The Annals of Probability
1983, Vol. 11, No. 3, 726–744

SOME POISSON APPROXIMATIONS USING COMPENSATORS

BY TIMOTHY C. BROWN

Monash University

We give new results on the total variation distance of the distribution of a point process on the line from that of a Poisson process. Both one dimensional and function space distances are considered. Additionally similar bounds for marked point processes are given, both for the finite and infinite mark space cases. The bounds are all given in terms of the compensator of the point process (with respect to an arbitrary filtration) and are analogues and extensions of discrete time results of Freedman (1974) and Serfling (1975). Some new techniques for discrete approximation of compensators are used in the proofs. Examples of the use of the bounds appear elsewhere (Brown and Pollett, 1982, Brown, 1981), but an application to compound Poisson approximation and thinning of point processes is given here.

1. Introduction, notation and statement of some results. There has been great interest recently in distributional limit theorems for continuous time stochastic processes which are related to martingales (for example Rebolledo, 1978, Brown, 1978, 1979, Kabanov, Liptser and Shiryayev, 1980, Jacod and Memin, 1980). The conditions in these theorems are usually in terms of the predictable characteristics of the processes (see Jacod, 1979 for general definitions). A particularly simple case is the one for which rates of convergence are studied here. Namely, a *point process* $(N, \mathscr{F})$ is formed from a *filtration* $\{\mathscr{F}(t)\}_{t\geq 0}$ (a right continuous, increasing set of σ-fields) and an unbounded sequence of stopping times

$$0 < T_1 < T_2 < \cdots$$

with

$$N(t) = \quad \text{number of} \quad T_i \leq t.$$

[There are three ways to describe the system; by the stopping times T_i, by the process $\{N(t)\}_{t\geq 0}$, or by the random measure generated by N. We shall use all three and will not distinguish between N viewed as a random measure or as a process, provided that context makes the meaning clear.] The predictable characteristic of $(N, \mathscr{F})$ is its *compensator* A which is the unique predictable process making $N - A$ a local martingale. Moreover, for any nonnegative predictable process Y

$$E\left(\int Y\,dN\right) = E\left(\int Y\,dA\right), \tag{1.1}$$

an equation which justifies many of the calculations here. The compensator depends on the filtration, but the results here allow the filtration to be larger than that generated by the process.

In Brown (1978), (1979) it is shown that point processes, whose compensators converge in distribution to a continuous deterministic function, do themselves converge in distribution to a Poisson process (more general results are in Kabanov, Liptser and Shiryayev, 1980, Brown, 1979 and Jacod and Memin, 1980). On the other hand, there is previous work on the existence of certain (discrete time) Poisson *approximation* theorems for sums of dependent Bernoulli random variables (Freedman, 1974, Serfling, 1975). We may embed these theorems in our framework by assuming that $T_1, T_2, \cdots$ only take integer values. Then

$$N(t) = \textstyle\sum_{i=1}^{[t]} X_i$$

Received May 1982.

AMS subject classifications. 60G55, 60G44, 60G07.

Key words and phrases. Point process, marks, Poisson process, compensator, discrete approximations.

where $X_1, X_2, \cdots$ take values only in $\{0, 1\}$. If $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \cdots$ are an increasing sequence of σ-fields, and $\mathscr{F}(t) = \mathscr{F}_{[t]}$, then it is easily seen that the compensator A of $(N, \mathscr{F})$ is given by

$$A(t) = \sum_{i=1}^{[t]} p_i$$

where

$$p_i = E(X_i \mid \mathscr{F}_{i-1}).$$

The approximation theorems mentioned above give bounds on the *total variation distance*, d, between the distribution of sum of the X_i and a Poisson distribution; for any two probability measures $\mathscr{P}$ and $\mathscr{Q}$ with the same domain

$$d(\mathscr{P}, \mathscr{Q}) = \sup \mid \mathscr{P}(A) - \mathscr{Q}(A) \mid$$

where the supremem is taken over all sets A in the common domain. Freedman (1974) gives the result that for any integer stopping time τ

$$d(N(\tau), \text{Poisson } (\mu)) \leq 2\delta_1 + (\nu' - \mu) + \alpha\delta_2 \tag{1.2}$$

where $\delta_1, \delta_2 > 0$ and $\nu' \geq \mu$ are such that

$$P(\mu \leq \textstyle\sum_{i=1}^{\tau} p_i \leq \nu' \quad \text{and} \quad \sum_{i=1}^{\tau} p_i^2 \leq \delta_2) \geq 1 - \delta_1$$

and $\alpha > 1$ is a function of δ_2. [Here and below we write $d(X, \mathscr{Q})$, for a random element X and probability $\mathscr{Q}$, to mean the total variation distance between the distribution of X and $\mathscr{Q}$.] Serfling (1975) gives the result that for any fixed integer n,

$$d(N(n), \text{Poisson } (\mu)) \leq \textstyle\sum_{i=1}^{n} E\,|\,p_i - \mu_i\,| + \sum_{i=1}^{n} \{E(p_i)\}^2 \tag{1.3}$$

where $\mu_1, \cdots, \mu_n$ are constants adding to μ. It is a simple matter to see from (1.2) what the continuous time analogue should be, since

$$\textstyle\sum_{i=1}^{\tau} p_i = A(\tau) \tag{1.4}$$

and

$$\textstyle\sum_{i=1}^{\tau} p_i^2 = \sum_{s \leq \tau} \Delta A^2(s) \tag{1.5}$$

where $\Delta A(s) = A(s) - A(s-)$. While this analogue is not so clear in (1.3), it has the advantage of reducing to the well-known bound np^2 when $N(n)$ is Binomial (n, p), whereas (1.2) in the same case is $\alpha np^2 > np^2$. Another difference is that (1.2) is in terms of probabilities, while (1.3) concerns expectations; although the former has theoretical advantages, it seems that the latter is easier to use in practice (see, for example, Brown and Silverman, 1979).

In the present paper, we produce bounds for the general continuous time point process departure from Poissonity. Moreover, in the discrete time case, one bound here improves that of Freedman (1974), while another is similar to, but different from, that of Serfling (1975). This latter here is also in terms of the quantities in (1.4) and (1.5), thus providing a unified theory.

Using the same methods, we can extend the bounds mentioned above in a number of directions. Firstly, we may consider a *marked point process* $(\mathbf{N}, \mathscr{F})$ where we have a second sequence

$$Z_1, Z_2, \cdots$$

of random elements on some measurable space, $(E, \mathscr{E})$; here Z_i is to be interpreted as a mark attached to T_i. Then, $\mathbf{N}$ becomes a random measure on $\mathbb{R}^+ \times E$ which is defined, for $t \geq 0$ and F in $\mathscr{E}$, by

$$\mathbf{N}((0, t] \times F) = \quad \text{number of} \quad T_i \leq t \quad \text{such that} \quad Z_i \in F.$$

If $E = \{1, 2, \cdots, k\}$ we call $(\mathbf{N}, \mathscr{F})$ a *k-type point process* and $\mathbf{N}$ is equivalent to the vector

$(N_1, N_2, \cdots, N_k)$ where

$$N_j(t) = N((0, t] \times \{j\}).$$

In this case, the compensator of $(\mathbf{N}, \mathscr{F})$ is the vector $\mathbf{A} = (A_1, A_2, \cdots, A_k)$ of compensators of $(N_1, \mathscr{F}), (N_2, \mathscr{F}), \cdots, (N_k, \mathscr{F})$. We can now state the first main result.

THEOREM 1. *If $(N, \mathscr{F})$ is a point process with compensator A, $\mu \geq 0$ and τ is any stopping time, then*

(a) $$d(N(\tau), \text{Poisson } (\mu)) \leq \delta_1 + \delta_2 + \delta_3 + (\nu' - \nu)$$

provided $\nu \leq \mu \leq \nu'$, $\delta_1, \delta_2, \delta_3$ (≥ 0) are constants such that

$$P(\nu \leq A(\tau) \leq \nu') \geq 1 - \delta_1$$

and

$$P(\textstyle\sum \Delta A^2(s) \leq \delta_2) \geq 1 - \delta_3.$$

Further,

(b) $$d(N(\tau), \text{Poisson } (\mu)) \leq E\,|A(\tau) - \mu| + E\{\textstyle\sum_{s\leq\tau} \Delta A^2(s)\}.$$

(c) *If $(\mathbf{N}, \mathscr{F})$ is a k-type point process with compensator $\mathbf{A} = (A_1, A_2, \cdots, A_k)$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_k) \geq 0$ and τ is a stopping time, then*

$$d(\mathbf{N}(\tau), \text{Poisson } (\boldsymbol{\mu})) \leq \textstyle\sum_{i=1}^{k} E\,|A_i(\tau) - \mu_i| + E\{\sum_{s\leq\tau} \Delta A^2(s)\}$$

where here, and from now on, Poisson$(\cdot)$ with a vector argument means a vector of independent *Poisson distributions and*

$$A = \textstyle\sum_{i=1}^{k} A_i.$$

Part (a) may also be extended to the k-type case, as will be apparent from the proof. The discrete time case of Part (a) improves the bound of (1.2) by removing the constant $\alpha > 1$, allowing δ_1 and δ_3 to differ and permitting ν to be less than μ. The comparison between the discrete time case of Part (b) and the bound of (1.3) is less clear; while the first term of (b) is, for $\tau = n$,

$$E\,|\textstyle\sum_{i=1}^{n} p_i - \mu| \leq \sum_{i=1}^{n} E\,|p_i - \mu_i|,$$

the second term is

$$E(\textstyle\sum_{i=1}^{n} p_i^2) \geq \sum_{i=1}^{n} \{E(p_i)\}^2.$$

A second extension, which actually follows from Part (c), concerns functional forms of the continuous time bounds. For a stochastic process $\{X(s)\}_{s\geq 0}$ by X^t we mean the process *stopped at* $t \geq 0$, so that

$$X^t(s) = X(s \wedge t).$$

Below, we bound the distance of the distribution of a point process stopped at t from that of a Poisson process. For this purpose, the distribution of a Poisson process on $[0, t]$ with mean measure μ will be written Poisson (μ^t), while the distribution of a vector of *independent* Poisson processes on $[0, t]$ will be written as Poisson $(\boldsymbol{\mu}^t)$ if the marginals have mean measures $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_k)$. We also need the *total variation*, $|\cdot|$, of a signed measure. This is the sum of the measures of the whole space in a Jordan decomposition and so for probabilities $\mathscr{P}$ and $\mathscr{Q}$

$$d(\mathscr{P}, \mathscr{Q}) = \tfrac{1}{2}\,|\mathscr{P} - \mathscr{Q}|.$$

If m is a signed measure on $(0, \infty)$ and $t > 0$, then $|m|_t$ will denote the total variation of the measure restricted to $(0, t]$. Thus, in the theorems below $|A - \mu|_t$ means the random variable $\omega \rightarrow |A(\cdot, \omega) - \mu|_t$, with $A(\cdot, \omega)$ regarded as a measure on $(0, \infty)$.

COROLLARY 1. *Under the conditions of Theorem* 1, *if* μ *is a measure on* $(0, \infty)$ *and* $t > 0$ *is non-random, then*

(a) $$d(N^t, \text{Poisson }(\mu^t)) \le E\,|\,A - \mu\,|_t + E\{\textstyle\sum_{s\le t} \Delta A^2(s)\},$$

and if $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_k)$ *is a vector of measures on* $\mathbb{R}^+$, *then*

(b) $$d(\mathbf{N}^t, \text{Poisson }(\boldsymbol{\mu}^t)) \le \textstyle\sum_{i=1}^n E\,|\,A_i - \mu_i\,|_t + E\{\sum_{s\le t} \Delta A^2(s)\}.$$

We remark that often the bounds in both Theorem 1 and Corollary 1 are linear in t, so that Corollary 1 is in one sense a much more informative result; indeed the second terms in the bounds coincide, and in the case that A is differentiable the first term in Corollary 1 is usually easier to calculate. In this particular case, Corollary 1 has been applied to the analysis of Markov queueing networks (Brown and Pollett, 1981). Theorem 1 has been applied to sums of exchangeable (Brown, 1981) random variables. In this case it often yields smaller bounds that those of (1.3).

There are three main techniques used in the proofs. As in Freedman (1974) and Serfling (1975), *coupling* of random variables and processes plays a vital role. A pair of random elements (X, Y) is a *coupling* for a pair of distributions $\mathscr{P}$ and $\mathscr{Q}$, if the marginals of (X, Y) are $\mathscr{P}$ and $\mathscr{Q}$. It is easy to see that in this case

(1.6) $$d(\mathscr{P}, \mathscr{Q}) \le P(X \ne Y)$$

(Freedman, 1974). Secondly, stochastic calculus is used to compute expectations and probabilities and to construct couplings. The necessary result for the constructions is the *random time change* theorem (Papangelou, 1972, Liptser and Shiryayev, 1974, Theorem 18.10); if $(N, \mathscr{F})$ is a point process with continuous compensator A such that $A(\infty) = \infty$, then $N \circ \hat{A}$ is a Poisson process if

$$\hat{A}(t) = \inf\{z : A(z) \ge t\}.$$

Finally, we use discrete skeletons of continuous time processes. This is done because the couplings in the proof of Theorem 1 are difficult to construct in continuous time. I suspect that they can be made using the distributional theory of Knight's prediction process due to Aldous (1979); however, a number of relatively deep auxiliary results would be needed and, by contrast, the discrete approximation theorems are straightforward.

The organisation of this paper follows. In the next section, we show that arbitrarily good discrete L_1-approximations to compensators and their processes may be obtained, provided that one forms partitions from predictable stopping times, rather than fixed times. In Section 3, Theorem 1 is proved. In Section 4, we prove Corollary 1 and a number of other Corollaries. These cover the case where conditional intensities exist, the generalisation to approximating an *arbitrary* (right-continuous) increasing process by a Poisson process, a new convergence result, and approximations by compound Poisson distributions. Finally, in Section 5, we consider bounds for marked point processes whose mark space is a suitable topological space.

Since writing the first draft of this work, several connected papers have been drawn to my attention. Memin (1982) and Kabanov and Liptser (1982) give total variation results for the distance between two point processes in terms of their *intrinsic* compensators (i.e. the compensators of the processes together with their minimal filtrations). Kabanov and Liptser (1982) also treat the marked case and show that convergence in probability of the variations of the intrinsic compensators is equivalent to variation convergence of the distributions of the processes. While their results apply to non-Poisson approximations and may give a better result in the Poisson case, we note that many applications involve larger than minimal filtrations. Valkeila (1982) gives a result for approximation by a point process with independent increments; here the bound involves the Doléans-Dade exponential of the compensators.

2. Discrete approximations to compensators. In Brown (1978) discrete approximations are used to obtain Poisson convergence results for continuous time processes. A problem with this method is that although compensators can always be obtained as limits of discrete approximations along the dyadic rationals, the topology used is that of weak L_1 convergence. Dellacherie and Doléans-Dade (1970) provide a counterexample to the conjecture that (strong) L_1 convergence can be used in general. However, provided one is willing to allow an approximation via a random time set, then this problem disappears (Proposition 1 below). The problem arises only because of jumps in the compensator, and it is overcome by inserting the times of large jumps of the compensators as singleton members of the partition which defines the approximand. Because the resulting partition has predictability properties, it seems that for many purposes the partition is as good as a non-random one. Perhaps other martingale-type weak limit results may be derived via this reduction to the discrete case.

To be precise, we define an *($\mathscr{F}-$) predictable partition* $\mathbf{Q}$ *of* $[0, \infty]$ as a mapping from Ω to the collection of all sequences of intervals of $[0, \infty]$ (where singletons count as closed intervals) with the following properties:

(i) if $\mathbf{Q}(\omega) = (Q_1(\omega), Q_2(\omega), \cdots)$, then $\cup\, Q_i(\omega) = [0, \infty]$ and the $Q_i(\omega)$ are disjoint,
(ii) for each i, there exist *($\mathscr{F}-$)predictable* stopping times L and R so that Q_i is one of (L, R), $(L, R]$, $[L, R]$, $[L, R)$ or $\{L\}$.

With each predictable partition $\mathbf{Q}$ we associate an *increasing sequence of σ-fields* $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots$ as follows:
(i) if Q_i is open on the left, then

$$\mathscr{F}_{i-1} = \mathscr{F}(L),$$

(ii) if Q_i is closed on the left, then

$$\mathscr{F}_{i-1} = \mathscr{F}(L-).$$

These σ-fields depend on $\mathbf{Q}$, but we will not make this explicit from now on and hope that it is clear from the context which partition defines a particular sequence of σ-fields. Given a predictable partition $\mathbf{Q} = (Q_1, Q_2, \cdots)$ we define the *$\mathbf{Q}$-discrete approximation to the compensator of a supermartingale X at the time point ∞ as*

$$a(\mathbf{Q}) = \textstyle\sum_{i\geq 1} E(-X(Q_i) \mid \mathscr{F}_{i-1})$$

where $X(Q_i)$ is $X(R-) - X(L)$, $X(R) - X(L)$, $X(R) - X(L-)$, $X(R-) - X(L-)$, $\Delta X(L)$ (respectively) in the five possible cases for Q_i of (ii) above.

The proof of Proposition 1 is just an enlargement of the remark at the end of Murali-Rao (1969) concerning L_1 convergence of discrete approximations using dyadic rational partitions. Perhaps the reason that the Proposition appears to be new is that much martingale work in continuous time eschews reduction to the discrete. However, some unattractively heavy work would be necessary to prove Theorem 1 purely in continuous time. Indeed, the author believes that one of the attractive features of the proof of Theorem 1 is the interplay between discrete and continuous time.

As a final preliminary, we note that it is easily verified that the predictable partitions of $[0, \infty]$ form a directed set under pointwise set inclusion.

We state a result for Class D potentials first as it may be of general interest. Clearly, the same technique may be used for the increasing processes in the canonical decomposition of semimartingales, but we adhere to the Class D potentials, for ease of exposition. We will use the Proposition later only in its Corollary.

PROPOSITION 1. *Let X be a potential of class D with compensator A. Then*

$$a(\mathbf{Q}) \to_{L_1} A(\infty)$$

along the directed set of predictable partitions $\mathbf{Q}$ *of* $[0, \infty]$. *Explicitly, given* $\lambda, \mu > 0$ *we can choose a predictable partition* $\mathbf{Q}_0$ *so that for any partition,* $\mathbf{Q}$, *finer than* $\mathbf{Q}_0$

$$E\,|\,a(Q) - A(\infty)\,| \leq (\lambda\mu)^{1/2} + 2E\{A(\infty) : A(\infty) \geq \lambda\}. \tag{2.1}$$

PROOF. We prove (2.1) and then the convergence follows. For, given $\delta > 0$, the integrability of $A(\infty)$ implies that λ may be chosen large enough to guarantee that the second term of (2.1) is bounded by $\delta/2$ and then the choice of μ so that

$$\sqrt{\mu} < \delta/(2\sqrt{\lambda})$$

ensures that the bound is smaller than δ.

Let S be the predictable time given by

$$S = \inf\{z > 0 : A(z) \geq \lambda\}.$$

The process $XI[0, S)$ has compensator $A' = AI[0, S)$ and clearly the L_1 norms of both $A(\infty) - A'(\infty)$ and $a(Q) - a'(Q)$ are dominated by $E\{A(\infty) : A(\infty) \geq \lambda\}$. Thus the second term of (2.1) is the penalty for assuming that A is bounded by λ, an assumption which we make henceforward.

Let $S_0 = T_0 = 0$ and for $i \geq 1$, let

$$T_i = \inf\{z > T_{i-1} : \Delta A(z) > \mu/2\}$$

$$S_i = \inf\{z > S_{i-1} : A(z) - A(S_{i-1}) > \mu/2\}.$$

The predictable partition, $\mathbf{Q}_0$, is formed from $Q_1, Q_2, \cdots$ which are defined inductively as follows. Let $Q_1(\omega) = \phi$ and $Q_2(\omega) = \{0\}$. For the inductive step, suppose that $Q_{2i}(\omega) = \{L(\omega)\}$ and that $R(\omega) = \min\{S_j(\omega), T_j(\omega) : j \geq 1, S_j(\omega), T_j(\omega) \geq L(\omega)\}$.

Assuming that L is predictable, then so is R. We let $Q_{2i+1}(w) = (L(w), R(w))$and $Q_{2i+2} = \{R(\omega)\}$. Since, for each ω, only a finite number of $S_1(\omega), T_1(\omega), \cdots$ are finite, the set $\{Q_1(\omega), Q_2(\omega), \cdots\}$ is a partition of $[0, \infty]$.

To see that $\mathbf{Q}_0$ has the desired properties, let

$$a_i = E(-X(Q_i) \mid \mathscr{F}_{i-1}).$$

Then, if i is odd,

$$a_i = E(A(Q_i) \mid \mathscr{F}_{i-1})$$

and, if i is even, since $\Delta A(L)$ is $\mathscr{F}(L-)$ measurable for a stopping time L,

$$a_i = A(Q_i).$$

Since also $A(\infty) \leq \lambda$, $\{A(Q_i) - a_i\}_i$ is the difference sequence of a uniformly square integrable martingale and thus

$$E(A(\infty) - a(Q))^2 = \textstyle\sum_{i=1}^{\infty} E(A(Q_i) - a_i)^2.$$

By the previous sentence the right side equals

$$\textstyle\sum_{i=0}^{\infty} E(A(Q_{2i+1}) - a_{2i+1})^2 \leq \sum_{i=0}^{\infty} E(A(Q_{2i+1})^2) \leq E\{\lambda \max_i A(Q_{2i+1})\}$$

and the latter is bounded by $\lambda\mu$, by the construction. The fact that standard deviation exceeds mean absolute deviation now completes the proof for $\mathbf{Q} = \mathbf{Q}_0$ and the proof for general $\mathbf{Q}$ is similar but notationally more complex.

COROLLARY 2. *Given a bounded point process $(N, \mathscr{F})$ with compensator A and given a stopping time τ and a number $\varepsilon > 0$, there exists a sequence of Bernoulli random variables $X_1, X_2, \cdots$ and a sequence of σ-fields $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots$ to which $\{X_i\}$ is adapted and such that*

$$E|A(\tau) - \textstyle\sum_{i\geq 1} p_i| \leq \varepsilon \tag{2.2}$$

$$E|\textstyle\sum_{s\leq\tau} \Delta A^2(s) - \sum_{i\geq 1} p_i^2| \leq \varepsilon \tag{2.3}$$

and

$$P(\textstyle\sum_{i\geq 1} X_i \neq N(\tau)) \leq \varepsilon \tag{2.4}$$

where

$$p_i = E(X_i \mid \mathscr{F}_{i-1}). \tag{2.5}$$

PROOF. Let $M \geq 1$ be an upper bound for N and let $\delta < \varepsilon/\{4M^2\}$. Let $X(t) = E(N(\tau) \mid \mathscr{F}(t)) - N^\tau(t)$, so that $(X, \mathscr{F})$ is a potential of class D. Choose μ and λ so that $(\lambda\mu)^{1/2} + 2E(A(\tau): A(\tau) \geq \lambda) = \delta$. If $\mathbf{Q}_0$ is chosen as in the last proof, then for any partition $\mathbf{Q}$ finer than $\mathbf{Q}_0$

$$\| a(\mathbf{Q}) - A(\tau) \|_1 \leq \delta \tag{2.6}$$

and unless Q_i is a singleton

$$a_i = E(N^\tau(Q_i) \mid \mathscr{F}_{i-1}) < \delta.$$

Because N^τ is a bounded point process, there is a *non-random* partition $\mathbf{R}$ of $[0, \infty]$ so that if $\mathbf{R} = \{R_1, R_2, \cdots, R_n\}$ then each R_i is an interval and

$$P(\cup_{i=1}^n [N^\tau(R_i) > 1]) < \delta. \tag{2.7}$$

Let $\mathbf{Q}$ be a partition which is finer than both $\mathbf{Q}_0$ and $\mathbf{R}$ and which preserves the property of $\mathbf{Q}_0$ that its evenly indexed sets are singletons while its odd indexed sets are open intervals. Because $\mathbf{R}$ is non-random, we can also make $\mathbf{Q}$ predictable. Since $\mathbf{Q}$ is finer than $\mathbf{R}$, equation (2.7) remains valid with Q_i replacing R_i and ∞ replacing n. Thus, if $Y_i = N^\tau(Q_i)$ and

$$X_i = I[Y_i \geq 1]$$

then (2.4) is satisfied, by (2.7). Further,

$$\begin{aligned}\| \textstyle\sum_{i\geq 1} p_i - a(\mathbf{Q}) \|_1 &= E\{\textstyle\sum_{i\geq 1} E(Y_i : Y_i > 1 \mid \mathscr{F}_{i-1})\} \\ &\leq MP(\cup_{i=1}^\infty Y_i > 1) \leq \varepsilon/4M.\end{aligned}$$

Combining this with (2.6), we obtain (2.2).

Furthermore,

$$\begin{aligned}\| \textstyle\sum_{i\geq 1} p_i^2 - \sum_{i\geq 1} \{E(Y_i \mid \mathscr{F}_{i-1})\}^2 \|_1 &= E\{\textstyle\sum_{i\geq 1} (p_i - E(Y_i \mid \mathscr{F}_{i-1}))(p_i + E(Y_i \mid \mathscr{F}_{i-1}))\} \\ &\leq 2ME\{\textstyle\sum_{i\geq 1} E(Y_i : Y_i > 1 \mid \mathscr{F}_{i-1})\} \\ &\leq \varepsilon/2.\end{aligned}$$

But

$$\begin{aligned}\| \textstyle\sum_{i\geq 1} \{E(Y_i \mid \mathscr{F}_{i-1})\}^2 - \sum_{s\leq t} \Delta A^2(s) \|_1 &= \| \textstyle\sum_{i\geq 1} \{E(A(Q_{2i+1} \mid \mathscr{F}_{2i}))\}^2 - \sum_{\Delta A(s)\leq \mu} \Delta A^2(s) \|_1 \\ &< \delta E\{\textstyle\sum_{i\geq 1} E(A(Q_i \mid \mathscr{F}_{i-1}))\} + \delta E\{\textstyle\sum_{s\leq t} \Delta A(s)\} \\ &< 2\delta E\{A(\infty)\} < \varepsilon/2.\end{aligned}$$

Combining the last two series of inequalities, we obtain (2.3), and the proof is complete.

The next Proposition is a generalization of Proposition 1 to the vector case.

PROPOSITION 2. *Let $((X_1, \cdots, X_k), \mathscr{F})$ be a vector of potentials of class D with compensators $(A_1, \cdots, A_k)$. Then, given $\lambda, \mu > 0$, we can choose a predictable partition $\mathbf{Q}_0$ so that for any partition $\mathbf{Q}$ finer than $\mathbf{Q}_0$ and each j*

$$\| a_j(Q) - A_j(\infty) \|_1 \leq (\lambda\mu)^{1/2} + 2E\{A(t) : A(t) \geq \lambda\}$$

where $A = \sum_{j=1}^k A_j$.

PROOF. This is essentially the same as the proof of Proposition 1. The only difference is that we define

$$a_{ji} = E(X_j(Q_i) \mid \mathscr{F}_{i-1})$$

and, by a similar argument to that of Proposition 1,

$$E(A_j(\infty) - a_j(\mathbf{Q}_0))^2 \le E\{\lambda \max_i A_j(Q_{2i+1})\}. \tag{2.8}$$

But, since

$$A_j(Q_i) \le A(Q_i),$$

the right side of (2.8) is still bounded by $\lambda\mu$.

COROLLARY 3. *Let $((N_1, \cdots, N_k), \mathscr{F})$ be a bounded k-type point process with compensators $(A_1, \cdots, A_k)$. Given a stopping time τ and a number $\varepsilon > 0$, there exists a sequence of Bernoulli random vectors $\mathbf{X}_1 = (X_{11}, X_{21}, \cdots, X_{k1})$, $\mathbf{X}_2 = (X_{12}, X_{22}, \cdots, X_{k2})$, $\cdots$ and a sequence of σ-fields $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots$ to which $\{\mathbf{X}_i\}$ is adapted and such that for each j (2.2), (2.3), (2.4) and (2.5) hold with A_j, p_{ji}, X_{ji}, N_j replacing A, p_i, x_i, N (respectively).*

PROOF. This is essentially the same as that of Corollary 2 after we make the definitions

$$N = \sum_{j=1}^k N_j \quad \text{and} \quad A = \sum_{j=1}^k A_j,$$

so that $(N, \mathscr{F})$ is a point process with compensator A. Further, let

$$Y_{ji} = N_j^\tau(Q_i), \quad Y_i = \sum_{j=1}^k Y_{ji}, \quad a_{ji} = E(Y_{ji} \mid \mathscr{F}_{i-1})$$

and

$$X_{ji} = I[Y_{ji} \ge 1] I[\textstyle\sum_{\ell \ne j} Y_{\ell i} = 0].$$

The verification that these definitions work is left to the reader.

3. Proof of Theorem 1. We first show that Theorem 1 holds in the case where all compensators are assumed continuous. By the argument of page 309 of Brown (1981), we may also assume that each compensator has infinite limit at infinity. To prove (a) and (b), let

$$\hat{A}(z) = \inf\{s : A(s) > z\}.$$

By the random time change theorem (e.g. Liptser and Shiryayev, 1978, Theorem 18.10), $\hat{N} = N \circ \hat{A}$ is a unit rate Poisson process. Thus

$$d(N(\tau), \text{Poisson } (\mu)) \le P(N(\tau) \ne \hat{N}(\mu)). \tag{3.1}$$

For (b), we first use the fact that both $N(\tau)$ and $\hat{N}(\mu)$ take only non-negative integer values. Thus, the right side of (3.1) is dominated by

$$E|N(\tau) - \hat{N}(\mu)| = E\left\{\int Z\, dN + \int Z'\, dN\right\}$$

where $Z(z)$ and $Z'(z)$ are indicators of the events $[\hat{A}(\mu) < z \le \tau]$ and $[\tau < z \le \hat{A}(\mu)]$ (respectively). Because Z and Z' are left continuous and adapted, they are predictable. Hence, we may replace N by A in the right side of the last equation to obtain

$$d(N(\tau), \text{Poisson } (\mu)) \le E|A \circ \hat{A}(\mu) - A(\tau)|$$

which gives (a) since $A \circ \hat{A}(\mu) = \mu$.

For (a), we observe that under (a) the right side of (3.1) is dominated by

$$P(\hat{N}(\nu') - \hat{N}(\nu) \ge 1) + \delta_1$$

since $N(\tau)$ almost surely equals $\hat{N}(A(\tau))$ (Brown (1981), page 309). But the last expression is dominated by $(\nu' - \nu) + \delta_1$ by Lemma 7 of Freedman (1974).

For (c), we define $\hat{A}_j$ by replacing A by A_j in the definition of $\hat{A}$. By Theorem 2$'$ of Meyer (1970), $(\hat{N}_1, \cdots, \hat{N}_k) = (N_1 \circ \hat{A}_1, \cdots, N_k \circ \hat{A}_k)$ now consists of *independent* unit rate

Poisson processes. Thus, the required distance is bounded by

$$P(\cup_{j=1}^{k} \hat{N}_j(\mu_j) \neq N_j(\tau_j)) \leq \sum_{j=1}^{k} P(\hat{N}_j(\mu_j) \neq N_j(\tau_j))$$

which gives the desired bound after using the argument of (b).

To obtain the general bounds, we use the approximation techniques of the previous section. We shall show that the required distances are bounded by the expressions given plus fixed multiples of ε, where $\varepsilon > 0$ is arbitrary.

Firstly, we may assume $E(N(\tau)) < \infty$ in (a) and (b) and $E(N_i(\tau)) < \infty$ for all i in (c). In (b) and (c), the assumptions may be made, because in the contrary case the bounds are infinite. For (a), we define

$$\sigma = \inf\{z : A(z) > \nu'\}$$

and consider the process $N'(t)$ which takes the value $N(t)$ if $t < \sigma$ and is $N(\sigma-)$ for $t \geq \sigma$. If A' is the compensator of $(N', \mathscr{F})$, then we have

$$E(N'(\tau)) = E(A'(\tau)) = E\left(\int_{[0,\sigma)} dA\right) \leq \nu' < \infty.$$

Moreover, A' still satisfies the second hypothesis of (a) and

$$P(\nu \leq A'(\tau) \leq \nu') = P(A(\tau) \geq \nu)$$

while

$$d(N(\tau), N'(\tau)) \leq P(A(\tau) > \nu')$$

and we obtain

$$d(N(\tau), \text{Poisson } (\mu)) \leq P(A(\tau) > \nu') + (\nu' - \nu) + P(A'(\tau) < \nu) + \delta_2 + \delta_3$$

provided (a) is true for N', and this gives inequality (a) for N.

We find a jump time T_M of N (with $N = \sum_{i=1}^{k} N_i$ in (c)) such that

$$P(T_M \leq \tau) < \varepsilon.$$

By replacing $N(\tau)$ by $N(\tau \wedge T_M)$ we will change the distance by at most ε and the bounds by at most 2ε; the latter because the compensator of $(N(\cdot \wedge T_M), \mathscr{F})$ is $A(\cdot \wedge T_M)$. Since the process $N(\cdot \wedge T_M)$ is bounded by M, we may apply Corollaries 2 and 3 to it. To complete the proof of Theorem 1 it thus suffices to establish:

PROPOSITION 3. *Let $\{X_i\}$ be a sequence of Bernoulli random variables, which is adapted to an increasing sequence $\{\mathscr{F}_i\}$ of σ-fields. Let*

(3.2) $$S = \Sigma X_i$$

where here and for the rest of the Proposition the unmarked sums are over i in $\{1, 2, \cdots\}$ and let

(3.3) $$p_i = E(X_i \mid \mathscr{F}_{i-1}).$$

Then, assuming $E(S) < \infty$, (a) *and* (b) *of Theorem* 1 *hold with S replacing $N(\tau)$, Σp_i replacing $A(\tau)$ and Σp_i^2 replacing $\Sigma \Delta A^2(s)$.*

Suppose $\{\mathbf{X}_n\} = \{(X_{1n}, \cdots, X_{kn})\}$ is a sequence of Bernoulli random vectors and $\{\mathscr{F}_n\}$ is an increasing sequence of σ-algebras to which $\{\mathbf{X}_n\}$ is adapted. Let

$$\mathbf{S} = (\Sigma X_{1i}, \cdots, \Sigma X_{ki})$$

and

$$p_{ji} = E(X_{ji} \mid \mathscr{F}_{i-1}).$$

Then, assuming $E(\mathbf{S}) < \infty$, (c) of Theorem 1 holds with $\mathbf{S}$ replacing $\mathbf{N}(\tau)$, Σp_{ji} replacing $A_j(\tau)$ and $\Sigma(\sum_{j=1}^{k} p_{ji})^2$ replacing $\Sigma \Delta A^2(s)$.

PROOF. For a finite sequence of 0-1 random variables, part (a) was proved in Brown (1982). The proof described a certain construction which enabled use of the bound given here for point processes with continuous compensator. The same construction is used here, but, particularly as it has points in common with the constructions in Freedman (1974) and Serfling (1975), we refer the reader to Brown (1982) for details. Alternatively, the construction has much in common with the one given in full in Section 5.

After possibly extending the underlying probability space, there exists a sequence $\{\xi_i\}$ of 0-1 random variables and a point process $(\eta, \mathscr{G})$ with the following properties:

(i) $(\xi_i, \xi_2, \cdots)$ has the same distribution as $(X_1, X_2, \cdots)$,
(ii) $(\eta, \mathscr{G})$ has a continuous compensator α with

(iii) $$\alpha(\infty) = \Sigma q_i$$

where $(q_1, q_2, \cdots)$ has the same distribution as $(p_1, p_2, \cdots)$,

(iv) $\mathscr{G}(i) = \sigma(q_1, \xi_1, \cdots, q_i, \xi_i, q_{i+1}), i \geq 0$

and

$$\mathscr{G}(t) = \mathscr{G}[t] \vee \sigma(\eta(z), z \leq t);$$

(v) if Δ is the conditional probability that ξ_i differs from $\eta(i) - \eta(i-1)$ given $\mathscr{G}(i-1)$, then Δ is bounded by q_i^2.

Let f be a continuous, strictly increasing function from $[0, 1]$ onto $[0, \infty]$. Define $N(t) = \eta(f(t))$ and $\mathscr{F}(t) = \mathscr{G}(f(t))$ for $t \leq 1$. Because

$$E(N(1)) = E(\Sigma q_i) = E(S) < \infty,$$

the process $(N, \mathscr{F})$ does not explode on $[0, 1]$. Hence, it is a point process whose continuous compensator takes the value Σq_i at 1. Now

$$d(\Sigma X_i, \text{Poisson }(\mu)) \leq d(\Sigma X_i, N(1)) + d(N(1), \text{Poisson }(\mu)). \tag{3.4}$$

To establish part (b), we note that the argument already given for point processes with continuous compensators proves that the second term on the right of (3.4) is bounded by

$$E|\Sigma q_i - \mu| = E|\Sigma p_i - \mu|$$

from property (iii) of the construction. But the first term equals $d(\Sigma \xi_i, N(1))$ from property (i) and the latter is dominated by

$$P(\Sigma \xi_i \neq \Sigma Z_i) \leq E\{\Sigma P(\xi_i \neq Z_i \mid \mathscr{G}(i-1))\} \leq E\{\Sigma p_i^2\}$$

using properties (iv) and (iii) of the construction.

To establish (a), we define the $\{\mathscr{G}(i)\}$ stopping time

$$\sigma = \inf\{n : \textstyle\sum_1^{n+1} q_i^2 > \delta_2\}.$$

The argument already given shows that the second term of (3.4) is bounded by $(\nu' - \nu) + \delta_1$. Now

$$P(\Sigma \xi_i \neq \Sigma Z_i) \leq P(\sigma < \infty) + P(\Sigma \xi_i' \neq \Sigma z_i')$$

where $\xi_i' = Y_i I[\sigma \geq i] + Z_i' = Z_i I[\sigma \geq i]$. By property (iii) and assumption, the first term of this inequality is bounded by δ_3. After noting that $[\sigma \geq i] \in \mathscr{G}(i-1)$, property (iv) of the construction gives a bound for the second term as

$$E\{\Sigma P(\xi_i \neq Z_i \mid \mathscr{G}(i-1)) I[\sigma \geq i]\} \leq E\{\Sigma q_i^2 I[\sigma \geq i]\} \leq \delta_2$$

by the definition of σ. The proof of (b) is thus complete.

For part (c), we make a similar construction to that used in parts (a) and (b) with X_i

$= \sum_{j=1}^{k} X_{ji}$. However, we now have

$$\alpha(\infty) = \Sigma q_i$$

where $q_i = \sum_{j=1}^{k} q_{ji}$ and $(\mathbf{q}_1, \mathbf{q}_2, \cdots)$ has the same distribution as $(\mathbf{p}_1, \mathbf{p}_2, \cdots)$. Moreover, we make

$$\mathscr{G}(i) = \sigma(\mathbf{q}_1, \xi_1, \mathbf{q}_2, \xi_2, \cdots, \mathbf{q}_{i+1}).$$

In addition, for a particular i, let $Z_1, Z_2, \cdots$ be random variables, which conditional on $\mathscr{G}(i-1)$, are independent and identically distributed according to

$$P(Z_1 = j \mid \mathscr{G}(i-1)) = q_{ji}/q_i, \quad j \in \{1, \cdots, k\},$$

and further make $\{Z_m\}$ $\mathscr{G}(i-1)$-conditionally independent of ξ_i. We now define

$$\xi_{ji} = Y_i I[Z_1 = j]$$

so that

$$P(\xi_{ji} = 1 \mid \mathscr{G}(i-1)) = q_{ji}$$

and thus $(\xi_1, \xi_2, \cdots)$ has the same distribution as $(\mathbf{X}_1, \mathbf{X}_2, \cdots)$. Finally, we define a k-type point process $(\eta_1, \cdots, \eta_k)$ whose points are those of η and whose marks are assigned by the following rule: the mth point of $(i-1, i]$ is given the mark Z_m. Setting $\mathscr{G}(t) = \mathscr{G}([t]) \vee \sigma(\eta_j(z), z \le t, j \in \{1, \cdots, k\})$, the process $((\eta_1, \cdots, \eta_k), \mathscr{G})$ is thus a k-type point process whose continuous compensator $(\alpha_1, \cdots, \alpha_k)$ satisfies

$$\alpha_j(\infty) = \Sigma q_{ji}.$$

We now argue as before to see that the required distance is bounded by

$$P(\Sigma\xi_{ji} \ne \eta_j(\infty), \quad \text{for some} \quad j) + d((\eta_1(\infty), \cdots, \eta_k(\infty)), \text{Poisson } (\boldsymbol{\mu})).$$

The first term is dominated by the probability that ξ_i is not equal to $\eta(i) - \eta(i-1)$, for some i. For, if the contrary holds, $\eta(i) - \eta(i-1)$ is one or zero for all i and the same random variables are used to indicate which, if any, of ξ_{ji} and $\eta_j(i) - \eta_j(i-1)$ are one. Thus the same argument as in the univariate case gives the second term of the bound and the first is obtained from the argument given for continuous compensators.

4. Corollaries and extensions of Theorem 1. We first prove Corollary 1. To do this we need a simple Lemma.

LEMMA 1. *Let $\mathscr{P}$ and $\mathscr{P}'$ be two probability measures on $\mathscr{H} = \sigma(\mathscr{I})$ where $\mathscr{I}$ is a field and suppose that*

$$|\mathscr{P}(A) - \mathscr{P}'(A)| < \delta$$

for all A in $\mathscr{I}$. Then the same holds for all A in $\mathscr{H}$.

PROOF. Fix A in $\mathscr{H}$ and $\varepsilon > 0$. By the Caratheodory extension theorem, there is a set A' in $\mathscr{I}$ such that $A' \supset A$ and

$$\mathscr{P}(A') \le \mathscr{P}(A) + \varepsilon.$$

Thus,

$$\mathscr{P}'(A) \le \mathscr{P}'(A') \le \mathscr{P}(A') + \delta \le \mathscr{P}(A) + \varepsilon + \delta.$$

The Lemma now follows from the arbitrariness of ε and a symmetrical argument with $\mathscr{P}$ in place of $\mathscr{P}'$ and vice versa.

From the Lemma, for part (a), we need only show that, for arbitrary k and $0 = t_0 < t_1 < t_2 \cdots < t$, we have $d((N(t_0, t_1], \cdots, N(t_{k-1}, t_k]), \mathscr{I})$ bounded by the right-hand side of (a), where $\mathscr{I}$ is a vector of independent Poisson random variables with means $\mu(t_0, t_1]$,

$\cdots, \mu(t_{k-1}, t]$. To do this, let $j \in \{1, \cdots, k\}$, $t \geq 0$ and

$$N_j(t) = N(t \wedge t_j) - N(t \wedge t_{j-1}),$$

so that $((N_1, \cdots, N_k), \mathscr{F})$ is a k-type point process. Then, Theorem 1 part (c) gives a bound for the required distance of

$$E\{\textstyle\sum_{j=1}^k |A(t_{j-1}, t_j] - \mu(t_{j-1}, t_j]|\} + E\{\textstyle\sum_{s\leq t} \Delta A^2(s)\}.$$

By the definition of the variation norm, this is enough.

The argument for part (b) is very similar but notationally more complicated and is thus omitted.

Corollary 1 has the following immediate consequence when the processes have conditional intensities.

COROLLARY 4. *Under the conditions of Corollary* 1 *and the additional assumptions that*

$$A(t) = \int_0^t a(s)\, ds, \quad A_i(t) = \int_0^t a_i(s)\, ds$$

$$\mu(t) = \int_0^t \lambda(s)\, ds, \quad \mu_i(t) = \int_0^t \lambda_i(s)\, ds$$

the bounds of (a) *and* (b) *become*

$$E\left\{\int_0^t |a(s) - \lambda(s)|\, ds\right\}$$

and

$$\textstyle\sum_{j=1}^k E\left\{\int_0^t |a_j(s) - \lambda_j(s)|\, ds\right\},$$

respectively.

Theorem 1 can produce error bounds in Poisson approximations for general increasing processes. As in Brown (1981), starting with an increasing process $(N, \mathscr{F})$ and $0 \leq \varepsilon < 1$, we define the *ε-point process* $(N_\varepsilon, \mathscr{F})$ of $(N, \mathscr{F})$ by setting $N_\varepsilon(t)$ to be the number of $z \leq t$ such that $\Delta N(z) \in (1 - \varepsilon, 1 + \varepsilon]$. The *$\varepsilon$-remainder process* $(N^\varepsilon, \mathscr{F})$ (which is *not* $N - N_\varepsilon$) has $N^\varepsilon(t)$ equal to $N(t) - \int_0^t \Delta N(z)\, dN_\varepsilon(z)$. Working with the compensator A of $(N, \mathscr{F})$ (and not that of $(N_\varepsilon\, \mathscr{F})$, which is often awkward to calculate), it is not difficult to apply Lemma 1 of Brown (1981) to show that for a stopping time τ and $\mu \geq 0$

(4.1) $$d(N(\tau), \text{Poisson } (\mu)) \leq E|A(\tau) - \mu| + (1 - \varepsilon)^{-1} E\{\textstyle\sum_{s\leq\tau} \Delta A^2(s)\} + 2\varepsilon E\{A(\tau)\} + E\{N^\varepsilon(\tau)\} + P(N(\tau) \neq N_\varepsilon(\tau)).$$

[a key point in the proof is that

$$E\{\textstyle\sum_{s\leq\tau} \Delta A^2(s)\} = E\left\{\int_0^\tau \Delta A(s)\, dN(s)\right\}.]$$

This result is most likely to be of interest if N only takes integer values, in which case (taking $\varepsilon = 0$) we have

(4.2) $$d(N(\tau), \text{Poisson } (\mu)) \leq E|A(\tau) - \mu| + E\{\textstyle\sum_{s\leq\tau} \Delta A^2(s)\} + 2E(N(\tau) I[N \text{ has a jump of size 2 in } (0, \tau]]),$$

provided $E\{N(\tau)\} < \infty$. This bound may be readily applied to obtain a Poisson approximation result for U-statistics with $0-1$ summands. The bound has a similar form to that of Brown and Silverman (1979) (where the bound Serfling (1975) was used) but one obtains smaller universal constants with the above.

It is also quite easy to extend these arguments to approximations for the whole process, N^t. Thereby, (after truncation as in Brown, 1981) one obtains the Poisson case of the Cox convergence result of Brown (1981). We may also obtain a curious new convergence result. Namely, if $(N_n, \mathscr{F}_n)$ is a sequence of point processes with *continuous* compensators A_n and, for all $t \geq 0$, $A_n(t) \to \mu(t)$, in distribution, for an *arbitrary* right continuous increasing function μ, then the finite dimensional distributions of N_n converge to those of a Poisson process with mean function μ [Brown, 1981, requires μ to be continuous, while in Kabanov, Liptser and Shiryayev, 1981, the limiting process is not Poisson if μ is discontinuous]. The result is curious because in general it will not be true that N_n converges in the Skorohod J_1 topology. An example of this is obtained by considering a sequence of Poisson processes, whose continuous mean functions converge to a discontinuous function. In this case, each N_n will only have jumps of size 1, while the distributional limit will have a jump of size 2 with positive probability; this phenomenon is not consistent with J_1 convergence.

We can also use Theorem 1 and Corollary 1 to give approximations by compound Poisson distributions. To describe this, let $(\mathbf{N}, \mathscr{F})$ be a k-type point process with compensator $\mathbf{A}$. Fix a k-vector of reals, $\mathbf{Z}$. We can then obtain another stochastic process $f(\mathbf{N})$ defined by

$$f(\mathbf{N})(t) = \textstyle\sum_{i=1}^{k} Z_i N_i(t), \quad t \geq 0;$$

$f(\mathbf{N})$ jumps by Z_i at the time of each type i point of $\mathbf{N}$. If $\mathbf{\Pi}$ is a k-vector of independent Poisson processes, then $f(\mathbf{\Pi})(t)$ has a *compound Poisson distribution* and

$$f(\mathbf{\Pi})(t) =_d \textstyle\sum_{j=1}^{\Pi(t)} \xi_j$$

where $\Pi(t) = \Sigma \Pi_i(t)$ has a Poisson distribution and is independent of $\xi_1, \xi_2, \cdots$, which are also independent and identically distributed. The *parameters* of the compound Poisson distribution are defined to be the mean of $\Pi(t)$, the vector $\mathbf{z}$, and the probabilities p_i that ξ_1 takes value z_i.

Theorem 1 (c) gives

$$d(f(\mathbf{N})(t), f(\mathbf{\Pi})(t)) \leq \textstyle\sum_{i=1}^{k} E\,|A_i(t) - \mu_i| + \sum_{s \leq t} E\{\Delta A^2(s)\} \tag{4.3}$$

where $\mu_i = E\{\Pi_i(t)\}$, while Corollary 1 (b) gives

$$d(f(\mathbf{N})^t, f(\mathbf{\Pi})^t) \leq \textstyle\sum_{i=1}^{k} E\,|A_i - \mu_i|_t + \sum_{s \leq t} E\{\Delta A^2(s)\} \tag{4.4}$$

where μ_i is the mean measure of Π_i.

As an example of the use of (4.3) we consider a point process which is thinned by a Markov chain (see Isham, 1980, and Böker and Serfozo, 1982). Specifically, suppose η is a point process and that $\{X_i\}_{i \geq 0}$ is an independent stationary Markov chain with states 0 and 1. We let

$$\varepsilon = P(X_1 = 1 \mid X_0 = 0)$$

and

$$p = P(X_1 = 1 \mid X_0 = 1).$$

By means of a limit theorem, it is shown in Böker and Serfozo (1982) that, if ε is small and $\varepsilon\eta(t)$ is approximately μ (a constant) > 0, then

$$\eta'(t) = \textstyle\sum_{i=1}^{\eta(t)} X_i$$

has an approximate compound Poisson distribution with parameters μ, $(1, 2, 3, \cdots)$ and $((1-p), p(1-p), p^2(1-p), \cdots)$. Below, we find a bound on the error in this approximation. While the assumptions that the Markov chain is stationary and that it is independent of the point process are not necessary for the limit theorem to hold, the latter would appear to be essential to get an error expressed only in terms of $\eta(t)$, ε and p. The former is merely a convenience to keep the bound relatively simple.

We work with a jump process which is connected to η'. For $j \geq 1$, let

$$N_j(t) = \sum_{i=2}^{\eta(t)} (1 - X_{i-1})X_iX_{i+1} \cdots X_{i+j-1}(1 - X_{i+j}) + X_1 \cdots X_j(1 - X_{j+1})I[\eta(t) > 0]$$

so that $N_j(t)$ counts the number of success runs of length j starting on the trials $1, \cdots, \eta(t)$. If

$$Z(t) = \sum_1^\infty jN_j(t)$$

then

$$Z(t) = \eta'(t)$$

if for all $j \geq 1$ and all ℓ such that

$$\eta(t) - j + 1 \leq \ell \leq \eta(t)$$

we have

$$(1 - X_{\ell-1})X_\ell \cdots X_{\ell+j-1}(1 - X_{\ell+j}) = 0.$$

Hence

$$(4.5) \quad P(\eta'(t) \neq Z(t)) \leq \sum_{j=1}^\infty (j-1)\left(\frac{1-\varepsilon}{1-p+\varepsilon}\right)\varepsilon p^{j-1}(1-p) = \frac{\varepsilon p(1-\varepsilon)}{(1-p)(1-p+\varepsilon)}.$$

Let Y be a compound Poisson random variable with the parameters mentioned previously. To estimate $d(Z(t), Y)$ we use the bound of (4.3) with infinite sums replacing finite ones. There are two ways to justify the extension; it is possible to generalize Theorem 1 (c) to marked point processes with countably infinite mark space or, more simply, a truncation argument works because

$$E\{\textstyle\sum_{j=J}^\infty jN_j(t)\} < E\{\eta(t)\}p^J(J+1)/(1-p)$$

and we will assume $E\{\eta(t)\} < \infty$. Thus

$$d(Z(t), Y) \leq \sum_{j=1}^\infty E|A_j(t) - \mu p^{j-1}(1-p)| + E\{\textstyle\sum_{s\leq t} \Delta A^2(s)\}$$

where A_j is the compensator of $(N_j, \mathscr{F})$ for some history $\mathscr{F}$. We choose

$$\mathscr{F}(t) = \sigma(\eta) \vee \sigma(X_0, \cdots, X_{\eta(t)})$$

so that

$$A_j(t) = \{\textstyle\sum_{i=1}^{\eta(t)} \varepsilon p^{j-1}(1-p)(1-X_{i-1})\} + p \cdot p^{j-1}(1-p)X_0I(\eta(t) > 0).$$

Hence, after some manipulation using the fact that the stationary distribution is $((1-p)/(1-p+\varepsilon), \varepsilon/(1-p+\varepsilon))$

$$(4.6) \quad d(Z(t), Y) \leq E|\varepsilon\eta(t) - \mu| + (1-p+\varepsilon)^{-1}\varepsilon\{(2-p)E(\varepsilon\eta(t)) + p(1+p)P(\eta(t) > 0)\}.$$

Combining (4.5) and (4.6) we obtain a bound on the distance of the number of retained points, $\eta'(t)$, from Y of

$$E|\varepsilon\eta(t) - \mu| + (1-p+\varepsilon)^{-1}\varepsilon\{(2-p)E(\varepsilon\eta(t)) + p(1+p)P(\eta(t) > 0) + p(1-\varepsilon)/(1-p)\}.$$

Usually, the first term will be dominant; for example, if $\eta(t)$ is a unit rate renewal process evaluated at nt and $\varepsilon = n^{-1}$, then the first term is $O(\sqrt{n}^{-1})$, while the second is $O(n^{-1})$. This tends to confirm an intuition that the approximation here is roughly as good as approximating $\eta'(t)$ by a Poisson (μ) random variable, if $\varepsilon = p$. For in the latter case it is easy to see that a bound of

$$E|\varepsilon\eta(t) - \mu| + \varepsilon E(\varepsilon\eta(t))$$

results from Theorem 1.

We finally mention another possible extension for which we will not provide details. It would be possible to relax a little the assumption that points with different marks do not occur together in a marked point process. For example, in the setup of Proposition 1, one could consider vectors $\mathbf{X}_i$ such that X_{ji} is $\mathscr{F}_{i-1}$ conditionally independent of $X_{\ell i}$ for $\ell \neq j$.

5. Marked point processes with general mark space. In this section we give a bound on the distance of a finite point process on a suitable topological space from a Poisson process on that space. The finite point process is supposed to arise as the accumulation of the marks in a marked point process. The bound is in terms of the processes in a certain decomposition of the compensator of the marked point process. This decomposition is given in Jacod (1979), Theorem 3.15 and we follow this and Jacod (1975) for definitions and some notation. An important part of the proof of the bound is Lemma 2, which gives concrete interpretations of the parts of the decomposition. Corollary 8 gives a bound for the distance of the whole marked point process on $(0, t]$ from a time-homogeneous marked Poisson process.

Consider a random measure $(\boldsymbol{\eta}, \mathscr{F})$ on $\mathscr{B}(\mathbb{R}^+) \times \mathscr{E}$, where $(E, \mathscr{E})$ is a Lusin space. We define η^F to be the increasing process given for F in $\mathscr{E}$ and $t > 0$

$$\eta^F(t) = \boldsymbol{\eta}((0, t] \times F).$$

The stochastic process $\{\boldsymbol{\eta}(t)\}_{t\geq 0}$ will take value the measure $\boldsymbol{\eta}((0, t] \times \cdot)$ at time point t. By the construction of the compensator $\mathbf{A}$ of a marked point process $(\mathbf{N}, \mathscr{F}) = ((T_1, Z_1, \cdots), \mathscr{F})$ given in Theorem 3.15 of Jacod (1979), we have the decomposition

$$A^F(t) = \int_0^t B(s, F)\, dA^E(s) \tag{5.1}$$

where $B(\cdot, F)$ is predictable, $B(s, \cdot)$ is a probability measure on $\mathscr{E}$ and A^F is the compensator of $(N^F, \mathscr{F})$. Intuitively, A^E gives the accumulated conditional rate of occurrence of points, whereas B gives conditional distributions for the marks. More precisely, by Theorem 2.6 of Pitman (1981):

LEMMA 2. *The conditional distribution for Z_i given $\mathscr{F}(T_i-)$ is $B(T_i)$ on $[T_i < \infty]$.*

REMARK 1. By $B(T_i)$ on $[T_i < \infty]$ we mean the random measure defined for $\omega \in [T_i < \infty]$ as

$$\omega \mapsto B(\omega, T_i(\omega), \cdot).$$

To prove our bound, we will need the existence of *maximal couplings* on general spaces. I learned this from Kaijser (1981). Let $\mathscr{P}$ and $\mathscr{P}'$ be probability measures on some space $(S, \mathscr{S})$. Then there exists a probability measure $\tilde{\mathscr{P}}$ on $\mathscr{S} \times \mathscr{S}$ such that

$$d(\mathscr{P}, \mathscr{P}') = \tilde{\mathscr{P}}\{(x, y) : x \neq y\}. \tag{5.2}$$

[In fact, $\tilde{\mathscr{P}}$ has the explicit formula

$$\tilde{\mathscr{P}}(A \times B) = \mathscr{P}(A \cap B \cap F) + \mathscr{P}'(A \cap B \cap E) + \{d(\mathscr{P}, \mathscr{P}')\}^{-1}\{\alpha \times \beta(A \times B)\}$$

where $E \cup F$ is a Hahn decomposition for $\mathscr{P} - \mathscr{P}'$ and $\alpha - \beta$ is a Jordan decomposition of the same measure.] Provided $(S, \mathscr{S})$ is Polish, we may then realise $\tilde{\mathscr{P}}$ as the distribution of a random element of $S \times S$. This random element can be taken to be a function of a uniform random variable U and we will label it $MC(U; \mathscr{P}, \mathscr{P}')$ (Billingsley, 1968, page 26, Exercise 6). If $MC(U; \mathscr{P}, \mathscr{P}') = (X, Y)$ then (5.2) takes the form

$$d(\mathscr{P}, \mathscr{P}') = P(X \neq Y),$$

so that $MC(U; \mathscr{P}, \mathscr{P}')$ achieves equality in (1.6). We can now prove:

THEOREM 2. *Suppose* $(\mathbf{N}, \mathscr{F})$ *is a marked point process with compensator* $\mathbf{A}$ *and that* $A = A^E$ *and* B *are as described above. Let* μ *be a finite measure on* $\mathscr{E}$. *Then, if* A *is continuous,*

$$d(\mathbf{N}(t), \text{Poisson } (\mu)) \leq E\,|A(t) - \nu\,| + E\left\{\int_0^{\hat{A}(\nu)} d'(B(s), \mu')A(ds)\right\} \tag{5.3}$$

where $\nu = \mu(E)$, $\mu' = \mu/\nu$, $\hat{A}(t) = \inf\{z : A(z) > t\}$ *and* $d'(B(s), \mu')$ *is the stochastic process*

$$(\omega, s) \to d(B(\omega, s, \cdot), \mu').$$

REMARK 2. If A is nearly deterministic, $\hat{A}(\nu)$ will be close to some constant c. In this case, a convenient upper bound to the second term would be

$$E\left\{\int_0^c d'(B(s), \mu')A(ds)\right\} + E\,\{\sup_{s\leq A(\nu)\vee c} d'(B(s), \mu')|\,A(c) - t\,|\}$$

REMARK 3. One is tempted to write $d(B(s), \mu')$ for $d'(B(s), \mu')$ in the right side of the bound. However, by convention, $d(B(s), \mu')$ would mean the distance between the distribution of $B(s)$ (a distribution on measure space) and μ' (even though the latter is not meaningful). It will be important in the proof that $d'(B(s), \mu')$ is a predictable process; this follows from the fact that it equals $\frac{1}{2}\,|\,B(s) - \mu'\,|$ and the fact that the predictable signed random measures form a vector space (Jacod, 1979, Corollary 3.13).

PROOF. As in the proof of Theorem 1, we assume that $A(\infty) = \infty$. We proceed by constructing a new version $\boldsymbol{\eta} = (\tau_1, \zeta_1, \tau_2, \zeta_2, \cdots)$ of $\mathbf{N}$ and a Poisson process $\mathbf{\Pi} = (\sigma_1, \gamma_1, \sigma_2, \gamma_2, \cdots)$ on $\mathbb{R}^+ \times E$. The times $(\sigma_1, \sigma_2, \cdots)$ will be chosen to have the same joint distribution as $(A(T_1), A(T_2), \cdots)$; the fact that these latter give a unit rate Poisson process on $\mathbb{R}^+$ is just the random time change theorem (Liptser and Shiryayev, 1978, Theorem 18.10). The marks of $\mathbf{\Pi}$ all have distribution μ', so $\mathbf{\Pi}(\nu)$ has distribution Poisson (μ).

The basis of the construction is a set $U_1, U_2, U_3, \cdots$ of independent and identically distributed uniform random variables over (0, 1). By enlargement of the underlying probability space, if necessary, we can assume these independent of $\mathscr{F}(\infty)$. Let (q_1, σ_1) be a function of U_1 whose distribution is the same as that of $(B(T_1), A(T_1))$. Let $(\zeta_1, \gamma_1) = MC(U_2; q_1, \mu')$, i.e. conditional on U_1, (ζ_1, γ_1) has the maximally coupled joint law with marginals $q_1(U_1)$ and μ' respectively. Hence, since μ' is non-random, $\sigma_1 = \sigma_1(U_1)$ and γ_1 are independent. Further, by Lemma 2, (q_1, σ_1, ζ_1) has the same distribution as $(B(T_1), A(T_1), Z_1)$. Thus, by using conditional distributions, we may produce random variables τ_1, q_2, σ_2 which are functions of U_3, and such that

$$(q_1, \sigma_1, \zeta_1, \tau_1, q_2, \sigma_2) =_d (B(T_1), A(T_1), Z_1, T_1, B(T_2), A(T_2)).$$

Proceeding in the above way to produce $(\zeta_2, \gamma_2), \tau_2, q_3, \sigma_3, \cdots$ we obtain the processes $\boldsymbol{\eta}$ and $\mathbf{\Pi}$ with the required properties. Moreover, (η^E, Π^E) has the same distribution as $(N^E, N^E \circ \hat{A})$, so that the proof of Theorem 1 gives

$$P(\eta^E(t) \neq \Pi^E(\nu)) \leq E\,|\,A(t) - \nu\,|.$$

Thus, the probability that $\boldsymbol{\eta}(t)$ differs from $\mathbf{\Pi}(\nu)$ is dominated by the right side of the last equation plus

$$P(\zeta_i \neq \gamma_i \quad \text{for some} \quad i = 1, \cdots, \Pi^E(\nu) \quad \text{and} \quad \eta^E(t) = \Pi^E(\nu)).$$

Let $\mathscr{H}_i = \sigma(U_1, \cdots, U_{2i+1})$, $i = 0, 1, \cdots$. Denote $\Pi^E(\nu)$ by S. Then S is an $\{\mathscr{H}_i\}$ stopping time, since

$$S = \inf\{i : \sigma_{i+1} > \nu\}.$$

Thus the last probability is bounded by

$$E\{\sum_{i=1}^{S} I[\zeta_i \neq \gamma_i]\} = E\{\sum_{i=1}^{S} P(\zeta_i \neq \gamma_i \mid \mathscr{H}_{i-1})\} = E\{\sum_{i=1}^{S} d'(q_i, \mu')\}$$

by construction. The construction also ensured that

$$(S, q_1, q_2, \cdots) =_d (N \circ \hat{A}(\nu), B(T_1), B(T_2), \cdots)$$

and hence the right side of the previous equation is

$$E\{\sum_{i=1}^{N\circ\hat{A}(\nu)} d'(B(T_i), \mu')\} = E\left\{\int_0^{\hat{A}(\nu)} d'(B(s), \mu') N(ds)\right\}$$

which is the second term of the bound since $\hat{A}(\nu)$ is an $\mathscr{F}$-stopping time.

Corollary 5. *Using the same notation as Theorem* 2, *letting* m *be an arbitrary measure on* $(0, \infty)$ *and* Poisson $(m \times \mu^t)$ *be the distribution of a* Poisson *process on* $(0, t] \times E$ *having mean measure* $m \times \mu$ *restricted to* $(0, t] \times E$, *we have*

$$d(\mathbf{N}^t, \text{Poisson } (m \times \mu^t)) \leq E\,|A - \nu m|_t + E\left\{\int_0^{\hat{A}(\nu)} d'(B(s), \mu') A(ds)\right\},$$

where $\mathbf{N}^t$ *is the distribution of* $\mathbf{N}$ *restricted to* $(0, t] \times E$.

Proof. We use the same construction as in the proof of Theorem 2. The sets of the form $I_1 \cup \cdots \cup I_k$, where

$$I_j = (t_{j-1}, t_j] \times F_j, \quad F_j \in \mathscr{E}, \quad 0 = t_0 < \cdots < t_k = t$$

are a semiring which generates $\mathscr{B}((0, t]) \times \mathscr{E}$. Thus, by the argument of Lemma 1.4 of Kallenberg (1975), we need only bound

$$P[(\eta(I_1), \cdots, \eta(I_k)) \neq (\Pi(J_1), \cdots, \Pi(J_k))],$$

where $J_i = (s_{i-1}, s_i] \times F_i$, and $s_i = \nu m(0, t_i]$. This probability is bounded, as in Theorem 2, by the sum of two terms, the first of which is

$$P(\eta^E(t_{j-1}, t_j] \neq \Pi^E(s_{j-1}, s_j] \quad \text{for some} \quad j)$$

and thus is bounded by the expression in the statement of the Corollary using the (univariate) bound of Corollary 1. The second term coincides with that of Theorem 1 and the proof is complete.

Corollary 5 can even give better bounds than Corollary 1 in the case of a k-type point process. As a simple example consider a unit rate Poisson process on $\mathbb{R}^+$, $\Pi = (T_1, T_2, \cdots)$, and an independent stationary Markov chain $\{X_n\}$ which takes k values $\{1, 2, \cdots, k\}$. Let

$$\mathbf{N}((0, t] \times \{j\}) = \sum_{T_i \leq t} I[X_i = j], \quad j = 1, 2, \cdots, k.$$

If $\mathbf{A}$ is the compensator of $(\mathbf{N}, \mathscr{F})$, where $\mathscr{F}(t) = \sigma(T_1, \cdots, T_i, X_1, \cdots, X_i, \text{ for } T_i \leq t)$, then

$$\mathbf{A}((0, t] \times \{j\}) = \int_0^t P_{Y(s)j}\, ds$$

where (P_{ij}) is the transition matrix of $\{X_n\}$ and $Y(s)$ is the left continuous process $X_{N(s-)}$. The bound of Corollary 5 for the distance of $\mathbf{N}$ from a k-vector of independent Poisson processes with rates given by the stationary distribution $(p_1, p_2, \cdots, p_k)$ of $\{X_n\}$ is

$$E \int_0^t \left\{\frac{1}{2} \sum_{j=1}^{k} |P_{Y(s)j} - p_j|\right\} ds,$$

since $A(t) = t$. But this expression simplifies to

$$t/2 \sum_{i=1}^{k} \sum_{j=1}^{k} |P_{ij} - p_j| p_i$$

whereas the bound of Corollary 1 gives

$$E\left\{\sum_{j=1}^{k} \int_0^t |P_{Y(s)j} - p_j| \, ds\right\}$$

which is exactly twice the previous bound.

Theorem 2 and Corollary 5 can be used to give compound Poisson approximations for jump processes whose jump set is uncountable. Given a marked point process $(\mathbf{N}, \mathscr{F})$ with mark space $\mathbb{R}$ and compensator $\mathbf{A}$, we define $f(\mathbf{N})$ by

$$f(\mathbf{N})(t) = \sum_{T_i \le t} Z_i$$

where the T's are the times and the Z's the marks of the process. If $\mathbf{\Pi}$ is a Poisson process on $(0, \infty) \times \mathbb{R}$ with mean measure $m \times \mu$, then the analogue of (4.3) is, assuming A to be continuous,

$$d(f(\mathbf{N})(t), f(\mathbf{\Pi})(t)) < E|A(t) - \nu| + E\left\{\int_0^{\hat{A}(\nu)} d'(B(s), \mu')A(ds)\right\} \tag{5.4}$$

where $\nu = m(t)\mu(\mathbb{R})$ and other symbols on the right have the same meanings as in Theorem 2. Again assuming A to be continuous, the analogue of (4.4) is that the bound of Corollary 5 is also a bound for the total variation distance between the distributions of $f(\mathbf{N})^t$ and $f(\mathbf{\Pi})^t$.

Acknowledgment. The author is grateful to P. K. Pollett for conversations which stimulated much of the work here, and to the referee for some helpful suggestions.

REFERENCES

[1] ALDOUS, D. J. (1979). Weak convergence of stochastic processes for processes viewed in the Strasbourg manner. Preprint, University of California, Berkeley.
[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.
[3] BÖKER, F. and SERFOZO, R. (1982). Ordered thinnings of point processes and random measures. *Stochastic Process. Appl.,* to appear.
[4] BROWN, T. C. (1978). A martingale approach to the Poisson convergence of simple point processes. *Ann. Probab.* **6** 615–628.
[5] BROWN, T. C. (1979). Some distributional approximations for random measures. Ph.D. thesis, University of Cambridge.
[6] BROWN, T. C. (1982). Poisson approximations and exchangeable random variables. In *Exchangeability in Probability and Statistics.* North-Holland, Amsterdam.
[7] BROWN, T. C. and POLLETT, P. K. (1982). Some distributional approximations for Markov queueing networks. *Adv. Appl. Probab.* **14** 654–671.
[8] BROWN, T. C. and SILVERMAN, B. W. (1979). Rates of Poisson convergence for U-statistics. *J. Appl. Probab.* **16** 428–432.
[9] DELLACHERIE, C. and DOLEANS-DADE, C. (1970). Un contreexemple au problème des Laplaciens approĉhes. Séminaire de Probabilités V. *Lecture Notes in Mathematics* **191**. Springer-Verlag, Berlin.
[10] FREEDMAN, D. (1974). The Poisson approximation for dependent events. *Ann. Probab.* **2** 256–269.
[11] ISHAM, V. (1980). Dependent thinning of point processes. *J. Appl. Probab.* **17** 987–995.
[12] JACOD, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. verw. Gebiete* **31** 235–253.
[13] JACOD, J. (1979). Calcul Stochastique et Problèmes de Martingales. *Lecture Notes in Mathematics* **714**. Springer-Verlag, Berlin.
[14] JACOD, J. and MEMIN, J. (1980). Sur la convergence des sémimartingales vers un processus à accroissements indépendants. Séminaire de Probabilités XIV. *Lecture Notes in Mathematics* **784** 227–248. Springer-Verlag, Berlin.
[15] KABANOV, Y., LIPTSER, R. and SHIRYAYEV, A. (1980). Some limit theorems for simple point processes (martingale approach). *Stochastics* **3** 203–216.

[16] KABANOV, Y. and LIPTSER, R. (1982). On convergence in variation of the distributions of multivariate point processes. Unpublished manuscript.
[17] KAIJSER, T. (1981). On a new contraction condition for random systems with complete connections. *Rev. Roum. Math. Pure et Appl.* **26** 1075–1117.
[18] KALLENBERG, O. (1975). *Random Measures.* Academic, London.
[19] LIPTSER, R. and SHIRYAYEV, A. (1978). *Statistics of Random Processes. II.* Springer-Verlag, New York.
[20] MEMIN, J. (1982). Distance en variation et conditions de contignite pur des lois de processus ponctuels. Unpublished manuscript.
[21] Meyer, P. A. (1971). Démonstration simplifiée d'un théorème de Knight. Séminaire de Probabilités V. *Lecture Notes in Mathematics* **191**. Springer-Verlag, Berlin.
[22] MURALI-RAO, K. (1969). On decomposition theorems of Meyer. *Math. Scand.* **24** 66–78.
[23] PAPANGELOU, F. (1974). Integrability of expected increments of point processes and a related random change of scale. *Trans. Amer. Math. Soc.* **165** 483–506.
[24] PITMAN, J. W. (1982). Lévy systems and path decompositions. In *Seminar on Stochastic Processes*, 1981. Birkhäuser, Boston.
[25] REBOLLEDO, R. (1978). Sur les applications de la théorie des martingales à l'étude statistique d'une famille de processus ponctuels. In *Journées de Statistique des Processus Stochastiques. Lecture Notes in Mathematics* **636**. Springer-Verlag, Berlin.
[26] SERFLING, R. J. (1975). A general Poisson approximation theorem. *Ann. Probab.* **3** 726–731.
[27] VALKEILA, E. (1982). A general Poisson approximation theorem. *Stochastics* **7** 159–172.

DEPARTMENT OF MATHEMATICS
MONASH UNIVERSITY, CLAYTON
VIC. 3168, AUSTRALIA