$i,j \in E$, allows the definition of a covariance process $<m(i),m(j)>$ defined by the relation

$$(2.22) \qquad m(i)m(j) - <m(i),m(j)> \in \mathcal{M}_{loc}(F,P).$$

The process $<\cdot>$ is called the quadratic characteristic and as it turns out bears a definite relation to the compensator function A. In particular we have the following lemma.

Lemma 2.1. The family $\{m(i),\ i \in E\}$ of τ_∞-locally square integrable local martingales defined by (2.21) has the unique quadratic characteristic given by:

$$(2.23) \qquad <m(i),m(j)> = \begin{cases} \int_{(0,t]} (1-\Delta A_s(i))dA_s(i), & i = j \\ -\int_{(0,t]} \Delta A_s(i)dA_s(j), & i \neq j \end{cases}$$

for all $i,j \in E$ and $t \geq 0$. In particular we have $m(i)m(j) - <m(i),m(j)>$, is a τ_∞-local martingale.

Proof. For $i = j$ (2.23) follows immediately from corollary 18.12 and lemma 18.12, Lipster and Shiryayev (1978). For $i \neq j$ the proof of (2.23) can be obtained with the applications of theorem 5.2, lemma 18.7 and theorem 18.8, Lipster and Shiryayev (1978). In both cases uniqueness follows from the predictability of A. □

2.2.4

In applications of counting processes, such as the inference problems we consider here, there arises the need to treat integrals of the form

$$(2.24) \qquad M_t = \int_{(0,t]} X_s(dN_s - dA_s)$$

$$= \int_{(0,t]} X_s dm_s \quad \text{on} \quad \{\omega\colon\ t < \tau_\infty(\omega)\}$$

where $X = \{(X_t, F_t),\ t \geq 0\}$ is a predictable process and for each t the integral in (2.24) is interpreted for each $\omega \in \{\omega\colon\ t < \tau_\infty(\omega)\}$ as a pathwise Lebesgue-Stieltjes integral. The question arises as to when the integral at (2.24) agrees with the stochastic integral of X relative to m meaning therefore that $M = \{(M_t, F_t),\ t \geq 0\}$ is a (local) martingale. For the purposes of this discussion it suffices to recall theorem 18.7, Lipster and Shiryayev (1978) to answer this question.

<u>Lemma 2.2</u>. Suppose $X = \{(X_t, F_t),\ t \geq 0\}$ is a predictable process such that $P(|X_t| < \infty) = 1$, $t \geq 0$ and let $M = \{(M_{t\wedge\tau_\infty}, F_t),\ t \geq 0\}$. Then

(a) $E\int_0^{\tau_\infty} |X_s| dA_s < \infty$ implies M is a uniformly integral martingale,

(b) $p(\int_0^t |X_s| dA_s = \infty,\ t < \tau_\infty) = 0$ implies M is a τ_∞-local martingale.

The Lemma is merely a statement of theorem 18.7, Lipster and Shiryayev (1978) and will not be proved.

2.2.5

We will apply lemma 2.2 to exhibit a further decomposition of the counting process N. For each $t > 0$ we introduce the notation $B(t) = B\{(0,t]\}$ with $B(0) = 0$, where B is the Borel measure introduced at definition 1.1. The function B is a monotonic increasing function of local bounded variation (i.e. B is of bounded variation on bounded sets of the form $[0,t]$, $t > 0$).

Let $A = \{t: t > 0, \Delta B(t) > 0\}$, where $\Delta B(t) = B(t) - \lim_{s \uparrow t} B(s)$, denote the countable set of discontinuities of B and when convenient we identify this set with a sequence $\{t_k, k \geq 1\}$ such that $0 < t_1 < t_2 < \ldots < t_k < \ldots$. Consider the following decomposition of N.

Let χ_A denote the characteristic function of the set A and introduce the following decomposition of N:

$$N = N^c + N^d \tag{2.25}$$

where for each $t \geq 0$ and $i \in E$

$$\begin{aligned} N_t^d(i) &= \sum_{k: t_k \leq t} \Delta N_{t_k}(i) \\ &= \int_0^t \chi_A(s) dN_s(i). \end{aligned} \tag{2.26}$$

Thus $N_t^d(i)$ counts the frequency of events of type i to occur exactly at one of the $t_k \in A$ over the interval $(0,t]$. Obviously, N^c and N^d are counting processes.

Let $B^c = B - \Sigma\Delta B$ denote the continuous part of B. This allows a decomposition of B of the form:

$$B = B^c + B^d \tag{2.27}$$

where $B^d = \Sigma\Delta B = \int \chi_A dB$. Thus the compensator function A may be written

$$A = \int YdB^c + \int YdB^d \tag{2.28}$$

$$= A^c + A^d$$

with obvious correspondence. It is shown that N^c has compensator A^c and N^d compensator A^d as follows.

Observe that $\chi_A = \{\chi_A(t),\ t \geq 0\}$ is a deterministic and therefore predictable process and χ_A is bounded by one. Consider the Lebesgue-Stieltjes integral

$$m_t^d = \int_{(0,t]} \chi_A(s) m_s \tag{2.29}$$

$$= \sum_{k:t_k \leq t} (\Delta N_{t_k} - Y_{t_k} \Delta B(t_k))$$

$$= N_t^d - \int_0^t Y_s B^d\{ds\}, \quad t \geq 0.$$

By virtue of lemma 2.2 it follows that $N^d = \{(N_t^d, \mathcal{F}_t),\ t \geq 0\}$ is a Poisson-type counting process having unique compensator A^d. Similarly, N^c is a Poisson-type counting process having unique compensator A^c.

In closing this section we note that in general the function B may be decomposed as follows:

$$B = B^{ac} + B^{sc} + B^d \tag{2.30}$$

where B^d is defined above, B^{ac} is an absolutely continuous function and B^{sc} is a singular continuous function (absolute continuity and singularity are of course defined relative to Lebesgue measure, cf Royden (1968)). Let μ denote the Lebesgue measure and set $\beta = dB^{ac}/d\mu$ so that

$$B^{ac} = \int \beta d\mu. \tag{2.31}$$

The compensator A is therefore written

$$\begin{aligned} A &= \int Y\beta d\mu + \int YdB^{sc} + \int YdB^{d} \\ &= A^{ac} + A^{sc} + A^{d}. \end{aligned} \tag{2.32}$$

The interesting thing about (2.32) is that a counting process, N^{ac} say, generated by A^{ac} would possess an intensity process $Y\beta$ which is a multiplicative intensity as defined by Aalen (1978).

2.2.6

The decomposition of N given at (2.25) shows N to be the superposition of two Poisson-type counting processes N^c with continuous compensator A^c and N^d with purely discrete compensator A^d, the latter having event epochs exclusively in the set A. Thus the sequence $T = \{\tau_n, n \geq 1\}$ of event epochs of N is a mixture of the event epochs of N^c and N^d, which we denote by $T^c = \{\tau_n^c, n \geq 1\}$ and $T^d = \{\tau_n^d, n \geq 1\}$, respectively.

By virtue of the continuity of the compensator A^c of N^c the sequence T^c is a sequence of totally inaccessible stopping times (see for example, lemma 18.3, corollary 1, Lipster and Shiryayev (1978)). Also it is a straightforward demonstration to show that the sequence

$T^d = \{\tau_n^d, n \geq 1\}$ is an accessible (see definition 7.4, Metivier (1982)) sequence of stopping times. We remark that the sequence T^d is predictable if and only if the counting process N^d is predictable in which case we must have $\Delta N_{t_k}^d = Y_{t_k} \Delta B(t_k)$ P-a.s. for all $t_k \in \mathcal{A}$. In such a case estimating an unknown B^d from observable (N^d,Y) would be trivially accomplished. However, in many statistical applications such as we illustrate in connection with survival analysis T^d is accessible but not predictable.

In light of the preceding discussion we use the terminology accessible set when referring to the set $\mathcal{A}$. In some statistical applications where B is unknown, $\mathcal{A}$ may be unknown as well and require estimation. This topic is beyond the scope of the present endeavor.

2.2.7

Much about the behavior of N can be learned from characterization of the compensator A and vice versa. The model (2.21) which defines the family of Poisson-type counting processes exhibits an interplay between the process Y and the measure B in determining the rate of occurrence of events of different types in the point process. Observe that in the case $Y \equiv 1$ (i.e. a constant process) then (2.21) yields $A = B$ meaning N has deterministic compensator B. We noted earlier that in this case N is a process of independent increments. Thus an immediate role played by the general process Y is to introduce a form of stochastic dependence between the future behavior of N and elements of its past.

Next we prove a version of lemma 3.3, Aalen (1978) which relates the occurrence of events in N with the compensator A and relies on the decomposition of B shown at (2.30).

<u>Lemma 2.3</u>. Suppose the family of processes $\{Y(i),\ i \in E\}$ is bounded by an integrable random variable. Then for each $n = 1,2,\ldots$ and $i \in E$ we have the following:

(i) $\lim_{h\downarrow 0} h^{-1}E(N_{t+h}(i) - N_t(i)|F_t) = Y_{t+}(i)\beta_i(t+)$;

(ii) $\lim_{h\downarrow 0} h^{-1}P(N_{t+h}(i) - N_t(i) = 1|F_t) = Y_{t+}(i)\beta_i(t+)$;

(iii) $\lim_{h\downarrow 0} h^{-1}P(N_{t+h}(i) - N_t(i) > 1|F_t) = 0$;

(iv) For each k such that $t_k \in A$

$$\lim_{h\downarrow 0} P(N_{t_k}(i) - N_{t_k-h}(i) = 1|F_{t_k-h})$$

$$= \lim_{h\downarrow 0} E(N_{t_k}(i) - N_{t_k-h}(i)|F_{t_k-h}) = Y_{t_k}(i)\Delta B_i(t_k);$$

where (i)-(iii) hold for t outside a set of μ-measure zero and on $\{(\omega,t):\ t \le \tau_n\}$, $\beta_i = dB_i^{ac}/d\mu$ and $\Delta B_i(t_k) = B_i(t_k) - B_i(t_k-)$.

<u>Proof</u>. The proof of (i)-(iii) can be obtained by the method of proof used in lemma 3.3(i)-(iii), Aalen (1978). First recall that for each $n = 1,2,\ldots$ $\{N_{t\wedge\tau_n} - A_{t\wedge\tau_n},\ t \ge 0\}$ is a martingale. Secondly, the functions B^{sc} and B^d are both singular and have derivative equal to zero a.e. relative to the Lebesgue measure. Thus for t outside a set of Lebesgue measure zero and on $\{(\omega,t),\ t \le \tau_n\}$ (so that $t \wedge \tau_n = t$) (i)-(iii) may be proved by the method of proof used in Lemma 3.3, Aalen (1978).

To prove (iv) recall the decomposition of A and N given by

$$A = A^c + A^d \tag{2.33}$$

$$N = N^c + N^d \tag{2.34}$$

where N^c has compensator A^c and N^d has compensator A^d. By virtue of the continuity of A^c it is easily seen that for $n = 1,2,...$ and $k \geq 1$

$$\begin{aligned} &\lim_{h \downarrow 0} E(N_{t_k \wedge \tau_n} - N_{(t_k \wedge \tau_n)-h} | \mathcal{F}_{t_k - h}) \\ &= \lim_{h \downarrow 0} [O(h) + E(N^d_{t_k \wedge \tau_n} - N^d_{(t_k \wedge \tau_n)-h} | \mathcal{F}_{t_k - h})] \\ &= \lim_{h \downarrow 0} E(N^d_{t_k \wedge \tau_n} - N^d_{(t_k \wedge \tau_n)-h} | \mathcal{F}_{t_{k-h}}) \end{aligned} \tag{2.35}$$

where $O(h) \downarrow 0$ as $h \downarrow 0$. Finally, since N^d has compensator $A^d = \int YdB^d$ and Y is predictable (hence $Y_{t_k} \in \mathcal{F}_{t_k -}$) we obtain (iv) whenever $(\omega, t_k) \in \{(\omega,t): t \leq \tau_n\}$. Note that in (2.35) we have conveniently suppressed the index i.

To obtain the result for $\lim_{h \downarrow 0} P(N_{t_k} - N_{t_k - h} = 1 | \mathcal{F}_{t_k - h})$ we introduce a family of random variables $\{S(h), h \geq 0\}$ where $S(h) > (t_k \wedge \tau_n) - h$ is the time of the first jump in N after $(t_k \wedge \tau_n) - h$. From what has been proved thus far it follows that as $h \downarrow 0$ $S(h) \uparrow S(0)$ where $S(0) \geq t_k \wedge \tau_n$ P-a.s. Hence, the result (iv) is proved as above via the method of proof of lemma 3.3(ii), Aalen (1978). This completes our proof. □

2.2.8

We began our discussion with a counting process $N = \{(N_t, F_t),\ t \geq 0\}$ having predictable compensator $A = \{(A_t, F_t),\ t \geq 0\}$ where A is a random measure of the form

$$A_t = \int_{(0,t]} Y_s B\{ds\}, \quad t \geq 0 \tag{2.36}$$

as shown in definition 2.1. In example 3, 2.1 we introduced the filtration $F^N = \{F_t^N,\ t \geq 0\}$ called the internal history of N where

$$F_t^N = F_0 \vee \sigma\text{-}(N_s,\ s \leq t), \quad t \geq 0. \tag{2.37}$$

More generally, we introduce the filtration $G = \{G_t,\ t \geq 0\}$ such that G satisfies

$$F_t^N \subseteq G_t \subseteq F_t, \quad \forall\ t \geq 0. \tag{2.38}$$

If we consider the counting process $\bar{N} = \{(N_t, G_t),\ t \geq 0\}$ it is natural to ask what is the form of the compensator $\bar{A} = \{(A_t, G_t),\ t \geq 0\}$ of $\bar{N}$ relative to G. In the general case this is a very difficult problem but as we shall see for the Poisson-type counting process this is not the case.

Lemma 2.4. Let the compensator of a point process $N = \{(N_t, F_t),\ t \geq 0\}$ be of the Poisson-type and given by the formula (2.36). Then

(2.39) $$\bar{A}_t = \int_{(0,t]} \bar{Y}_s B\{ds\}, \ t \geq 0$$

is the unique predictable compensator of a point process $\bar{N} = \{(N_t, G_t), \ t \geq 0\}$ where $\bar{Y}_t$ is a $P\{d\omega\}B\{dt\}$ measurable version of $E(Y_t | G_t)$. □

The proof of this lemma is given in Theorem 18.3, Lipster and Shiryayev (1978) for the case $G = F^N$ and for general G in the case where the intensity exist by Bremaud (1980) (see T14, Chapter II). However, it is easily seen that lemma 2.4 can be proved by the method used in Lipster and Shiryayev (1978) for general G without change.

2.2.9

The necessary and sufficient conditions for the likelihood of Poisson-type counting processes to have a convenient exponential form can be given for the self-exciting case (i.e. $F = F^N$). The form of the likelihood is shown by the process $\{Z_t, \ t \geq 0\}$ defined at equation (14), Jacod (1975) and the conditions are those which guarantee the validity of $E(Z_\infty) = 1$. Since this result is vital to the development of likelihood based inference procedures and the proof long and complicated we are inclined to consider this problem in a later paper.

3. ESTIMATION OF THE COMPENSATOR

3.1.1

We address the problem of estimation of the compensator from Poisson-type counting processes as defined in section 2. We assume that the process (N,Y) is observable and that the Borel measure B in equation (2.1) is unknown. The statistical problem is to estimate B based on observation of the process (N,Y) over a period of time.

For each $n = 1,2,\ldots$ suppose we observe (N^i,Y^i), $i = 1,\ldots,n$ independent processes where $N^i(j) = \{(N^i_t(j), \mathcal{F}^i_t), t \geq 0\}$ is a Poisson-type counting process with compensator $A^i(j) = \int Y^i(j)dB_j$, $j = 1,\ldots,m$ and $i = 1,\ldots,n$. For each n define $F^{(n)} = \{\mathcal{F}^{(n)}_t, t \geq 0\}$, $N^{(n)}$ and $Y^{(n)}$ as follows:

$$\mathcal{F}^{(n)}_t = \bigvee_{i=1}^{n} \mathcal{F}^i_t, \ \forall\ t \geq 0; \tag{3.1}$$

$$N^{(n)} = \sum_{i=1}^{n} N^i; \tag{3.2}$$

$$Y^{(n)} = \sum_{i=1}^{n} Y^i. \tag{3.3}$$

Denote $A^{(n)} = \int Y^{(n)}dB$ (componentwise) and observe that for each n $N^{(n)} - A^{(n)} \in \mathcal{M}_{loc}(F^{(n)},P^{(n)})$, where $P^{(n)} = P^1 \times P^2 \times \ldots \times P^n$. Note that $\Delta N^{(n)} > 1$ occurs with positive probability at the atoms of B (i.e. the accessible set $\mathcal{A}$) so that technically $N^{(n)}$ is not a counting process in our sense of the term but it is an integral-valued random measure over $(R^+,\sigma(R^+))$.

We remark that the sequence $(N^{(n)},Y^{(n)},A^{(n)})$ outlined above may arise in other ways and the general theory on weak convergence we appeal to in the next section can be adapted to this wider usage.

To continue, for each n we define the $F^{(n)}$-predictable process X^n by

$$X_t^n = (Y_t^{(n)})^{-1} 1(Y_t^{(n)} > 0), \ \forall\ t \geq 0 \tag{3.4}$$

where we use the convention $\frac{0}{0} := 0$. Next for each $j = 1,\dots,m$ consider the following Stieltjes integral

$$\hat{B}_t^n(j) = \int_0^t X_s^n(j) dN_s^{(n)}(j), \quad \forall\ t \geq 0 \tag{3.5}$$

and denote $\hat{B}^n = (\hat{B}^n(1),\dots,\hat{B}^n(m))$. We propose to use $\hat{B}^n$ as an estimator of B which for Poisson-type counting processes is the analogous estimator to that used by Aalen (1978) in connection with the multiplicative intensity model.

We observe that the indicator function $1(Y^{(n)} > 0)$ appearing in equation (3.4) becomes particularly relevant when with positive probability $B\{t\colon Y_t^{(n)} = 0\} > 0$. In this case we cannot estimate B on $\{t\colon Y_t^{(n)} = 0\}$ but rather we estimate $\tilde{B}^n$ where

$$\tilde{B}^n = \int 1(Y^{(n)} > 0) dB. \tag{3.6}$$

Thus in problems of this type $\tilde{B}^n$ plays the role of "parameter" and does not become known to us until the process $Y^{(n)}$ has been observed. This is a common occurrence in problems of estimation from stochastic processes and numerous other examples may be found (c.f. Aalen (1978)).

In lemma 3.1 (below) we show that for each $n = 1,2,\dots$ the error of estimation process $\hat{B}^n - \tilde{B}^n$ is a local martingale for $\hat{B}^n$ of (3.5) and $\tilde{B}^n$ of (3.6). First we recall that $N^{(n)} - A^{(n)}$ is a local martingale and therefore let $\{\tau_k^n, k \geq 1\}$ denote a localizing sequence of $F^{(n)}$-stopping times for this process (i.e. for each k $\{N^{(n)}_{t\wedge\tau_k^n} - A^{(n)}_{t\wedge\tau_k^n}, t \geq 0\}$ is a martingale). Also let $\tau_\infty^n = \lim_{k\uparrow\infty} \tau_k^n$ and consider the following lemma.

Lemma 3.1. Let $\hat{B}^n$ and $\tilde{B}^n$ be defined by (3.5) and (3.6), respectively.

i) If for each $n = 1,2,\ldots$

$\int_0^t 1(Y_s^{(n)}(j) > 0)B_j\{ds\} < \infty, \ \forall\ t < \tau_\infty^n, \quad j = 1,\ldots,m$ (a.s.-$P^{(n)}$),

then $\hat{B}^n - \tilde{B}^n \in M_{loc}(F^{(n)}, P^{(n)})$.

ii) If for each $n = 1,2,\ldots$

$\int_0^t X_s^n(j)B_j\{ds\} < \infty, \ \forall\ t < \tau_\infty^n, \ j = 1,\ldots,m$ (a.s.-$P^{(n)}$)

then $\hat{B}^n - \tilde{B}^n \in M_{loc}^2(F^{(n)}, P^{(n)})$ and for each $t \geq 0$

(3.7)

$$\langle B^n - \tilde{B}^n\rangle_t^{ij} = \begin{cases} \int_0^t X_s^n(i)B_i\{ds\} - \sum_{s\leq t} (X_s^n(i))^2\{\sum_{k=1}^n (Y_s^k(i))^2\}(\Delta B_i(s))^2, & i = j; \\ \text{and} & \\ -\sum_{k=1}^n \int_0^t X_s^n(i)X_s^n(j)Y_s^k(i)Y_s^k(j)\Delta B_i(s)B_j\{ds\}, & i \neq j \end{cases}$$

where for all $i,j \in E$, $\langle \hat{B}^n - \tilde{B}^n\rangle^{ij} = \langle \hat{B}^n(i) - \tilde{B}^n(i), \hat{B}^n(j) - \tilde{B}^n(j)\rangle$.

Proof. Observe that by virtue of the predictability of $Y^{(n)}$ X^n is a predictable process and that from equation (3.4) $P(|X_t^n| < \infty) = 1, \ \forall\ t \geq 0$. Secondly, observe that

$$\hat{B}^n - \tilde{B}^n = \int X^n(dN^{(n)} - dA^{(n)}) \tag{3.8}$$

$$= \sum_{k=1}^n \int X^n(dN^k - dA^k)$$

and that for each $j = 1,2,\ldots,m$

$$(3.9)\qquad \int_0^t 1(Y_s^{(n)}(j) > 0)B_j\{ds\} = \int_0^t X_s^n(j) \sum_{k=1}^n Y_s^k(j)B_j\{ds\}$$

$$= \sum_{k=1}^n \int_0^t X_s^n(j)Y_s^k(j)B_j\{ds\},\ \forall\ t \geq 0.$$

Finally recall that for each $k = 1,\ldots,m$ $N^k - A^k \in M_{loc}(F^{(n)},P^{(n)})$. Thus by (i), lemma 2.2(b) and equation 3.8 the conclusion $\hat{B}^n - \tilde{B}^n \in M_{loc}(F^{(n)},P^{(n)})$ easily follows.

To prove part (ii) we first observe that with probability one

$$(3.10)\qquad \int_0^t X_s^n(j)B_j\{ds\} = \int_0^t (X_s^n(j))^2 dA_s^{(n)}(j)$$

$$= \sum_{k=1}^n \int_0^t X_s^n(j)dA_s^k(j)$$

$$\geq \sum_{k=1}^n \int_0^t X_s^n(j)(1-Y_s^k(j)\Delta B_j(s))Y_s^k(j)B_j\{ds\}$$

for all $t \geq 0$, $j = 1,\ldots,m$ where the inequality results from the fact that $0 \leq \Delta A_s^k(j) \leq 1$, $\forall\ s \geq 0$ with probability one. By (ii) the right hand side of (3.10) is almost surely finite for all $t < \tau_\infty^n$ and thus by theorem 18.8, Lipster and Shiryayev (1978) we obtain $\int X^n(dN^i - dA^i) \in M_{loc}^2(F^{(n)},P^{(n)})$, $i = 1,\ldots,n$ and therefore their sum $\hat{B}^n - \tilde{B}^n \in M_{loc}^2(F^{(n)},P^{(n)})$. Finally, by the assumption of independence, the family $\{N^i - A^i,\ i = 1,\ldots,m\}$ is mutually orthogonal so that by lemma 2.1, theorem 18.8 (Lipster and Shiryayev (1978)) we obtain (3.7). □

3.1.2

Consistency. We consider the problem of determining in what sense $\hat{B}^n$ is to be considered a close approximation to B on $\{t\colon Y_t^{(n)} > 0\}$. Thus the concept of consistency we develop here is defined in terms associated with the error of estimation process $\hat{B}^n - \tilde{B}^n$.

For use in theorem 3.1 (below) we define two modes of uniform consistency of the estimator $\hat{B}$. In each mode we assume that for each $n = 1,2,\ldots$ $\hat{B}^n - \tilde{B}^n \in \mathcal{M}_{loc}(F^{(n)}, P^{(n)})$ and let $\{\tau_k^n, k \geq 1\}$ be a localizing sequence of $F^{(n)}$-stopping times with $\tau_\infty^n = \lim_{k\uparrow\infty} \tau_k^n$.

<u>Definition 3.1</u>. For each k such that $1 \leq k \leq \infty$ and

$$(3.11) \qquad \sup_{s \leq t} |\hat{B}^n_{s \wedge \tau_k^n} - \tilde{B}^n_{s \wedge \tau_k^n}| \xrightarrow{P} 0 \quad \text{as} \quad n \uparrow \infty, \ \forall\, t \geq 0$$

we call $\{\hat{B}^n, n \geq 1\}$ a <u>local τ_k-uniformly consistent</u> estimator for $\{\tilde{B}^n, n \geq 1\}$. □

<u>Definition 3.2</u>. For each k such that $1 \leq k \leq \infty$ and

$$(3.12) \qquad \sup_{s \leq \tau_k^n} |\hat{B}^n_s - \tilde{B}^n_s| \xrightarrow{P} 0 \quad \text{as} \quad n \uparrow \infty$$

we call $\{\hat{B}^n, n \geq 1\}$ a <u>τ_k-uniformly consistent</u> estimator for $\{\tilde{B}^n, n \geq 1\}$. □

The relationship between these modes of consistency is obviously that τ_k-uniform consistency implies local τ_k-uniform consistency. Further, if in either mode consistency is obtained for some k it is also obtained for $k-1, k-2, \ldots, 1$.

Observe that in each mode of consistency we can only assert that $\hat{B}^n$ is a close approximation to B on $\{t: Y_t^{(n)} > 0\}$ and then confined only to subintervals of R^+. Implicit in these definitions is the possibility of replacing $\tilde{B}^n$ with B when conditions guarantee that, for example, $Y_{\cdot}^{(n)} > 0$ uniformly on the subintervals in question in (3.11) and (3.12). In most applications such conditions will generally be of importance. For example in random censorship models in survival analysis where $Y^{(n)}$ may be interpreted as a risk process and B as the cumulative hazard rate we may have restrictions on the censoring mechanism which produce the required conditions.

The idea behind these modes of consistency, particularly local uniform consistency, is that according to definition 3.1 and 3.2 we may only achieve consistency of the estimator $\hat{B}$ over bounded intervals of the form $[0,t]$ and not necessarily over $[0,\infty)$. We apply these definitions in the next theorem where all operations (e.g. $|\cdot|$) on vector valued processes are to be interpreted componentwise.

Theorem 3.1. Suppose $\{\hat{B}^n, n \geq 1\}$ is a sequence of estimators defined at (3.5) for $\{\tilde{B}^n, n \geq 1\}$ defined at (3.6). Consider the following conditions.

[α] For $k < \infty$ and $t \in R^+$

(1) $E|\hat{B}^n_{t\wedge\tau_k^n} - \tilde{B}^n_{t\wedge\tau_k^n}| \to 0$ as $n \uparrow \infty$;

(2) $E|\hat{B}^n_{\tau_k^n} - \tilde{B}^n_{\tau_k^n}| \to 0$ as $n \uparrow \infty$;

and for $k = \infty$ we require (1) or (2) with the additional condition

(3) $E\int_0^{\tau_\infty^n} 1(Y_s^{(n)} > 0)B\{ds\} = E\int_0^{\tau_\infty^{n}-} 1(Y_s^{(n)} > 0)B\{ds\} < \infty.$

[β] For all k and $t \in R^+$

(1) $\langle\hat{B}^n-\tilde{B}^n\rangle^{ii}_{t\wedge\tau_k^n} \overset{P}{\to} 0$ as $n \uparrow \infty$, $i = 1,\ldots,m$;

(2) $\langle\hat{B}^n-\tilde{B}^n\rangle^{ii}_{\tau_k^n} \overset{P}{\to} 0$ as $n \uparrow \infty$, $i = 1,\ldots,m$.

a) If for each $n = 1,2,\ldots$ $\hat{B}^n - \tilde{B}^n \in \mathcal{M}_{loc}(F^{(n)}, P^{(n)})$, then $[\alpha](1)$ and $[\alpha](2)$ imply $\{\hat{B}^n\}$ is local τ_k-uniformly consistent and τ_k-uniformly consistent, respectively.

b) If for each $n = 1,2,\ldots$ $\hat{B}^n - \tilde{B}^n \in \mathcal{M}^2_{loc}(F^{(n)}, P^{(n)})$, then $[\beta](1)$ and $[\beta](2)$ imply $\{\hat{B}^n\}$ is local τ_k-uniformly consistent and τ_k-uniformly consistent, respectively.

Proof. First consider the case $k < \infty$. Assume that $\hat{B}^n - \tilde{B}^n \in \mathcal{M}_{loc}(F^{(n)}, P^{(n)})$ and that (1) of $[\alpha]$ holds. Since $|\hat{B}^n_{\cdot\wedge\tau^n_k} - \tilde{B}^n_{\cdot\wedge\tau^n_k}|$ is a submartingale theorem 3.4, Doob (1953) implies that for each real λ and $t \geq 0$

$$(3.13) \qquad \lambda P(\sup_{s\leq t} |\hat{B}^n_{s\wedge\tau^n_k} - \tilde{B}^n_{s\wedge\tau^n_k}| \geq \lambda) \leq E|\hat{B}^n_{t\wedge\tau^n_k} - \tilde{B}^n_{t\wedge\tau^n_k}|.$$

Since λ is arbitrary (1) of $[\alpha]$ implies that $\hat{B}$ is local τ_k-uniformly consistent.

Next suppose that (2) of $[\alpha]$ holds. Let $\{t_\ell, \ell \geq 1\}$ be a deterministic sequence such that $t_\ell \uparrow \infty$ as $\ell \uparrow \infty$ and $t_\ell \wedge \tau^n_k \uparrow \tau^n_k$ almost surely. The monotone convergence theorem and Fatou's lemma, Royden (1968) together with (3.13) imply that

$$(3.14) \qquad \lambda P(\sup_{s\leq\tau^n_k} |\hat{B}^n_s - \tilde{B}^n_s| \geq \lambda) \leq \liminf_{\ell} E|\hat{B}^n_{t_\ell\wedge\tau^n_k} - \tilde{B}^n_{t_\ell\wedge\tau^n_k}|.$$

To verify that the right hand side of (3.14) tends to zero as $n \uparrow \infty$, recall that $\{|\hat{B}^n_{t_\ell\wedge\tau^n_k} - \tilde{B}^n_{t_\ell\wedge\tau^n_k}|, \ell \geq 1\}$ is a uniformly integrable family of random variables such that $|\hat{B}^n_{t_\ell \wedge \tau^n_k} - \tilde{B}^n_{t_\ell \wedge \tau^n_k}| \to |\hat{B}_{\tau^n_k} - \tilde{B}^n_{\tau^n_k}|$ a.s. $-P^{(n)}$.

Thus proposition 5.19, Breiman (1968) implies that

$$(3.15) \qquad \lim_{\ell \uparrow \infty} E|\hat{B}^n_{t_\ell \wedge \tau^n_k} - \tilde{B}^n_{t_\ell \wedge \tau^n_k}| = E|\hat{B}^n_{\tau^n_k} - \tilde{B}^n_{\tau^n_k}|.$$

On combining (3.15), (3.14), (2) of $[\alpha]$ and the arbitrariness of λ it follows that $\hat{B}$ is τ_k-uniformly consistent.

Next assume, as in (b), that $\hat{B}^n_{\cdot \wedge \tau^n_k} - \tilde{B}^n_{\cdot \wedge \tau^n_k} \in M^2(F^{(n)}, P^{(n)})$. Therefore $(\hat{B}^n_{\cdot \wedge \tau^n_k} - \tilde{B}^n_{\cdot \wedge \tau^n_k})^2 - \langle \hat{B}^n - \tilde{B}^n \rangle_{\cdot \wedge \tau^n_k} \in M(F^{(n)}, P^{(n)})$ and for any finite $F^{(n)}$-stopping τ

$$(3.16) \qquad E\{(\hat{B}^n_{\tau \wedge \tau^n_k} - \tilde{B}^n_{\tau \wedge \tau^n_k})^2\} = E\{\langle \hat{B}^n - B^n \rangle_{\tau \wedge \tau^n_k}\}.$$

Further for $i = 1, \ldots, m$ $(\hat{B}^n(i) - \tilde{B}^n(i))^2$ and $\langle \hat{B}^n - \tilde{B}^n \rangle^{ii}$ are nonnegative with $\langle \hat{B}^n - \tilde{B}^n \rangle^{ii}$ being nondecreasing and $F^{(n)}$-predictable. Hence according to the Lenglart inequality, Lenglart (1977)

$$(3.17) \qquad P\left(\sup_{s \le t \wedge \tau^n_k} |\hat{B}^n_s(i) - \tilde{B}^n_s(i)| > a^{1/2} \right)$$
$$\le \frac{1}{a} E(\langle \hat{B}^n - \tilde{B}^n \rangle^{ii}_{t \wedge \tau^n_k} \wedge b) + P(\langle \hat{B}^n - \tilde{B}^n \rangle^{ii}_{t \wedge \tau^n_k} \ge b)$$

for all real $a > 0$ and $b > 0$ and $t \in R^+$.

Assume that (1) of $[\beta]$ holds so that b in (3.17) can be taken arbitrarily small and therefore the right hand side of (3.17) can be made arbitrarily small for any a. Hence, (1) of $[\beta]$ and (3.17) implies $\hat{B}$ is local τ_k-uniformly consistent.

Next assume that (2) of $[\beta]$ holds and replace t in (3.17) with a sequence $\{t_\ell, \ell \geq 1\}$ such that $t_\ell \uparrow \infty$ as $\ell \uparrow \infty$ so that $t_\ell \wedge \tau_k^n \uparrow \tau_k^n$ a.s. $-P^{(n)}$. Recall that $\langle\hat{B}^n-\tilde{B}^n\rangle^{ii}$ is a nondecreasing process and that in particular $\langle\hat{B}^n-\tilde{B}^n\rangle^{ii}_{t_\ell\wedge\tau_k^n} \uparrow \langle\hat{B}^n-\tilde{B}^n\rangle^{ii}_{\tau_k^n}$ a.s. $-P^{(n)}$. Thus an application of the monotone convergence theorem to (3.17) (using $\{t_\ell\}$) it follows that (2) of $[\beta]$ implies that $\hat{B}$ is τ_k-uniformly consistent.

It remains to consider the case $k = \infty$. First observe that $\langle\hat{B}^n-\tilde{B}^n\rangle^{ii}$ is a nondecreasing process and consider a deterministic sequence $\{t_\ell, \ell \geq 1\}$ such that $t_\ell \wedge \tau_\ell^n \to \tau_\infty^n$ a.s. $-P^{(n)}$. First apply (3.17) with $\{\tau_\ell^n, \ell \geq 1\}$ replacing τ_k^n and for fixed $t \geq 0$. Using the fact that $\langle\hat{B}^n-\tilde{B}^n\rangle^{ii}_{\tau_\infty^n-} = \langle\hat{B}^n-\tilde{B}^n\rangle^{ii}_{\tau_\infty^n}$ a.s. $-P^{(n)}$ and letting $\ell \to \infty$ then (1) of $[\beta]$, Fatou's lemma and the monotone convergence theorem imply that $\hat{B}$ is local τ_∞-uniformly consistent. Secondly, apply (3.17) with $\{\tau_\ell^n, \ell \geq 1\}$ replacing τ_k^n and $\{t_\ell, \ell \geq 1\}$ replacing t. Again using $\langle\hat{B}^n-\tilde{B}^n\rangle^{ii}_{\tau_\infty^n-} = \langle\hat{B}_n-\tilde{B}^n\rangle^{ii}_{\tau_\infty^n}$ a.s. $-P^{(n)}$ and letting $\ell \to \infty$ we have that (2) of $[\beta]$ implies $\hat{B}$ is τ_∞-uniformly consistent.

Observe that the added condition (3) of $[\alpha]$ (for $k = \infty$) implies, according to lemma 2.2, that $\hat{B}_{\cdot\wedge\tau_\infty^n} - \tilde{B}^n_{\cdot\wedge\tau_\infty^n}$ is a uniformly integrable martingale so that (1) and (3) of $[\alpha]$ implies, via equation (3.13), that $\hat{B}$

is local τ_∞-uniformly consistent. Finally, take a sequence $\{t_\ell, \ell \geq 1\}$ such that $t_\ell \uparrow \infty$ and note that condition (3) implies that $\{\hat{B}^n_{t_\ell \wedge \tau^n_\infty} - \tilde{B}^n_{t_\ell \wedge \tau^n_\infty}, \ \ell \geq 1\}$ is a uniformly integrable family. Thus equation (3.13) and proposition 5.19, Breiman (1968) (using $|\hat{B}^n_{\tau^n_\infty -} - \tilde{B}^n_{\tau^n_\infty -}| = |\hat{B}^n_{\tau^n_\infty} - \tilde{B}^n_{\tau^n_\infty}|$ - a.s.) together with conditions (1) and (3) of $[\alpha]$ imply that $\hat{B}$ is τ_∞-uniformly consistent. This completes the proof. □

3.1.3.

<u>Weak Convergence to a Gaussian process</u>. We investigate conditions so that the normalized error of estimation process $n^{1/2}(\hat{B}^n - \tilde{B}^n)$ converges weakly as $n \to \infty$ to a Gaussian process of independent increments. To prove this result we appeal to the functional central limit theorems for semimartingales developed by Jacob, Klopotowski and Mémin (1982) and Lipster and Shiryayev (1980).

Theorem 3.2 (below) considers the case where $\hat{B}^n - \tilde{B}^n$ is of class $M^2_{loc}(F^{(n)}, P^{(n)})$ for each $n = 1,2,\ldots$. In this setting it is natural to consider (see section 3.1.4) estimation of the limiting quadratic characteristic function (i.e. asymptotic covariance function) associated with the weak limit of $n^{1/2}(\hat{B}^n - \tilde{B}^n)$. It turns out that the convergence criteria we give for this case appear to be those we can often verify in applications (see for example section 4).

In theorem 3.2 (below) the limit process X is a zero mean Gaussian process of independent increments and is defined as follows. Let $\xi = \{\xi_t, t \geq 0\}$ be a continuous m-dimensional zero mean Gaussian process and let $C = (C^1, \ldots, C^m)$ be a continuous function defined on $R^+ = [0,\infty)$. We assume that for each i,j, $1 \leq i,j \leq m$

$$\langle\xi\rangle_t^{ij} = \begin{cases} C^i(t), & i = j; \\ 0, & i \neq j. \end{cases} \tag{3.18}$$

Secondly, let $U = \{U_j,\ j \geq 1\}$ be a family of independent zero mean Gaussian random variables in R^m such that U_j has distribution measure Φ_j and covariance matrix $\{\sigma_j^{\ell k},\ 1 \leq \ell,k \leq m\}$. Let $\{t_j,\ j \geq 1\}$ be a countable sequence such $0 < t_1 < t_2 < \ldots < t_j$ and suppose $\{t_j\}$ is identified with the accessible set A defined in section 2.2.5. We assume that for each $t \geq 0$, $\sum_{j:t_j \leq t} U_j$ converges almost surely and that $\{U_j\}$ is independent of ξ.

Finally, we define the process $X = \{X_t,\ t \geq 0\}$ by

$$X_t = \xi_t + \sum_{j:t_j \leq t} U_j, \quad t \geq 0 \tag{3.19}$$

so that for each i,j, $1 \leq i,j \leq m$,

$$\langle X\rangle_t^{ij} = \begin{cases} C^i(t) + \sum_{k:t_k \leq t} \sigma_k^{ii}, & i = j; \\ \sum_{k:t_k \leq t} \sigma_k^{ij}, & i \neq j; \end{cases} \tag{3.20}$$

and for all $t \geq 0$. Thus X is a zero mean Gaussian process of independent increments having quadratic characteristic given by (3.20).

<u>Theorem 3.2</u>. Let A denote the accessible set and identify A with the sequence $\{t_\ell,\ \ell \geq 1\}$. Let X be the Gaussian process defined according to equation (3.19) and let $n^{1/2}(\hat{B}^n - \tilde{B}^n)$ be the normalized error of estimation process defined by (3.5) and (3.6). Define a process $Z_\cdot^n = n^{1/2}(Y_\cdot^{(n)})^{-1} 1(Y_\cdot^{(n)} > 0)$ where $Y^{(n)}$ is defined at (3.3) and consider the following conditions:

$[\alpha]$ $\forall\ t \geq 0,\ j = 1,2,\ldots,m,\quad \forall\ \epsilon \in (0,1]$

$$\int_0^t (Z_s^n(j))^2 1(|Z_s^n(j)| > \epsilon)\, dA_s^{nc}(j) \underset{n\uparrow\infty}{\overset{P}{\to}} 0;$$

$[\beta]$ $\forall\ t \geq 0,\ j = 1,2,\ldots,m$

$$\int_0^t (Z_s^n(j))^2 dA_s^{nc}(j) \underset{n\uparrow\infty}{\overset{P}{\to}} C^j(t);$$

$[\gamma]$ For each $t_\ell \in A$ and $u \in R^m$

$$E(e^{iu\cdot W_\ell^n} | F_{t_\ell -}^{(n)}) \underset{n\uparrow\infty}{\overset{P}{\to}} E(e^{iu\cdot U_\ell}) \quad (i = \sqrt{-1}),$$

where $W_\ell = Z_{t_\ell}^n (\Delta N_{t_\ell}^{(n)} - \Delta A_{t_\ell}^{(n)})$;

$[\delta]$ $\forall\ t \geq 0$ and $i,\ 1 \leq i \leq m$

1) $$\sum_{\ell: t_\ell \leq t} (\Delta < n^{1/2}(\hat{B}^n - \tilde{B}^n)>_{t_\ell}^{ii})^2 \underset{n\uparrow\infty}{\overset{P}{\to}} \sum_{\ell: t_\ell \leq t} (\sigma_\ell^{ii})^2;$$

where $A^{nc} = A^n - \Sigma\Delta A^n$ and $\Delta<n^{1/2}(\hat{B}^n-\tilde{B}^n)>_t = n[<\hat{B}^n-\tilde{B}^n>_t - <\hat{B}^n-\tilde{B}^n>_{t-}]$ where $<\hat{B}^n-\tilde{B}^n>$ is defined at (3.7). Also for all $x,u \in R^m$ $u\cdot x = \Sigma_{i=1}^m u_i x_i$.

Let $D(R^+,R^m)$ denote the space of right continuous functions from R^+ into R^m having left hand limits. If for $n = 1,2,\ldots$ $n^{1/2}(\hat{B}^n-\tilde{B}^n) \in M_{loc}^2(F^{(n)},P^{(n)})$ then the conditions $[\alpha]$, $[\beta]$, $[\gamma]$ and $[\delta]$ imply that $n^{1/2}(\hat{B}^n-\tilde{B}^n)$ converges in law to X in $D(R^+,R^m)$ endowed with the Skorokhod topology as $n \uparrow \infty$.

<u>Proof</u>. To prove the theorem we proceed by showing that the finite dimensional distributions of $n^{1/2}(\hat{B}^n-\tilde{B}^n)$ converge to those of X and to show that the family of probability measures $\{P^{(n)},\ n \geq 1\}$ is relatively compact.

For each $n = 1,2,\ldots$ let $M^n = n^{1/2}(\hat{B}^n-\tilde{B}^n)$ and consider the process M^{nc} defined as follows:

$$(3.21) \qquad M_t^{nc} = \int_0^t (1-\chi_A)(s)dM_s^n, \quad t \geq 0$$

where χ_A denotes the characteristic function of the accessible set A (i.e. $\chi_A(t) = 0$ or 1 as $t \notin A$ or $t \in A$) and where (3.21) is to be interpreted as a Lebesgue-Stieltjes integral. Since χ_A is a bounded predictable process it follows from lemma 2.2 that for each $n = 1,2,\ldots$ M^{nc} defines a locally square integrable local martingale.

It is easily argued that the local martingale M^{nc} is a so-called purely discrete local martingale (see for example Shiryayev (1981) for a definition). Thus there exists a random counting measure μ^{nc} having predictable compensator ν^{nc} such that

$$(3.22) \qquad \mu^{nc}([0,t]\times B) - \nu^{nc}([0,t]\times B), \ t \geq 0$$

defines a local martingale for each Borel set $B \in \sigma(R^m)$. Further for each $i = 1,\ldots,m$ $M^{nc}(i)$ has the representation

$$(3.23) \qquad M_t^{nc}(i) = \int_0^t \int_{R^m} x^i(\mu^{nc}-\nu^{nc})(ds,dx), \ t \geq 0$$

where x^i denotes the i^{th} coordinate of $x \in R^m$.

Now let i denote $\sqrt{-1}$ and for each $u \in R^m$ define the complex predictable process $\hat{A}(M^{nc},u)$ by

$$(3.24) \qquad \hat{A}(M^{nc},u)_t = \int_{R^m} (e^{iu\cdot x} - 1 - iu\cdot x)\nu^{nc}([0,t]\times dx), \quad t \geq 0$$

where $u\cdot x = \Sigma\, u^k x^k$ (i.e. the ordinary inner product).

The process M^{nc} is readily shown to be quasi-left continuous so that in particular $\Delta\hat{A}(M^{nc},u)_t = 0$ for all $t \geq 0$ with probability 1. Therefore, according to theorem 3.4, Jacod, Klopotowski and Mémin (1982), the finite dimensional distributions of M^n converge to those of X follows from $[\gamma]$ upon showing in addition that for all $t \geq 0$

$$(3.25) \qquad \hat{A}(M^{nc},u)_t \overset{P}{\to} -\frac{1}{2}\sum_r (u^r)^2 C^r(t) \quad \text{as} \quad n \uparrow \infty.$$

To prove (3.25) we first observe that $\{M^{nc}(1),\ldots,M^{nc}(m)\}$ are orthogonal local martingales so that for each $u \in R^m$

$$\{(u\cdot M^{nc})_t^2 - \sum_{r=1}^{m} (u^r)^2 \int_0^t (Z_s^n(r))^2 dA_s^{nc}(r),\ t \geq 0\} \in \mathcal{M}_{loc}(F^{(n)},P^{(n)}).$$

The second term (above) is the predictable compensator for $(u\cdot M^{nc})^2$ and may also be written in terms of the measure ν^{nc} as follows:

$$(3.26) \qquad \sum_{r=1}^{m} (u^r)^2 \int_0^t (Z_s^n(r))^2 dA_s^{nc}(r) = \int_{R^m} (u\cdot x)^2 \nu^{nc}([0,t]\times dx), \quad t \geq 0.$$

Therefore, in view of condition $[\beta]$, to prove (3.25) it suffices to show that for $t \geq 0$

$$(3.27) \qquad \hat{A}(M^{nc},u)_t + \frac{1}{2}\int_{R^m} (u\cdot x)^2 \nu^{nc}([0,t]\times dx) \overset{P}{\to} 0 \quad \text{as} \quad n \uparrow \infty.$$

But for each $u \in R^m$ this is precisely the method used by Lipster and Shiryayev (1980) to show that the finite dimensional distributions of a process $u\cdot M^{nc}$ converge to those of $u\cdot\xi$ where $u\cdot\xi$ is a continuous

Gaussian process. Thus it is seen that conditions $[\alpha]$ and $[\beta]$ imply, by the method of proof of theorem 1, part (i), Lipster and Shiryayev (1980) (see conditions (A), (B) and equation (65)), that the limit at (3.27) holds. Hence, by equation (3.26) the limit at (3.25) holds as well.

Since each t_ℓ is a predictable stopping time conditions $[\alpha]$, $[\beta]$ and $[\gamma]$ have been shown to imply conditions i)-iv), theorem 3.4, Jacod, Klopotowski and Mémin (1982). Hence, the finite dimensional distributions of $M^n = n^{1/2}(\hat{B}^n - \tilde{B}^n)$ converge to those of X.

It remains to show that the sequence of laws on D induced by M^n is tight. To achieve this we apply a criterion for relative compactness developed by Jacod and Mémin (1980).

Define the $F^{(n)}$-predictable nondecreasing process

$$(3.28) \qquad G_t^n = \sum_{k=1}^{m} \langle M^n \rangle_t^{kk}, \quad t \geq 0$$

where $\langle M^n \rangle^{kk} = n \langle \hat{B}^n - \tilde{B}^n \rangle^{kk}$ is defined at (3.7). Secondly, define G^∞ by

$$(3.29) \qquad G_t^\infty = \sum_{k=1}^{m} [C^k(t) + \sum_{j:t_j \leq t} \sigma_j^{kk}], \quad t \geq 0,$$

where $\{C^k\}$ and $\{\sigma_j^{kk}\}$ are defined at (3.20).

Let $L(M^n)$ denote the probability law on D induced by the process M^n. First observe that according to Rebolledo (1980) (see bottom page 271) it is no restriction to consider the sequence of processes $\{M^n\}$ as being defined on a common probability space (Ω, F, P) yet adapted to different filtrations $\{F^{(n)}\}$. Secondly, we note that G^{∞} is a deterministic process and in light of conditions $[\beta]$ and $[\delta]$

$$G_t^n \xrightarrow{P} G_t^{\infty}$$

(3.30) and

$$\sum_{0<s\leq t} (\Delta G_s^n)^2 \xrightarrow{P} \sum_{0<s\leq t} (\Delta G_s^{\infty})^2 \quad \text{as} \quad n \uparrow \infty, \ \forall\ t \geq 0.$$

Hence, by theorem 1.8, Jacod and Mémin (1980) $\{L(M^n),\ n \geq 1\}$ is relatively compact and therefore $M^n = n^{1/2}(\hat{B}^n - \tilde{B}^n)$ converges weakly to X in $D(R^+,R^m)$ endowed with the Skorokhod topology. □

In theorem 3.2 we consider weak convergence over the whole of $R^+ = [0,\infty)$. This is in contrast to Aalen (1978) who restricts his counting processes to the interval $[0,1]$. In many statistical applications we may have to content ourselves with convergence restricted to some well defined bounded subinterval I of R^+. In fact in section 4 we illustrate an example of this where I is determined by limitations imposed on observations by a censoring mechanism in a survival analysis setting. In cases such as these the essential features of theorem 3.2 remain unchanged if we replace R^+ with I and $D(R^+,R^m)$ with $D(I,R^m)$.

3.1.4

Estimating the asymptotic covariance function. In this section we consider the problem of estimating the asymptotic covariance function or specifically the quadratic characteristic $\langle X \rangle$ (see equation 3.20) of the weak limit process X. So as to ease the exposition we introduce a number of simplifying assumptions which should not obscure the generality within which this problem may be treated.

We define the processes V^n and W^n as follows

(3.31)
$$V^n_s(i) = n(Y^{(n)}_s(i))^{-1} 1(Y^{(n)}_s(i)/n > 0)$$

$$W^n_s(i,j) = n^{-1} \sum_{k=1}^{n} Y^k_s(i) Y^k_s(j)$$

for all $s \geq 0$, $i,j \in E$ and $n = 1,2,\ldots$. For each $n = 1,2,\ldots$ define the process $\hat{\sigma}^n$ as follows:

$$(3.32)\quad \hat{\sigma}^n_t(i,j) = \begin{cases} \int_0^t V^n_s(i) d\hat{B}_s(i) - \int_0^t (V^n_s(i))^2 W^n_s(i,i) \Delta\hat{B}_s(i) d\hat{B}_s(i), & i = j; \\ - \int_0^t V^n_s(i) V^n_s(j) W^n_s(i,j) \Delta\hat{B}_s(i) d\hat{B}_s(j), & i \neq j \end{cases}$$

for all $t \geq 0$, $n = 1,2,\ldots$. We propose $\hat{\sigma}^n = \{\hat{\sigma}^n_t,\ t \geq 0\}$ as an estimator of the quadratic characteristic $\langle X \rangle$ of the weak limit X.

In theorem 3.3 (below) we exhibit sufficient conditions to show that $\hat{\sigma}^n$ converges in a suitable sense to $\langle X \rangle$ as n tends toward infinity. First we introduce some preliminary assumptions. Let $G_i > 0$ be positive B-measurable functions defined on R^+ such that $\inf_{s \geq 0} G_i(s) > 0$. Also let F_{ij} be functions defined on R^+ so that each is B-measurable $\sup_{s \geq 0} F_{ij}(s) < \infty$ and $(F_{ii}(t)/G_i(t)) \Delta B_i(t) \leq 1$ for all $t \geq 0$ and $1 \leq i,j \leq m$. Let B^c denote the measure $B - \Sigma\Delta B$ and suppose that

$$(3.33)\qquad \begin{aligned} &1)\quad \int_0^t G_i^{-1}(s)\,dB_i^c(s) = C^i(t), \quad t \geq 0, \quad i = 1,\dots,m \\ &2)\quad \sigma_{t_k}^{ij} = \begin{cases} G_i^{-1}(t_k)\Delta B_i(t_k)(1-G_i^{-1}(t_k)F_{ii}(t_k)\Delta B_i(t_k)), & i = j; \\ -G_i^{-1}(t_k)G_j^{-1}(t_k)F_{ij}(t_k)\Delta B_i(t_k)\Delta B_j(t_k), & i \neq j, \end{cases} \end{aligned}$$

where $t_k \in A$. Equation (3.33) therefore allows an explicit expression of the quadratic characteristic $\langle X\rangle$. Consider the following theorem.

Theorem 3.3. Suppose that $n^{1/2}(\hat{B}^n - \tilde{B}^n)$ converges weakly to a Gaussian process X of independent increments and having quadratic characteristic $\langle X\rangle$ (see equation 3.20) in accordance with theorem 3.2. Let $\{\hat{\sigma}^n,\ n \geq 1\}$ be the sequence of estimators for $\langle X\rangle$ defined at (3.33) and consider the following conditions.

[α] Suppose that for each i,j, $1 \leq i,j \leq m$

i) $Y_{(i)}^{(n)}/n \xrightarrow{P} G_i$ uniformly on $[0,\infty)$

ii) $W^n(i,j) \xrightarrow{P} F_{ij}$ uniformly on $[0,\infty)$

where G_i and F_{ij} satisfy (3.33) and $W^n(i,j)$ is defined at (3.31);

[β] $\{\hat{B}^n,\ n \geq 1\}$ is locally τ_k-uniformly consistent for $\{\tilde{B}^n,\ n \geq 1\}$ for some $k \geq 1$.

If condition [α] and [β] hold then for each $t \geq 0$

$$\sup_{s:s\leq t\wedge\tau_k^n} |\hat{\sigma}_s^n(i,j) - \langle X\rangle_s^{ij}| \xrightarrow{P} 0 \quad \text{as} \quad n \uparrow \infty$$

where $1 \leq i,j \leq m$.

Proof. We first observe that since $G_i > 0$ condition (i) of [α] implies easily that $\{\hat{B}^n,\ n \geq 1\}$ is locally τ_k-uniformly consistent for B whenever [β] holds.

To prove the theorem it suffices to observe that according to the definition of $\hat{\sigma}^n(i,j)$ each term is either of the form $\int H^n_{ij} d\hat{B}^n_i$ or $\int H^n_{ij} \Delta\hat{B}^n_j d\hat{B}^n_i$ where in each case H^n_{ij} converges uniformly to a function h on $[0,\infty)$ and in the second case $\Delta\hat{B}^n_j$ is locally τ_k-consistent for ΔB_j. Furthermore, on bounded intervals of the form $[0,t]$ we have $\sup_{s:s\leq t} h(s)$ and $\sup_{s:s\leq t} h(s)\Delta B_j(s)$ both finite. Thus,

$$(3.34)\quad \sup_{s:s\leq t\wedge\tau^n_k} \left|\int_0^s H^n_{ij}(s)d\hat{B}^n_i - \int_0^s h(s)dB_i(s)\right| \leq \sup_{s:s\leq t\wedge\tau^n_k} \int_0^s \left|H^n_{ij}(s)-h(s)\right| d\hat{B}^n_i$$

$$+ \sup_{s:s\leq t\wedge\tau^n_k} \left|\int_0^s h(s)(d\hat{B}^n_i(s)-d\tilde{B}_i(s))\right|$$

with a corresponding inequality for the alternate case. From the conditions $[\alpha]$ and $[\beta]$ of the theorem and the comments above it is easily seen that the right hand side of (3.34) may be made arbitrarily small in probability. Therefore upon applying these expressions in appropriate order to $|\hat{\sigma}^n(i,j) - \langle X\rangle^{ij}|$ the conclusion of the theorem readily follows. □

As similar remark applies to theorem 3.3 as was made at the end of theorem 3.2. That is, the limitations of a specific applied problem may allow us to confine our attention to bounded sub-intervals of R^+. For example, in the next section we consider a function G, analogous in that situation to the present function G_i, which is taken to be nonzero on an interval $[0,\tau]$ or $[0,\tau)$ where $\tau > 0$ is arbitrary.

4. APPLICATION TO SURVIVAL ANALYSIS

We include under the heading survival analysis any application to life-testing, medical clinical trials, biological experimentation involving the observation of independent and possibly censored positive random variables. The specific model we consider is described as follows:

Random Censorship Model: For each $n = 1,2,\ldots$ X_i and U_i $i = 1,\ldots,n$ are $2n$ independent positive random variables with X_i or U_i almost surely finite for each i. Assume X_i has distribution function F and U_i has distribution function L, which in general may be defective. The observable random variables $\tilde{X}_i$ and δ_i are defined by $\tilde{X}_i = X_i \wedge U_i$ and $\delta_i = 1(X_i \le U_i)$. When $\delta_i = 0$ we say the observation X_i has been right censored at U_i. □

To analyze problems arising in the random censorship model we define the following processes for each $i = 1,\ldots,n$:

$$
\begin{aligned}
N_t^i &= 1(\tilde{X}_i \le t,\ \delta_i = 1), \quad t \ge 0 \\
Y_t^i &= 1(\tilde{X}_i \ge t), \quad t \ge 0 \qquad (4.1) \\
M_t^i &= N_t^i - \int_0^t Y_s^i B\{ds\}, \quad t \ge 0
\end{aligned}
$$

where B is a deterministic function given by

$$
B(t) = \int_0^t (1-F(s-))^{-1} F\{ds\}, \quad t \ge 0 \qquad (4.2)
$$

and is the so-called cumulative hazard or risk function which uniquely determines the distribution F.

The process $N^i = \{N_t^i, t \geq 0\}$ is called a simple counting process and is equal to 0 until the ith observation time elapses and has not been censored and is equal to 1 thereafter. The process $Y^i = \{Y_t^i, t \geq 0\}$ is called a risk process and is equal to 1 as long as the ith observation remains under observation (i.e. the ith observation continues to survive and has not been censored).

Let $(\Omega^n, \mathcal{F}^n, P^n)$ be a probability space on which X_i, U_i are defined, $i = 1,\ldots,n$. We define a filtration $F^n = \{\mathcal{F}_t^n, t \geq 0\}$ as follows

$$(4.3)\quad \mathcal{F}_t^n = \mathcal{F}_0^n \vee \sigma\text{-}(1(\tilde{X}_i \leq s),\ \delta_i 1(\tilde{X}_i \leq s),\ s \leq t,\ i = 1,\ldots,n),\ t \geq 0$$

where $\mathcal{F}_0^n$ contains the P^n-null sets of $\mathcal{F}^n$ and their subsets. According to theorem 3.1.1, Gill (1980) the process $M^i = \{(M_t^i, \mathcal{F}_t^n), t \geq 0\}$ is a square integrable martingale so that N^i has compensator $A^i = \{(A_t^i, \mathcal{F}_t^n), t \geq 0\}$ which satisfies

$$(4.4)\qquad A^i\{0\} = \int_0 Y_s^i B\{ds\} \quad \text{for any Borel set} \quad 0 \in \sigma(R^+)$$

where by virtue of its left-continuity, the process Y^i is F^n-predictable and B is a Borel measure. Thus according to definition 2.1 N^i is a Poisson-type counting process with parameter B.

We consider the statistical problem of estimating the cumulative risk function B. In the case where F has a density f relative to Lebesgue measure then $\alpha = f/(1-F)$ defines the intensity of the counting process N^i and the theory of Aalen (1978) would apply. Gill (1980), on the other hand, uses the counting process approach to study this problem but omits the assumption that the density exists. Hence, the theory of inference in the random censorship model developed by Gill (1980) arises as a special case of our theory for Poisson-type counting processes.

To estimate B we define the processes

$$M_t^{(n)} = \sum_{i=1}^{n} M_t^i, \quad t \geq 0$$

(4.5)
$$Y_t^{(n)} = \sum_{i=1}^{n} Y_t^i, \quad t \geq 0$$

$$J_t^n = 1(Y_t^{(n)} > 0), \quad t \geq 0.$$

Then the estimator $\hat{B}^n$ is given by

(4.6)
$$\hat{B}_t^n = \int_0^t (Y_s^{(n)})^{-1} J_s^n dN_s^{(n)}, \quad t \geq 0$$

where $N^{(n)} = \Sigma_{i=1}^n N^i$.

The estimator $\hat{B}^n$ in this case is called the empirical cumulative hazard function and is studied by Gill (1980) under a more general random censorship model than is defined above. In chapter 4 he considers the asymptotic properties of $\hat{B}^n$ including consistency and weak convergence to a Gaussian process of independent increments. We are interested here to give a simpler proof of weak convergence for these processes based on theorem 3.2.

Define the process $\tilde{B}^n$ by

(4.7)
$$\tilde{B}_t^n = \int_0^t J_s^n B\{ds\}, \quad t \geq 0$$

and consider the following lemma.

Lemma 4.1. For each $n = 1,2,\ldots$ let $\hat{B}^n$ and $\tilde{B}^n$ be defined by (4.6) and (4.7), respectively. Then we have that $\hat{B}^n - \tilde{B}^n$ is of class $\mathcal{M}^2(F^n, P^n)$.

Proof. Observe that for each $t \geq 0$ and $\omega \in \Omega^n$

$$\hat{B}^n_t - \tilde{B}^n_t = \int_0^t Y_s^{(n)} J_s^n dM_s^{(n)} \tag{4.8}$$

where $M^{(n)}$ is a square integrable martingale defined at (4.5) and the integral is interpreted as an ordinary Lebesgue-Stieltjes integral. Since $Y^{(n)}$ is an integer valued process it follows that

$$\sup_{t: t \geq 0} (Y_t^{(n)})^{-1} J_t^n \leq 1$$

where the convention $\frac{0}{0} = 0$ has been invoked. Further $(Y^{(n)})^{-1} J^n$ is a predictable process so that $\hat{B}^n - \tilde{B}^n \in \mathcal{M}^2(F^n, P^n)$ follows from lemma 2.2. □

In theorem 4.11, page 56, Gill (1980) it is proved that $\hat{B}$ converges uniformly to B on sets of the form $[0,t]$ for which $Y_t^n \overset{P}{\to} \infty$ as $n \to \infty$. We proceed by showing that the normalized error of estimation process $n^{1/2}(\hat{B}^n - \tilde{B}^n)$ converges weakly to a Gaussian process.

Theorem 4.1. Let Y^n/n converge uniformly on $[0,\infty)$ to a function y in probability as $n \to \infty$ and let $I = \{t: y(t) > 0\}$. Let X be a zero-mean Gaussian process of independent increments with characteristic function

$$\langle X\rangle_t = \int_0^t (y(s))^{-1}(1-\Delta B(s))B\{ds\}, \quad t \in I.$$

Then $n^{1/2}(\hat{B}^n-\tilde{B}^n)$ converges weakly to X in the space $D(I,R)$ endowed with the Skorokhod topology.

Proof. If $u = \sup\{t: y(t) > 0\}$ we assume $u > 0$ otherwise there is nothing to prove. Thus $I = [0,u)$ when $y(u) = 0$ or $I = [0,u]$ when $y(u) > 0$. By lemma 4.1 $n^{1/2}(\hat{B}^n-\tilde{B}^n)$ is a square integrable martingale on I so we apply Theorem 3.2 with $m = 1$. By lemma 3.1 we have

$$\langle n^{1/2}(\hat{B}^n-\tilde{B}^n)\rangle_t = \int_0^t n(Y_s^{(n)})^{-1}J_s^n(1-\Delta B(s))B\{ds\}, \quad t \in I. \tag{4.9}$$

It is easily shown that $n(Y_s^{(n)})^{-1}J_s^n \overset{P}{\to} (y(s))^{-1}$ uniformly over I as $n \to \infty$, thus proving condition $[\beta]$ and $[\delta]$ of theorem 3.2 with $c = \int y^{-1}dB^c$ and $\sigma_j^2 = y(t_j)(1-\Delta B(t_j))\Delta B(t_j)$ where $\{t_j\}$ is the set of discontinuities of B. To prove $[\alpha]$ we observe that for each $t \in I$ and $\epsilon \in (0,1]$ we have

$$n \int (Y^{(n)})^{-1}((Y^{(n)})^{-1}J^n > \epsilon)dB^c \tag{4.10}$$

gives the left-hand-side of $[\alpha]$ in theorem 3.2. Since y is strictly positive on I and $n^{-1}Y^{(n)} \overset{P}{\to} y$ we have $Y^{(n)} \overset{P}{\to} \infty$ on I as $n \to \infty$ so that (4.10) converges to zero in probability over I as $n \to \infty$. Thus $[\alpha]$ is proved.

To complete the proof it remains to show that the law of the random variable $Z_j^n = n^{1/2}(Y_{t_j}^{(n)})^{-1}J_{t_j}^n(\Delta N_{t_j}^{(n)} - Y_{t_j}^{(n)}\Delta B(t_j))$ conditional on $\mathcal{F}_{t_j-}^n$ converges in probability to the law of a Gaussian random variable U_j with zero-mean and variance $\sigma_j^2 = y(t_j)(1-\Delta B(t_j)\Delta B(t_j)$. But according to theorem 3.1.1, Gill (1980) $\Delta N_{t_j}^{(n)}$ conditional on $F_{t_j -}^n$ is a Binomial random variable with parameters $Y_{t_j}^{(n)}$ and $\Delta B(t_j)$ from which it is easy to show that the Laplace transform $\varphi^n(s,\omega)$ of Z_j^n conditional on $F_{t_j -}^n$ converges in probability to $\exp\{-\frac{1}{2} s^2(y(t_j)\Delta B(t_j)(1-\Delta B(t_j))^{-1}\}$ as $n \to \infty$ for all real s. On combining each of these results theorem 4.1 follows from theorem 3.1. □

The estimator $\hat{B}^n$ can be used to construct the product-limit or Kaplan-Meier estimator of the distribution function F. This problem has been considered by Gill (1980) and more recently by Phelan (1985) using methods developed for Poisson-type counting processes. The proofs of weak convergence in Phelan (1985) and here are to be compared with those of Gill (1980) which rely on an interesting but elaborate construction. Our proof is based on adaptations of Jacod, et al. (1982) and Lipster and Shiryayev (1980) as was anticipated by Gill (1980) (see page 78).

5. FURTHER TOPICS

Extensions of this research have been considered by the author. These include a family of one and two-sample hypotheses tests, likelihoods for Poisson type counting processes and applications to estimation from Markov renewal processes. Some of this work is included in the author's Ph.D. thesis and is to appear elsewhere.

The author is currently considering likelihood based inference procedures for Poisson type counting processes which will include semiparametric Cox-type regression models as well as fully parametric models. Applications of these models to consumer decision processes are to appear.

References

[1] Aalen, O. (1978). Nonparametric inference for a family of counting processes. Ann. Math. Statist., 6, 707-726.

[2] Aalen, O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scand. J. Statist., 5, 141-150.

[3] Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study. Ann. Math. Statist., 10, 1100-1120.

[4] Borgan, O. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. Scand. J. Statist., 11, 1-16.

[5] Cox, D.R. (1972). Regression models and life-tables. J. Roy. Statist. Soc. B, 34, 187-220.

[6] Gill, R.D. (1980). Censoring and Stochastic Integrals. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

[7] Gill, R.D. (1983). Large sample behavior of the product-limit estimator on the whole line. Ann. Math. Statist., 11, 49-58.

[8] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. Z. Wahrs. verw. Geb., 31, 235-253.

[9] Jacod, J., Klopotowski, A. and Mémin, J. (1982). Central limit theorem and functional convergence to a process of independent increments: martingale method. Inst. Henri Poincare, 18, 1-45.

[10] Jacod, J. and Memin, J. (1980). Un nouveau critere de compacite relative pour des suites de processus. Sem. Prob. Rennes, 79, 1-27.

[11] Lenglart, E. (1977). Relation of domination between two processes. Am. Inst. H. Poincare, B, XIII, 171-179.

[12] Lipster, R.S. and Shiryayev, A.N. (1974). Statistics of Random Processes I and II. Springer-Verlag, New York.

[13] Lipster, R.S. and Shiryayev, A.N. (1980). A functional central limit theorem for semimartingales. Theory of Prob. and its Appls., 25, 667-688.

[14] Metivier, M. (1982). Semimartingales: a Course on Stochastic Processes, Walter de Gruyter, New York.

[15] Murali-Rao, K. (1969). On decomposition theorems of Meyer. Math. Scand., 24, 66-78.

[16] Phelan, M.J. (1985). Weak convergence for martingale related to a counting process with applications to survival analysis. Technical Report in progress, Cornell University.

[17] Prentice, R.L. and Self, S.G. (1982). Asymptotic distribution theory for Cox-type regression models with general relative risk form. Ann. Math. Statist., 11, 804-813.

[18] Rebolledo, R. (1980). Central limit theorems for local martingales. Z. Wahrs. Ver. Geb., B, 51, 269-286.

[19] Royden, H.L. (1968). Real Analysis. Macmillan, New York.

[20] Shiryayev, A.N. (1981). Martingales: recent developments and applications. Int. Statist. Rev., 49, 199-233.

[21] Slud, E.V. (1984). Sequential linear rank tests for two-sample censored survival data. Ann. Math. Statist., 12, 551-571.