Z. Wahrscheinlichkeitstheorie verw. Gebiete
38, 261–277 (1977)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete


# Weak Convergence of Stochastic Integrals Related to Counting Processes

Odd O. Aalen*
University of California, Berkeley and University of Copenhagen

## Table of Contents



## 1. Introduction

During the last few years a number of authors have developed a theory of point processes on the positive real line based on the modern theory of martingales and stochastic integrals. The development seems to have started with the thesis of Brémaud (1972), and it was continued with the papers of Boel, Varaiya and Wong (1975a, b), Brémaud (1974), Dolivo (1974), Segall and Kailath (1975a, b), Varaiya (1975) and possibly others of whom the author may not be aware. Jacod (1973, 1975) has taken a somewhat different approach and his work is an important supplement to that of the above mentioned authors.

The point of view taken in this theory concentrates on the counting aspect of the point process. One defines the counting process as the process counting the events along the time axis.

The present author is concerned with statistical applications of this counting process theory. In my Ph.D. dissertation (Aalen, 1975) I have studied one possible model for such applications which contains certain Markov process models as special cases. The bulk of that theory will be published elsewhere. The

* Research supported by The Norwegian Research Council for Science and the Humanities

Author's address: Rosenborggaten 6, Oslo 3, Norway

present paper contains an asymptotic result which is basic for that theory and which also may have an independent interest.

We will proceed to give a short description of a part of the counting process theory. For a fuller account one should consult one of the papers mentioned above.

Let $\Omega$ be an abstract space. A multivariate stochastic process $\{\mathbf{N}(t)=(N_1(t),\ldots,N_k(t));\ t\in[0,\ 1]\}$ defined on $\Omega$ is called a (multivariate) counting process when the sample paths $\mathbf{N}(t,\omega)=(N_1(t,\omega),\ldots,N_k(t,\omega))$ have the following properties for each $\omega\in\Omega$:

(i) $N_i(t,\omega)$; $i=1,\ldots,k$; are right-continuous increasing step functions. Each has a finite number of jumps with each jump positive and equal to 1. Also, for each $i$, we have $N_i(0,\omega)=0$.

(ii) The functions $N_i(t,\omega)$ and $N_j(t,\omega)$ never jump at the same time when $i\neq j$.

Let $\mathcal{N}_t$ be the $\sigma$-algebra on $\Omega$ generated by $\{\mathbf{N}(s),\ s\leqq t\}$. According to Corollary 2.5 of Boel et al. (1975a) the family $\{\mathcal{N}_t,\ t\in[0,1]\}$ is right-continuous, i.e. $\bigcap_{h>0}\mathcal{N}_{t+h}=\mathcal{N}_t$ for every $t$. Clearly, the family is also increasing. Put $\mathcal{N}=\mathcal{N}_1$.

Let $\{\mathcal{F}_t,\ t\in[0,1]\}$ be another right-continuous increasing family of $\sigma$-algebras such that $\mathcal{N}_t\subset\mathcal{F}_t$ for each $t$. Put $\mathcal{F}=\mathcal{F}_1$. Throughout the paper all processes will be studied relative the $\sigma$-algebras $\mathcal{F}_t$. Let $P$ denote a probability measure on $(\Omega,\mathcal{F})$.

We call (see e.g. Meyer 1966, D33) a nonnegative random variable $T$ a stopping time if the event $\{T\leqq t\}$ belongs to $\mathcal{F}_t$ for every $t\in[0,1]$.

Following Boel et al. (1975a) we give the following formal definition of the jump times $T_n$ of $\mathbf{N}$:

$$T_0\equiv 0,\qquad T_{n+1}=\inf\{t\,|\,t>T_n,\mathbf{N}(t)\neq\mathbf{N}(T_n)\}.$$

$T_n$ is for all $n$ a stopping time relative to $\{\mathcal{N}_t\}$ and hence also relative to $\{\mathcal{F}_t\}$ (*ibid.*, Corollary 2.2).

We will assume throughout the paper that $\mathbf{N}$ satisfies the following basic property (*ibid.*, Section III):

(1.1) **Assumption.** *The jump times $T_n$ are totally inaccessible relative to $\{\mathcal{F}_t\}$, i.e. for each n and any sequence $S_1, S_2, \ldots$ of stopping times*

$$P\{\lim_k S_k=T_n \text{ and } S_k<T_n \text{ for all } k\}=0.$$

See Meyer (1966, Section VII.5) for a discussion of this concept and its meaning. We also assume:

(1.2) **Assumption.** $EN_i(1)<\infty\ \ i=1,\ldots,k$.

We say that a process $X(t)$ defined on $\Omega$ is *adapted to* $\{\mathcal{F}_t\}$ if $X(t)$ is measurable with respect to $\mathcal{F}_t$ for all $t\in[0,1]$. We need the following definitions and concepts (see Kunita and Watanabe (1967), Meyer (1967, 1971)).

(1.3) **Definition.** Let $M$ be a real-valued process defined on $[0,1]$ and adapted to $\{\mathcal{F}_t\}$. Assume that $M$ has sample functions that are right-continuous and have left-hand limits. $M$ is called a *square integrable martingale* if $\sup_t E(M(t)^2) < \infty$ and $E(M(t) | \mathscr{F}_s) = M(s)$ whenever $s < t$. The space of these processes is denoted by $\mathscr{M}^2$.

Let $M_1, M_2$ be two elements of $\mathscr{M}^2$ and let $\langle M_1, M_2 \rangle$ be the process defined in Theorem 1.1 of Kunita and Watanabe (1967). If $M \in \mathscr{M}^2$, then $\langle M, M \rangle$ is the quadratic variation process of $M$ as defined by Meyer (1966). Kunita and Watanabe defined $M_1$ and $M_2$ as being orthogonal if $\langle M_1, M_2 \rangle = 0$. This is equivalent to $M_1 M_2$ being a martingale with respect to $\{\mathscr{F}_t\}$, which is just Meyer's definition of orthogonality.

The following result is basic in the counting process theory:

(1.4) **Theorem.** *There exists an increasing, continuous, k-variate process* $\mathbf{A}(t) = (A_1(t), \ldots, A_k(t))$ *adapted to* $\{\mathscr{F}_t\}$ *and with* $\mathbf{A}(0) = 0$, *such that*

(i) $M_i = N_i - A_i \in \mathscr{M}^2 \quad i = 1, \ldots, k,$

(ii) $\langle M_i, M_i \rangle = A_i \quad i = 1, \ldots, k,$

(iii) $\langle M_i, M_j \rangle = 0$ *whenever* $i \neq j$.

A version of this theorem formulated by means of local martingales is given by Proposition 3.2 and Lemma 3.1 of Boel et al. (1975a). Our Assumption 1.2 allows us to make the stronger statement using square integrable martingales (cfr. Theorem 2.4.7 of Dolivo (1974)).

Throughout this paper we will assume:

(1.5) **Assumption.** *Each sample function of the process* $A_i$ *is absolutely continuous with respect to Lebesgue measure on* $[0, 1]$.

From this assumption it follows that there exists a process $\Lambda(t) = (\Lambda_1(t), \ldots, \Lambda_k(t))$ which is predictable (see e.g. Meyer, 1971) with respect to $\{\mathscr{F}_t\}$ and satisfies $A_i(t) = \int_0^t \Lambda_i(s)\, ds$, $i = 1, \ldots, k$. $\Lambda$ is called the *intensity process* of $\mathbf{N}$ with respect to $\{\mathscr{F}_t\}$. Results on the existence of counting processes with a given intensity process can be found in the references mentioned above, see in particular Boel et al. (1975b, Theorem 3.3 and Prop. 3.4) and Jacod (1975).

The object of the present paper is to study weak convergence of stochastic integrals with respect to the martingales $M_i$ of Theorem 1.4. Such integrals have been studied mainly by Meyer and Doléans-Dadé, see Meyer (1967, 1971) and Doléans-Dadé and Meyer (1970). As shown in Aalen (1975) such integrals play an important role in certain parts of the inference theory for counting processes. An example is given in Section 3 of the present paper.

Let $\{H(t), t \in [0, 1]\}$ be a predictable process. Let $M \in \mathscr{M}^2$. $L^2(M)$ is defined as the class of all predictable processes $H$ satisfying

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

The stochastic integral of an element $H$ of $L^2(M)$ with respect to $M$ on the interval $[0, t]$ is defined e.g. in Meyer (1971). We denote it by

$$\int_0^t H(s)\, dM(s).$$

The corresponding stochastic process that arises by letting $t$ run through $[0, 1]$ is denoted by $\int H\, dM$. We summarize the properties of this process (see e.g. Doléans-Dadé and Meyer (1970)):

(1.6) $\int H\, dM \in \mathcal{M}^2$.

(1.7) $\langle \int H\, dM, \int H\, dM \rangle = \int H^2\, d\langle M, M \rangle$

(1.8) Let $K \in \mathcal{M}^2$, $G \in L^2(K)$. Then $\langle \int H\, dM, \int G\, dK \rangle = \int HG\, d\langle M, K \rangle$.

Integrals with respect to processes which are not elements of $\mathcal{M}^2$ are always supposed to be Lebesgue-Stieltjes integrals.

Let $M_1, \ldots, M_k$ be the martingales of Theorem 1.4 and define $\mathbf{H} = (H_1, \ldots, H_k)$ where $H_i$ is an element of $L^2(M_i)$ for $i = 1, \ldots, k$. (1.6)–(1.8) implies:

(1.9) $\int H_i\, dM_i \in \mathcal{M}^2 \qquad i = 1, \ldots, k,$

(1.10) $\langle \int H_i\, dM_i, \int H_i\, dM_i \rangle = \int H_i^2 \Lambda_i\, dt \qquad i = 1, \ldots, k,$

(1.11) $\langle \int H_i\, dM_i, \int H_j\, dM_j \rangle = 0 \qquad$ whenever $i \neq j$.

We will consider only certain sub classes of the families $L^2(M_i)$ $i = 1, \ldots, k$. First we will make a requirement which guarantees that the stochastic integrals $\int_0^t H_i(s)\, dM_i(s)$, $i = 1, \ldots, k$, coincide with the corresponding Lebesgue-Stieltjes integrals. By Proposition 3 of Doléans-Dadé and Meyer (1970) and Section 2 of Jacod (1975) this will be the case if the following requirement holds:

*Requirement A.* $E \int_0^1 |H_i(s)|\, dN_i(s) < \infty. \qquad i = 1, \ldots, k.$

As a preparation for the asymptotic results of the next section we will also make another, rather technical, requirement. It is needed in order to make our proofs work, but it is probably not really necessary for the weak convergence result to hold. Nevertheless, it is quite easily verified in most of the cases of application which the author has in mind. Put

$$\bar{N}(t) = \sum_{i=1}^{k} N_i(t), \qquad \bar{\Lambda}(t) = \sum_{i=1}^{k} \Lambda_i(t).$$

*Requirement B.* There exists a non-decreasing process $\{\Phi(t),\ t \in [0, 1]\}$ defined on $(\Omega, \mathcal{F})$ such that $E\,\Phi(1) < \infty$ and the following holds:

(i) $|H_i^2(t)\Lambda_i(t) - H_i^2(s)\Lambda_i(s)| \leqq \Phi(t) - \Phi(s)$ for $0 \leqq s < t \leqq 1$ and $i = 1, \ldots, k$,

(ii) $H_i^2(t)\Lambda_i(t) \leqq \Phi(1)$ for $0 \leqq t \leqq 1$ and $i = 1, \ldots, k$,

(iii) $E[\Phi(1)\bar{N}(1)] < \infty$,

(iv) $E\left[\Phi(1) \int_0^1 \bar{\Lambda}(s)\, ds\right] < \infty$.

In accordance with well known invariance principles for martingales it is reasonable to believe that under certain conditions the processes $\int H_i\,dM_i$, $i=1,\ldots,k$, will converge weakly to independent normal processes, each with independent increments. In the next section we will give a precise statement of such a result.

## 2. Main Result

Let $(\mathbf{N}_n, \Lambda_n)$, $n=1,2,\ldots$, be a sequence of processes of the kind defined in the previous section. Write $\mathbf{N}_n=(N_{1,n},\ldots,N_{k,n})$ and similarly for $\Lambda_n$. Assumptions 1.1, 1.2 and 1.5 shall hold for each $n$. Let $\mathbf{M}_n=(M_{1,n},\ldots,M_{k,n})$ be defined for each $n$ as in Theorem 1.4. Let $\mathbf{H}_n=(H_{1,n},\ldots,H_{k,n})$ be a sequence of processes such that for each $n$ $H_{i,n}$ is a member of $L^2(M_{i,n})$, $i=1,\ldots,k$, satisfying Requirements A and B. Of course, the process $\Phi$ is allowed to vary with $n$.

Let $D$ be the space of real functions on $[0,1]$ which are right-continuous and have left hand limits. The members of $\mathcal{M}^2$ are random elements of $D$. Let $D$ be equipped with the Skorohod topology (Billingsley (1968), Chapter 3). Let $D^m$ be the Cartesian product of $D$ with itself $m$ times, and let $D^m$ be equipped with the corresponding product topology. When we talk about weak convergence of stochastic processes, we will always mean weak convergence of random elements of $D^m$ with respect to this product topology. We denote weak convergence by $\Rightarrow$. See Billingsley (1968) for the general theory of weak convergence.

Let $W$ be the Wiener process. Using the notation of Section 1, we see that $\langle W,W\rangle(t)=t$, hence $g\in L^2(W)$ if $g\in L^2(0,1)$, and so in that case $\int g\,dW$ is well-defined as a stochastic integral. It is a normal process with independent increments.

We put $\bar{N}_n=\sum_{i=1}^{k} N_{i,n}$ and $\bar{\Lambda}_n=\sum_{i=1}^{k}\Lambda_{i,n}$, and let $\xrightarrow{P}$ denote convergence in probability. We also write for each $n$:

$$Y_{i,n}(t)=\int_0^t H_{i,n}(s)\,dM_{i,n}(s) \qquad i=1,\ldots,k,$$

$$\mathbf{Y}_n=(Y_{1,n},\ldots,Y_{k,n}).$$

**Theorem 2.1.** *Make the following assumptions:*

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

$$\int_0^t H_{i,n}^2(s)\,\Lambda_{i,n}(s)\,ds \xrightarrow{P} \int_0^t g_i^2(s)\,ds \qquad \textit{for } 0\leq t\leq 1 \textit{ and } i=1,\ldots,k.$$

(b) *Let $S_{i,n}^{(m)}$, $m=1,2,\ldots,N_{i,n}(1)$, be the successive jump times of the process $N_{i,n}$.*

*We assume:*

$$\sum_{i=1}^{k} E\left\{\sum_{m=1}^{N_{i,n}(1)} [H_{i,n}^2(S_{i,n}^{(m)})\,I(|H_{i,n}(S_{i,n}^{(m)})|>\varepsilon)]\right\}\to 0 \qquad \textit{for every } \varepsilon>0.$$

*Let* $W_1, \ldots, W_k$ *be independent Wiener processes and put* $Y_i(t) = \int_0^t g_i(s)\, dW_i(s)$ *and* $\mathbf{Y} = (Y_1, \ldots, Y_k)$. *Then*

$$\mathbf{Y}_n \Rightarrow \mathbf{Y}.$$

*Remark.* We have

$$\langle Y_{i,n}, Y_{i,n} \rangle = \int H_{i,n}^2 \Lambda_{i,n}(s)\, ds.$$

Hence condition (a) serves to stabilize the quadratic variation process of $Y_{i,n}$. It is reasonable that such a stabilization is needed since $Y_i$ has a deterministic quadratic variation process.

Condition (b) is a sort of Lindeberg condition. It guarantees that the jumps (or discontinuities) of the processes disappear in the limit. Note that this condition implies that the $H$-processes become small. This means that the process $\Lambda_n$ grows large when $n$ increases, and hence that the number of jumps in $\mathbf{N}_n$ grows large. This is obviously required in order to get a normal process in the limit.

Theorem 2.1 is a revised version of Theorem 4.1 of Aalen (1975).

## 3. An Example

In this section we will apply Theorem 2.1 to a simple Markov chain example. We will start by defining a special bivariate counting process and then explain the connection with a Markov chain.

Let $n_1$ and $n_2$ be given nonnegative integers and let $\alpha_1(t)$ and $\alpha_2(t)$ defined on $[0,1]$ be nonnegative left-continuous and uniformly bounded functions with right-hand limits. Also, assume that $\alpha_1$ and $\alpha_2$ are of bounded variation. Let $\mathbf{N} = (N_1, N_2)$ be a bivariate counting process and let $\{\mathscr{F}_t\}$ be the $\sigma$-algebras generated by $\mathbf{N}$. We want $\mathbf{N}$ to have an intensity process $\Lambda = (\Lambda_1, \Lambda_2)$ of the following form:

$$\Lambda_i(t) = \alpha_i(t)\, \Gamma_i(t) \qquad i = 1, 2$$

where

$$\Gamma_1(t) = n_1 - N_1(t-) + N_2(t-), \qquad \Gamma_2(t) = n_2 + N_1(t-) - N_2(t-).$$

Jacod (1975, Theorem 3.6) shows that there exists a unique measure $P$ on $\mathscr{F}$ giving $\mathbf{N}$ the intensity process $\Lambda$ in the local martingale sense of Jacod's paper. That our Assumption 1.1 is fulfilled follows from Theorem 2.4.7 of Dolivo (1974). Our Assumption 1.2 follows from the fact that $\Lambda_1$ and $\Lambda_2$ are bounded by constants and hence $N_1$ and $N_2$ are dominated by Poisson processes. Theorem 2.4.8 of Dolivo (1974) then proves that $\Lambda$ is an intensity process in the sense of Theorem 1.4 and Assumption 1.5 of the present paper.

The counting process $\mathbf{N}$ can be thought of as arising from a Markov chain in the following way: Assume that we have a Markov process on the state space

$\{1,2\}$. The infinitesimal transition probabilities (or forces of transition) from state 1 to 2 and from 2 to 1 are $\alpha_1(t)$ and $\alpha_2(t)$ respectively. Assume that $n_1+n_2$ "particles" are moving around on the state space independently of each other, $n_1$ starting in state 1 and $n_2$ starting out in state 2. Then $N_1$ and $N_2$ can be thought of as counting the transitions from state 1 to 2 and from 2 to 1 respectively. $\Gamma_1(t)$ and $\Gamma_2(t)$ denote the numbers of particles in state 1 and 2 respectively at time $t$.

This connection between the Markov chain and the counting process $\mathbf{N}$ can be treated rigorously but that will not be done here. Aalen (1975, Section 5 D) gives a more rigorous discussion.

We will now pass to the stochastic integrals we have in mind. Define for $i=1,2$:

$$J_i(t)=I(\Gamma_i(t)>0),$$
$$H_i(t)=J_i(t)\,\Gamma_i t)^{-1},$$

where we define $0\cdot\infty=0$.
$H_1$ and $H_2$ are left-continuous processes with right-hand limits and bounded by the constant 1. Hence they are elements of $L^2(M_i)$, $i=1,2$, respectively, where $M_1$ and $M_2$ are the martingales corresponding to $N_1$ and $N_2$ (see Theorem 1.4). Hence the following stochastic integrals are well defined:

$$M_i^*(t)=\int_0^t H_i(s)\,dM_i(s) \qquad i=1,2.$$

Requirement A is trivially fulfilled in this case by the boundedness of $H_1$ and $H_2$. Hence the integral can alternatively be taken as a Lebesgue-Stieltjes integral.

The reason for being interested in the $M_i^*$ is the following: We can write $(i=1,2)$:

$$M_i^*(t)=\hat{\beta}_i(t)-\beta_i(t)$$

where

$$\beta_i(t)=\int_0^t J_i(s)\,\alpha_i(s)\,ds,$$

$$\hat{\beta}_i(t)=\int_0^t H_i(s)\,dN_i(s).$$

From the martingale property of $M_i^*$ it follows that

$$E\,\hat{\beta}_i(T)=E\,\beta_i(T)$$

for any stopping time $T\leqq 1$. Hence the process $\hat{\beta}_i$ is an unbiased estimator of $\beta_i$. One can say that $\beta_i$ is an empirical cumulative force of transition. Such estimators are studied in a general framework in Aalen (1975). They generalize the empirical cumulative hazard rate of Nelson (1969) and are related to the empirical survival function of Kaplan and Meier (1958).

We will show how the theorem of the previous section can be applied to the $M_i^*$.

We will first verify Requirement B. We have ($i=1,2$):

$$H_i^2(t)\Lambda_i(t)=\alpha_i(t)J_i(t)\Gamma_i(t)^{-1}.$$

Remember that the $\alpha_i$ are supposed to be of bounded variation. Put

$$c=\sup_{i,t}\alpha_i(t)$$

and let $\psi(t)$ be a nonnegative non-decreasing function satisfying

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

for $0\leqq s<t\leqq 1$ and $i=1,2$.

We define:

$$\Phi(t)=c(N_1(t)+N_2(t))+(n_1+n_2)\psi(t).$$

The processes $H_i^2\Lambda_i$ have jumps at the same times as $N_1+N_2$. The size of these jumps are bounded by $c$. Apart from the jumps the variation in $H_i^2\Lambda_i$ is only due to the variation in $\alpha_i$. Hence parts (i) and (ii) of Requirement B are fulfilled. Parts (iii) and (iv) are consequences of the boundedness of $\Lambda$ and the implication that $N_1$ and $N_2$ are dominated by Poisson processes.

Let $W_1$ and $W_2$ be independent Wiener processes. Let $P_{ij}(t)$ denote the probability of being in state $j$ at time $t$ for a particle starting out in state $i$ at time 0. We can now prove the next proposition as a consequence of Theorem 2.1. It should be remembered that the stochastic processes in the statement of the proposition and in its proof depend on $n_1$ and $n_2$ even though this is not shown in the notation.

(3.1) **Proposition.** *Assume that $n_1$ and $n_2$ increase to infinity in such a way that*

$$\frac{n_1}{n_2}\to a$$

*where $a$ is a positive constant. Then the processes $\sqrt{n_i}(\hat\beta_i-\beta_i)$, $i=1,2$, converge weakly to the normal processes*

$$\int g_i\,dW_i \qquad i=1,2$$

*where*

$$g_1(t)^2=\alpha_1(t)(P_{11}(t)+a^{-1}P_{21}(t))^{-1},$$
$$g_2(t)^2=\alpha_2(t)(P_{22}(t)+aP_{12}(t))^{-1}.$$

*Proof.* Our situation is the same as that treated in Theorem 2.1 if one identifies $H_{i,n}$ with $n_i^{\frac{1}{2}}H_i$ and $H_{i,n}^2\Lambda_{i,n}$ with $n_iH_i^2\Lambda_i$, which can alternatively be written $n_i\alpha_iH_i$.

Define $K_i^1 = n_i - N_i(1)$, $i=1,2$. These variables have binomial $(n_i, p_i)$ distributions where

$$p_i = \exp\left(-\int_0^1 \alpha_i(s)\,ds\right) \qquad i=1,2.$$

Define $K_i = 1$ if $K_i^1 = 0$ and $K_i = K_i^1$ otherwise. We will use the random variables $K_1$ and $K_2$ to verify the assumptions of Theorem 2.1. For each $t$ the following holds:

$$\frac{1}{n_1}\Gamma_1(t) \xrightarrow{p} P_{11}(t) + a^{-1} P_{21}(t),$$

$$\frac{1}{n_2}\Gamma_2(t) \xrightarrow{p} P_{22}(t) + a P_{12}(t).$$

This implies:

$$n_i H_i^2 \Lambda_i(t) \xrightarrow{p} g_i(t)^2 \qquad i=1,2.$$

The left hand side is bounded above by $c\, n_i/K_i$ which converges in probability to $c\, p_i^{-1}$. Hence, Assumption (a) of Theorem 2.1 follows from the general form of the Lebesgue convergence theorem given in Royden (1968, Sect. 4.4, Thm. 16).

We also have to check Assumption (b) of Theorem 2.1 with $H_{i,n}$ in that expression substituted by $n_i^{\frac{1}{2}} H_i$. This process is bounded above by $n_i^{\frac{1}{2}}/K_i$ and hence it is enough to prove the following for every $\varepsilon > 0$:

$$E[N^* n_i K_i^{-2} I(n_i^{\frac{1}{2}} K_i^{-1} > \varepsilon)] \to 0 \qquad i=1,2,$$

where $N^* = N_1(1) + N_2(1)$. By Hölder's inequality it is enough to prove:

$$E\left(\frac{N^*}{n_i}\right)^3 E\left(\frac{n_i}{K_i}\right)^6 P(n_i^{\frac{1}{2}} K_i^{-1} > \varepsilon) \to 0 \qquad i=1,2.$$

The first factor on the left hand side is bounded when $n_1$ and $n_2$ increases since $N^*$ is dominated by a Poisson random variable. The second factor can be shown to be bounded by a straightforward extension of Lemma 4.2, part (i) of Aalen (1976). The third factor converges to 0. □

## 4. Proof of the Main Result

Theorem 2.1 will be proved by means of Theorem A.2 of the Appendix:

Let $\{v_n\}$ be an increasing sequence of nonnegative integers such that $v_n \to \infty$. Define for all $i$ and $n$:

$$v_n(t) = j \quad \text{if} \quad \frac{j-1}{v_n} < t \leqq \frac{j}{v_n}, \qquad j=1,\ldots,v_n,$$

$$Z_{n,j}^{(i)} = Y_{i,n}\left(\frac{j}{v_n}\right) - Y_{i,n}\left(\frac{j-1}{v_n}\right),$$

$$U_n^{(i)}(t) = \sum_{j=1}^{v_n(t)} Z_{n,j}^{(i)}, \qquad \mathbf{U}_n = (U_n^{(1)}, \ldots, U_n^{(k)}).$$

Using the terminology of the appendix, we see that $\{Z_{n,j}^{(i)}\}$, $i=1,\dots,k$, are $k$ orthogonal martingale difference arrays. By Theorem A.2 we then just have to verify that, for some sequence $\{v_n\}$, conditions (i) and (ii) of Theorem A.1 hold for each array with $f=g_i$ for array number $i$. When we have verified this we can conclude

$$\mathbf{U}_n \Rightarrow \mathbf{Y}$$

for this sequence $\{v_n\}$. Now it follows from the definition of $\mathbf{U}_n$ that

$$|U_n^{(i)}(t) - Y_{i,n}(t)| \leqq \sup_j Z_{n,j}^{(i)}.$$

Hence, to get the conclusion of the theorem, we also have to verify:

$$\sup_j Z_{n,j}^{(i)} \xrightarrow{P} 0, \qquad i=1,\dots,k.$$

However, by part (a) of the proof of Corollary 3.8 in McLeish (1974), the last condition follows from what corresponds to condition (i) of Theorem A.1.

We now fix $i$ and suppress it from the notation in the rest of this section. We also suppress $n$ from the notation. It should however be kept in mind that almost all quantities occurring below, including the process $Z$, depend on $n$, and that all limits are taken with respect to $n$. $E_j$ is defined in the appendix.

The existence of a sequence $\{v_n\}$ such that conditions (i) and (ii) of Theorem A.1 holds will now be proved in the form of two lemmas.

**Lemma 4.1.** *There exists a sequence* $\{v_n\}$ *such that*

$$A = \sum_{j=1}^{v} E_{j-1}[Z_j^2 I(|Z_j| > \varepsilon)] \xrightarrow{P} 0$$

*for every* $\varepsilon > 0$.

*Proof.* Let $T_j^* > (j-1)/v$ be the time of the first jump of the process $N_i$ after time $(j-1)/v$. Put $T_j = \min(T_j^*, j/v)$. From Section 1 we know that $T_j$ is a stopping time. We have:

$$Z_j = \int_{(j-1)/v}^{T_j} H\,dM + \int_{T_j}^{j/v} H\,dM.$$

Hence we can rewrite $Z_j$ as

$$Z_j = V_{1,j} + V_{2,j} + V_{3,j}$$

where

$$V_{1,j} = -\int_{(j-1)/v}^{T_j} H\Lambda\,ds,$$

$$V_{2,j} = H(T_j) I(T_j < j/v),$$

$$V_{3,j} = \int_{T_j}^{j/v} H\,dM.$$

By Lemma 1 of the Appendix it is enough to prove:

(i) There exists a sequence $\{v_n\}$ such that

$$\sum_{j=1}^{v} E_{j-1}[V_{ij}^2 I(|V_{ij}|>\varepsilon)] \xrightarrow{p} 0, \qquad i=1,2,3$$

for every $\varepsilon>0$.

For $i=2$ (i) is a consequence of Assumption (b). We will now prove (i) for $i=1$. Clearly, it is enough to prove $A_1 \xrightarrow{p} 0$ where

$$A_1 = \sum_{j=1}^{v} E_{j-1} V_{1,j}^2.$$

Since $A_1$ is nonnegative it follows from a version of Čebyšev's inequality that it is enough to prove that $EA_1 \to 0$. We have:

$$EA_1 \leqq \sum_{j=1}^{v} E[(\int_{D_j} H\Lambda\, ds)^2]$$

where $D_j=((j-1)/v, j/v)$. Hölder's inequality and Requirement B gives us:

$$(\int_{D_j} H\Lambda\, ds)^2 = (\int_{D_j} (H\Lambda^{\frac{1}{2}})(\Lambda^{\frac{1}{2}})\, ds)^2$$

$$\leqq \int_{D_j} H^2\Lambda\, ds \int_{D_j} \Lambda\, ds \leqq \frac{1}{v}\, \Phi(1) \int_{D_j} \Lambda\, ds.$$

Hence

$$EA_1 \leqq E \sum_{j=1}^{v} (\int_{D_j} H\Lambda\, ds)^2 \leqq \frac{1}{v}\, E\left(\Phi(1) \int_0^1 \Lambda\, ds\right).$$

By part (iv) of Requirement B it follows that a sequence $\{v_n\}$ can be chosen so that $EA_1 \to 0$.

Hence it remains to prove (i) for $i=3$. It is enough to prove that $A_2 \xrightarrow{p} 0$ where

$$A_2 = \Sigma E_{j-1} \left(\int_{T_j}^{j/v} H\, dM\right)^2.$$

By (1.10) we have:

$$A_2 = \Sigma E_{j-1} \int_{T_j}^{j/v} H^2 \Lambda\, ds$$

which further gives us:

$$A_2 \leqq \Sigma E_{j-1} \left[ I(N(D_j) \geqq 1) \int_{D_j} H^2 \Lambda \right]$$

where $N(D_j)=N(j/v)-N((j-1)/v)$.

By Requirement B we have:

$$A_2 \leqq \frac{1}{v} \Sigma E_{j-1}(\Phi(1) I(N(D_j) \geqq 1)].$$

Hence

$$EA_2 \leqq \frac{1}{v} \Sigma E[\Phi(1) I(N(D_j) \geqq 1)]$$

$$= \frac{1}{v} E[\Phi(1) \Sigma I(N(D_j) \geqq 1)]$$

$$\leqq \frac{1}{v} E[\Phi(1) N(1)].$$

From part (iii) of Requirement B it follows that a sequence $\{v_n\}$ can be chosen so that $EA_2 \to 0$, and hence $A_2 \xrightarrow{p} 0$. □

We have checked condition (i) of Theorem A.1 and will now check condition (ii).

**Lemma 4.2.** *There exists a sequence* $\{v_n\}$ *such that for each* $t \in [0, 1]$

$$B(t) = \sum_{j=1}^{v(t)} E_{j-1}(Z_j^2) \xrightarrow{p} \int_0^t g^2(s)\, ds.$$

*Proof.* By (1.10) we have

$$B(t) = \sum_{j=1}^{v(t)} E_{j-1} \int_{D_j} H^2 \Lambda\, ds$$

$$= \sum_{j=1}^{v(t)} \int_{D_j} H^2 \Lambda\, ds + \sum_{j=1}^{v(t)} \int_{D_j} [E_{j-1}(H^2 \Lambda) - H^2 \Lambda]\, ds.$$

We denote the two parts of the last expression as $B_1(t)$ and $B_2(t)$ respectively. We have

$$B_1(t) = \int_0^t H^2 \Lambda\, ds + \int_t^{v(t)/v} H^2 \Lambda\, ds.$$

Denote the two parts of this expression as $B_3(t)$ and $B_4(t)$ respectively. By (a) we have

$$B_3(t) \xrightarrow{p} \int_0^t g^2\, ds.$$

Let $\delta > 0$ be given. When $n$ is large enough we have:

$$B_4(t) \leqq \int_t^{t+\delta} H^2 \Lambda\, ds.$$

The right side converges in probability to $\int_t^{t+\delta} g^2\,ds$ which goes to 0 when $\delta\to 0$. Hence

$$B_4(t) \xrightarrow{p} 0.$$

Hence we only need to show that $B_2(t)\xrightarrow{p} 0$. We have:

$$|B_2(t)| \leqq \sum_{j=1}^{\nu(t)} \int_{D_j} E_{j-1}\left|H^2(s)\Lambda(s)-H^2\left(\frac{j-1}{\nu}\right)\Lambda\left(\frac{j-1}{\nu}\right)\right| + \sum_{j=1}^{\nu(t)} \int_{D_j}\left|H^2(s)\Lambda(s)-H^2\left(\frac{j-1}{\nu}\right)\Lambda\left(\frac{j-1}{\nu}\right)\right|.$$

It is enough to show that $E|B_2(t)|\to 0$. By Requirement B we have:

$$\begin{aligned} E\,|B_2(t)| &\leqq 2\sum_{j=1}^{\nu(t)} \int_{D_j} E\left|H^2(s)\Lambda(s)-H^2\left(\frac{j-1}{\nu}\right)\Lambda\left(\frac{j-1}{\nu}\right)\right| \\ &\leqq \frac{2}{\nu} E \sum_{j=1}^{\nu(t)}\left(\Phi\left(\frac{j}{\nu}\right)-\Phi\left(\frac{j-1}{\nu}\right)\right) \\ &\leqq \frac{2}{\nu} E(\Phi(1)-\Phi(0)). \end{aligned}$$

Hence a sequence $\{\nu_n\}$ can be chosen so that $E\,|B_2(t)|\to 0$. □

## Appendix

In this section we give a multivariate generalization of an invariance principle due to McLeish (1974, Corollary 3.8) and based on a central limit theorem due to Dvoretsky (1972, Theorem 2.2).

Let $\{X_{n,i};\ i=1,\dots,\nu_n;\ n=1,2,\dots\}$ be a triangular array of random variables defined on a probability space $(\Omega,\mathscr{F},P)$. Let $\{\mathscr{F}_{n,i};\ i=1,\dots,\nu_n;\ n=1,2,\dots\}$ be a triangular array of sub-$\sigma$-algebras of $\mathscr{F}$ such that $X_{n,i}$ is $\mathscr{F}_{n,i}$ measurable and $\mathscr{F}_{n,i-1}\subset\mathscr{F}_{n,i}$. We will denote $E(U\,|\,\mathscr{F}_{n,i})=E_{n,i}\,U$. We call $\{X_{n,i};\ i=1,\dots,\nu_n;\ n=1,2,\dots\}$ a martingale difference array if $E_{n,i-1}X_{n,i}=0$ almost surely for all $n$ and $i$.

Let $\nu_n(t)$ be integer-valued, non-decreasing, right-continuous functions defined on $[0,1]$, and such that $\nu_n(0)=0$ and $\nu_n(1)=\nu_n$. Define

$$U_n(t)=\sum_{i=1}^{\nu_n(t)} X_{n,i}.$$

Let $W$ be the Wiener process. The stochastic integral $\int g\,dW$ is well defined whenever $\int_0^1 g^2(s)\,ds<\infty$, i.e. whenever $g\in L^2(0,1)$.

Let $\xrightarrow{p}$ denote convergence in probability. Let $\Rightarrow$ denote weak convergence for random elements of $D[0,1]$ with respect to the Skorohod topology (see Billingsley (1968), Chapter III). The following theorem is a consequence of Corollary 3.8 of McLeish (1974):

**Theorem A.1.** *Make the following assumptions:*

(i) $\sum_{i=1}^{\nu_n} E_{n,i-1} X_{n,i}^2 I(|X_{n,i}|>\varepsilon) \xrightarrow{P} 0$ *for all* $\varepsilon>0$.

(ii) *There exists a nonnegative function* $f \in L^2(0,1)$ *such that*

$$\sum_{i=1}^{\nu_n(t)} E_{n,i-1} X_{n,i}^2 \xrightarrow{P} \int_0^t f^2(s)\,ds$$

*for all* $t \in [0,1]$.
*Then*

$$U_n \Rightarrow \int f\,dW.$$

We will now prove a multivariate generalization of this theorem. Let $\{X_{n,i}^{(j)};\ i=1,\dots,\nu_n;\ n=1,2,\dots\}$, $j=1,2,\dots,k$ be $k$ martingale difference arrays with respect to the array of $\sigma$-algebras $\{\mathscr{F}_{n,i}\}$. We assume

$$E_{n,i-1}(X_{n,i}^{(j)} X_{n,i}^{(l)})=0 \qquad \text{whenever } j \neq l.$$

In accordance with the terminology of Section 1, we say that the martingale difference arrays are *orthogonal.*

Define

$$U_n^{(j)}(t)=\sum_{i=1}^{\nu_n(t)} X_{n,i}^{(j)},$$

$$\mathbf{U}_n=(U_n^{(1)},\dots,U_n^{(k)}).$$

Let $W_1,\dots,W_k$ be independent Wiener processes.

In the following theorem $\Rightarrow$ denotes weak convergence of random elements in the cartesian product of $D[0,1]$ with itself $k$ times, equipped with the product topology corresponding to the Skorohod topology.

**Theorem A.2.** *Assume that each of the arrays* $\{X_{n,i}^{(j)}\}$, $j=1,\dots,k$, *satisfy assumptions* (i) *and* (ii) *of Theorem* A.1. *Denote the function* $f$ *of assumption* (ii) *corresponding to* $\{X_{n,i}^{(j)}\}$ *by* $f_j$. *Then*

$$\mathbf{U}_n \Rightarrow \left(\int f_1\,dW_1,\dots,\int f_k\,dW_k\right).$$

We need the following simple lemma:

**Lemma A.1.** *Let* $c_1, c_2, \dots, c_k$ *and* $x_1, x_2, \dots, x_k$ *be real numbers and define*

$$y=\sum_{j=1}^{k} c_j x_j.$$

*Assume that*

$$\sum_{j=1}^{k} |c_j| \leqq 1.$$

*Then for any* $\varepsilon > 0$

$$y^2 I(|y|>\varepsilon) \leqq \sum_{j=1}^{k} x_j^2 I(|x_j|>\varepsilon).$$

*Proof.* Clearly:

$$|y| \leqq \sum_{j=1}^{k} |c_j|\,|x_j| \leqq \max_{1\leqq j\leqq k} |x_j|.$$

$y^2 I(|y|>\varepsilon)$ is non-decreasing in $|y|$. Hence:

$$y^2 I(|y|>\varepsilon) \leqq \max_j |x_j|^2 I(\max_j |x_j|>\varepsilon) \leqq \sum_{j=1}^{k} x_j^2 I(|x_j|>\varepsilon). \quad \square$$

*Proof of Theorem A.2.* By Theorem A.1

$$U_n^{(j)} \Rightarrow \int f_j \, dW_j \qquad j=1,\ldots,k.$$

Since $D[0,1]$ is separable and complete in the Skorohod topology (Billingsley (1968), Section 14), each sequence $\{U_n^{(j)}\}$ is tight by Prohorovs theorem (*ibid.*, Theorem 6.2). It is easily seen that this implies the tightness of $\{\mathbf{U}_n\}$. Hence we only have to show convergence of finite-dimensional distributions. We use the Cramér-Wold method (Billingsley (1968), Theorem 7.7).

Let $c_1(t), \ldots, c_k(t)$ be integrable functions satisfying

$$\sum_{j=1}^{k} |c_j(t)| \leqq 1.$$

Define

$$V_n(t) = \sum_{j=1}^{k} \int_0^t c_j(s)\, dU_n^{(j)}(s)$$

where the integrals are Stieltjes integrals. We can write

$$V_n(t) = \sum_{i=1}^{\nu_n(t)} Y_{n,i}$$

where $\{Y_{n,i}\}$ is the martingale difference array defined by

$$Y_{n,i} = \sum_{j=1}^{k} c_j(\nu_n^{-1}(i))\, X_{n,i}^{(j)}$$

where

$$\nu_n^{-1}(i) = \inf\{t: \nu_n(t) \geqq i\}.$$

By Lemma A.1, $\{Y_{n,i}\}$ satisfies assumption (i) of Theorem A.1. Hence we must check assumption (ii) of Theorem A.1. By the assumed orthogonality we have

$$\sum_{i=1}^{\nu_n(t)} E_{n,i-1} Y_{n,i}^2 = \sum_{i=1}^{\nu_n(t)} E_{n,i-1} \left[ \sum_{j=1}^{k} c_j(\nu_n^{-1}(i)) X_{n,i}^{(j)} \right]^2$$
$$= \sum_{i=1}^{\nu_n(t)} \sum_{j=1}^{k} c_j^2(\nu_n^{-1}(i)) E_{n,i-1}(X_{n,i}^{(j)})^2.$$

Put $Z_n^{(j)}(t) = \sum_{i=1}^{\nu_n(t)} E_{n,i-1}(X_{n,i}^{(j)})^2$. Then the above expression assumes the form

$$\sum_{j=1}^{k} \int_0^t c_j^2(s)\, dZ_n^{(j)}(s).$$

The assumption on the $c$-functions implies $\sum c_j^2(t) \leqq 1$. By the assumption of the theorem

$$Z_n^{(j)}(t) \xrightarrow{P} \int_0^t f_j^2(s)\, ds.$$

Hence

$$\sum_{i=1}^{\nu_n(t)} E_{n,i-1} Y_{n,i}^2 \xrightarrow{P} \sum_{j=1}^{k} \int_0^t c_j^2(s) f_j^2(s)\, ds.$$

To apply the Cramér-Wold theorem, we can let $c_j$, $j=1,\ldots,k$, be step functions. Hence the appropriate finite-dimensional convergence follows. □

*Acknowledgement.* The main result of the present paper is a thoroughly revised version of a part of the author's Ph.D. dissertation. I am very grateful to my advisor, Professor Lucien Le Cam, who suggested that the theory of McLeish, (1974) would be useful, and with whom I had several enlightening discussions.

## References

1. Aalen, O.: Statistical inference for a family of counting processes. Ph.D. dissertation, Univ. of California, Berkeley. Reprinted by the Copenhagen University Institute of Mathematical Statistics (1975)
2. Aalen, O.: Nonparametric inference in connection with multiple decrement models. Scandinavian J. Statist. **3**, 15–27 (1976)
3. Billingsley, P.: Convergence of Probability Measures. New York: Wiley 1968
4. Boel, R., Varaiya, P., Wong, E.: Martingales on jump processes I: Representation results. SIAM J. Control, **13**, 999–1021 (1975a)
5. Boel, R., Varaiya, P., Wong, E.: Martingales on jump processes II: Applications. SIAM J. Control, **13**, 1022–1061 (1975b)
6. Brémaud, P.: A martingale approach to point processes. Memorandum ERL-M345, Electronics Research Laboratory, University of California, Berkeley, (1972)
7. Brémaud, P.: 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**, 519–542. Berlin-Heidelberg-New York: Springer 1974

8. Doléans-Dadé, C., Meyer, P.A.: Intégrales stochastiques par rapport aux martingales locales. Seminaire de probabilités IV. Lecture Notes in Math. **124**, 77–107. Berlin-Heidelberg-New York: Springer 1970
9. Dolivo, F.-B.: Counting processes and integrated conditional rates: A martingale approach with application to detection. Technical Report, College of Engineering, University of Michigan. (1974)
10. Dvoretsky, A.: Central limit theorems for dependent random variables. Proc. 6th Berkeley Sympos. Math. Statist. Probab. Univ. Calif. 513–535 (1972)
11. Jacod, J.: On the stochastic intensity of a random point process over the half-line. Technical Report 51, Department of Statistics, Princeton University (1973)
12. Jacod, J.: Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. Z. Wahrscheinlichkeitstheorie verw. Gebiete **31**, 235–253 (1975)
13. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. **53**, 457–481 (1958)
14. Kunita, H., Watanabe, S.: On square integrable martingales. Nagoya Math. J. **30**, 209–245 (1967)
15. McLeish, D.L.: Dependent central limit theorems and invariance principles. Ann. Probability **2**, 620–628. (1974)
16. Meyer, P.A.: Probability and Potential. Waltham. Massachussetts: Blaisdell 1966
17. Meyer, P.A.: Intégrales stochastiques, I. II, III et IV. In: Lecture Notes in Math. **39**, 77–162, Berlin-Heidelberg-New York: 1967
18. Meyer, P.A.: Square integrable martingales, a survey. Lecture Notes in Mathematics **190**, 32–37. Berlin-Heidelberg-New York: Springer 1971
19. Nelson, W.: Hazard plotting for incomplete failure data. J. Qual. Tech. **1**, 27–52. (1969)
20. Royden, H.L.: Real analysis. Second Edition. New York: Macmillan 1968
21. Segall, A., Kailath, T.: The modeling of randomly modulated jump processes. IEEE Transactions on Information Theory **IT-21** (2), 135–143 (1975a)
22. Segall, A., Kailath, T.: Radon-Nikodym derivatives with respect to measures induced by discontinuous independent-increment processes. Ann. Probability **3**, 449–464 (1975b)
23. Varayia, P.: The martingale theory of jump processes, IEEE Transactions, AC-20, **1**, (1975)

Received June 10, 1976; in revised form December 9, 1976

**Note Added in Proof.** In the first part of the proof of Proposition 3.1 we apply the general form of the Lebesgue convergence theorem (denoted *GL*) given in Royden (1968, Sect. 4.4, Thm. 16). Professor S. Johansen has pointed out that this theorem can not be applied directly in our situation. We have to introduce the following argument:

Put $X_i(t)=|n_i H_i^2(t)\Lambda_i(t)-g_i^2(t)|$. In order to verify Assumption (a) of Theorem 2.1 it suffices to prove for $i=1,2$:

$$\int_0^1 EX_i(t)\,dt\to 0$$

By the strong law of large numbers $X_i(t)\to 0$ a.s. for each $t$. We also have $X_i(t)\leqq c\,n_i/K_i+c\,p_i^{-1}$ for all $t$. The right hand side converges almost surely to $2c\,p_i^{-1}$ and its expectation converges to the same quantity (see e.g. Lemma 4.2 of Aalen (1976)). By an application of *GL* we can conclude $EX_i(t)\to 0$ for all $t$. Furthermore $EX_i(t)\leqq c\,E(n_i/K_i)+c\,p_i^{-1}$ which is bounded. Hence an application of the dominated convergence theorem gives us the conclusion.