ON BASIC RESULTS OF POINT PROCESS THEORY

M. R. LEADBETTER
UNIVERSITY OF NORTH CAROLINA

1. Introduction

There are many existing approaches to the theory of point processes. Some of these—following the original work of Khinchin [9] are "analytical" and others (for example, [15], [8]) quite abstract in nature. Here we will take a position somewhat in the middle in describing the development of some of the basic theory of point processes in a relatively general setting, but by using largely the simple techniques of proof described for the real line in [11]. We shall survey a number of known results—giving simple derivations of certain existing theorems (or their adaptations in our setting) and obtain some results which we believe to be new. Our framework for describing a general point process will be essentially that of Belyayev [2], while that for Section 4 concerning Palm distributions is developed from the approach of Matthes [14].

First we give the necessary background and notation. There are various essentially equivalent ways of defining the basic structure of a point process. For example, for point processes on the line, one may consider the space of integer valued functions $x(t)$ with $x(0) = 0$, which increase by a finite number of jumps in any finite interval. The events of the process then correspond to jumps of $x(t)$. One advantage of such a specification is that multiple events fit naturally into the framework.

To define point processes on an arbitrary space T, it is often appropriate to consider the "sample points" ω to be subsets of T. This is the point of view taken in [18], where each ω is itself a countable subset of the real line, the set of points "where events occur." Sometimes, however, a point process arises from some existing probabilistic situation (such as the zeros of a continuous parameter stochastic process) and one may wish to preserve the existing framework in the discussion. A convenient structure for this is the following, used in [2]. Let $(\Omega, \mathscr{F}, P)$ be a probability space and $(T, \mathscr{T})$ a measurable space (T is the space "in which the events will occur"). For each $\omega \in \Omega$, let S_ω be a subset of T. If for each $E \in \mathscr{T}$

$$N(E) = N_\omega(E) = \text{card } (E \cap S_\omega) \tag{1.1}$$

is a (possibly infinite valued) random variable, then S_ω is called a *random set* and the family $\{N(E): E \in \mathscr{T}\}$ a *point process*. The "events" of the process are, of course, the points of S_ω.

The model may be generalized slightly to take account of *multiple events*—that is, the possible occurrence of more than one event at some $t \in T$. The definition of $N(E)$ as card $(E \cap S_\omega)$ shows that $N(E)$ is an integer valued measure (on the subsets of T) for each ω. As a measure, $N_\omega(\cdot)$ has its mass confined to S_ω and $N_\omega(\{t\}) = 1$ for each $t \in S_\omega$.

To allow multiple events, we may simply redefine $N_\omega(E)$ to be an integer valued measure with all its mass confined to S_ω, with $N_\omega(\{t\}) \geqq 1$ for each $t \in S_\omega$ and such that $N_\omega(E)$ is a random variable for each $E \in \mathscr{T}$. If $N_\omega(\{t\}) > 1$, we say a *multiple event* occurs at t. If there is zero probability that any $t \in S_\omega$ is multiple, we say the process is *without multiple events.*

If we say a process may have multiple events, we shall be referring to this framework and shall write $M(E)$ for the number of multiple events in E. In such a case, we shall write $N^*(E)$ for card $\{S_\omega \cap E\}$ and refer to $N^*(E)$ as the number of events in E *without regard to their multiplicities.* (Of course, $N(E)$ is the total number of events in E.)

In the manner just described, a point process may be regarded as a special type of *random measure.* This concept has been developed in considerable generality for stationary cases (see, for example, [15]), but this generality will not be pursued here.

Another method of taking account of multiple events is to replace each $t \in S_\omega$ by a pair (t, k_t), where k_t is a "mark" associated with t denoting the multiplicity. This again is capable of considerable generalization by considering rather arbitrary kinds of "marks" and the appropriate additional measure theoretic structure. These ideas have been developed by Matthes (see, for example, [14]) for stationary point processes on the real line and provide an elegant framework for obtaining results, for example, in relation to Palm distributions. In such cases, the marks are chosen to be highly dependent on the set S_ω (for example, translates of S_ω). At the same time, most results of interest can be obtained by using essentially these techniques, but without explicit reference to marks. Hence, we here use the framework previously explained.

For stationary point processes on the real line, there are several important basic theorems. Included among these are (writing $N(s, t)$ for the number of events in $(s, t]$):

(i) the theorem of Khinchin regarding the existence of the *parameter* $\lambda = \lim_{t \downarrow 0} Pr\,\{N(0, t) \geqq 1\}/t$;

(ii) Korolyuk's theorem which, in its sharpest form, says that for a stationary point process without multiple events, λ is equal to the *intensity* $\mu = \mathscr{E}N(0, 1)$ that is, the mean number of events per unit time; λ and μ may be infinite; if multiple events may occur, we replace μ by $\mathscr{E}N^*(0,1)$;

(iii) for the regular (orderly, ordinary) case (that is, when $Pr\,\{N(0, t) > 1\} = o(t)$ as $t \downarrow 0$) multiple events have probability zero;

(iv) "Dobrushin's lemma"—a converse to (iii)—stating that if $\lambda < \infty$ and multiple events have probability zero, then the process is regular.

Various analogues of these results have been studied for nonstationary point processes on the real line in [20], [4], [5], largely by using properties of Burkhill integrals. A clarifying and general viewpoint has been more recently given by Belyayev [2]. Specifically in [2], generalizations of the two constants λ, μ are made in terms of *measures* $\lambda(\cdot)$, $\mu(\cdot)$ on the space T, instead of in terms of point functions. The *principal measure* $\mu(\cdot)$ is defined (as customarily) on $\mathscr{T}$ simply by $\mu(E) = \mathscr{E}N(E)$—countable additivity of N guaranteeing countable additivity of $\mu(\cdot)$. On the other hand, the *parametric measure* $\lambda(\cdot)$ is defined in [2] by

$$(1.2) \qquad \lambda(E) = \sup\left\{\sum_1^\infty Pr\{N(E_i) > 0\} : E_i \in T, E_i \text{ disjoint}, \bigcup_1^\infty E_i = E\right\}.$$

It is easily shown that $\lambda(\cdot)$ is a measure and it is clear that $\lambda(E) \leqq \mu(E)$ for all $E \in \mathscr{T}$. For a stationary point process on the real line, $\lambda(E) = \lambda m(E)$ and $\mu(E) = \mu m(E)$, where m denotes Lebesgue measure.

Using these definitions, it is possible to extend the basic results quoted above to apply to point processes which may be nonstationary, on spaces T more general than the line (including any Euclidean space). These generalizations are systematically described in Section 2. In Section 3, stationarity is discussed in general terms (with particular reference to Khinchin's theorem) when T is a topological group. In both these sections, the general lines of development are those of [2], with adaptation of the results in presenting a somewhat different viewpoint, and with emphasis on simplicity of proofs obtained by direct analogy with those of [11].

Finally, in Section 4, we discuss some basic results relative to Palm distributions (and their expressions as limits of conditional probabilities), for stationary point processes on the real line. The approach is essentially that of [14] (without explicit reference to marks), again with emphasis on the simplicity of proofs obtained from the techniques of [11].

2. The basic general theorems

The notation already developed will be used throughout this section. We shall systematically obtain the generalizations of the basic theorems referred to in Section 1. This development follows the same general lines as [2] but with differences of detail and perspective.

All that is to be said in *general* relative to Khinchin's theorem concerning the existence of the intensity, is contained in Belyayev's definition of the parametric measure (1.2) given in Section 1. (For special cases, when T has a group structure and the point process is stationary, it is possible to say more that is directly analogous to the real line case—mention of this will be made later.)

It is shown in [2] that the truth of the generalized version of Korolyuk's theorem, namely, $\lambda(E) = \mu(E)$ for all $E \in \mathscr{T}$ (for a point process without multiple events or $\lambda(E) = \mathscr{E}N^*(E)$ if multiple events may occur), depends on the structure of T rather than on any stationarity assumption. The proof given directly generalizes that of [11] for stationary processes on the real line. This is most clearly seen for a nonstationary process on the real line ($\mathscr{T}$ then being the Borel sets). For then if E is an interval $(a, b]$, we may divide E into n equal subintervals E_{ni}, $i = 1, \cdots, n$, and write $\chi_{ni} = 1$ if $N(E_{ni}) \geqq 1$, $\chi_{ni} = 0$ otherwise. Assuming there are no multiple events, it is easily seen that $N_n = \Sigma_{i=1}^n \chi_{ni} \rightarrow N(E)$ with probability one, as $n \rightarrow \infty$, and hence by Fatou's lemma,

$$\mu(E) \leqq \liminf \mathscr{E} N_n = \liminf \sum_{i=1}^{n} Pr\{\chi_{ni} = 1\} \leqq \lambda(E). \tag{2.1}$$

But $\lambda(E) \leqq \mu(E)$, and hence, $\lambda(E) = \mu(E)$ for all E of the form $(a, b]$. Thus, $\lambda(E) = \mu(E)$ for all Borel sets E, provided μ is σ-finite.

For the above proof to be useful when T is a more general space, we require T to have sets playing the role of intervals. A suitable definition of such a class of sets is given by Belyayev [2] and called a "fundamental system of dissecting sets" for T. Here we shall use a somewhat different definition to achieve the desired results. Specifically, we here say that a class $\mathscr{C} = \{E_{nk}: n, k = 1, 2 \cdots\}$ of sets $E_{nk} \in \mathscr{T}$ is a *dissecting system* for T if

(i) $\mathscr{C}$ is a "determining class" (see [3]) for σ-finite measures on $\mathscr{T}$; that is, two σ-finite measures equal on $\mathscr{C}$ are equal on $\mathscr{T}$ (for example, $\mathscr{C}$ may be a semiring generating $\mathscr{T}$);

(ii) for any given set $E \in \mathscr{C}$, there is corresponding to each $n = 1, 2, 3, \cdots$ a set I_n of integers such that

(a) E_{nk} are disjoint subsets of E for $k \in I_n$ with $E - \cup_{k \in I_n} E_{nk} \subset F_n \in \mathscr{T}$, where $F_n \downarrow \varnothing$, the empty set, as $n \rightarrow \infty$, and

(b) given any two points $t_1 \neq t_2$ of E, for all sufficiently large values of n (that is, for all $n \geqq$ some $n_0(t_1, t_2)$), there are sets $E_{nk_1}, E_{nk_2}, k_1, k_2 \in I_n(k_1 \neq k_2$, such that $t_1 \in E_{nk_1}, t_2 \in E_{nk_2}$.

For example, for the real line, we may take E_{nk} to be any interval $(a, b]$ with rational endpoints and of length $1/n$. We note also that the requirement in (ii) (a) that $\lim F_n = \varnothing$ may be replaced by $Pr\{N(\lim F_n) = 0\} = 1$, but this, of course, depends on the process as well as the structure of T.

The proof of Korolyuk's theorem given for the real line now generalizes at once to apply to a point process on a space T possessing a dissecting system. This is easily seen from the following lemma.

LEMMA 2.1. *Consider a point process on a space T possessing a dissecting system $C = \{E_{nk}\}$. With the above notation for $E \in C$, $k \in I_n$, write $\chi_{nk} = 1$ if $N(E_{nk}) > 0$, $\chi_{nk} = 0$ otherwise. Let*

$$N_n = \sum_{k \in I_n} \chi_{nk}. \tag{2.2}$$

Then

$$N_n \rightarrow N^*(E) \leqq \infty \text{ with probability one,} \tag{2.3}$$

as $n \rightarrow \infty$.

Further, if $\chi^*_{nk} = 1$ when $N(E_{nk}) > 1$ and $\chi^*_{nk} = 0$ otherwise, and if $N(E) < \infty$ with probability one, then

$$\sum_{k \in I_n} \chi^*_{nk} \to M(E) \text{ with probability one,} \tag{2.4}$$

as $n \to \infty$.

PROOF. It is clear that $N_n \leqq N(E)$. On the other hand, if $\infty \geqq N(E) \geqq m$, there are points $t_1, \cdots, t_m$, where events occur. For large n, these are eventually contained in different sets E_{nk} and hence $N_n \geqq m$. Equation (2.3) follows by combining these results.

The second part follows by noting that since (with probability one) only a finite number of distinct events occur, they are eventually contained in different E_{nk} sets when n is large, and thus for such n, $N^*_n = M(E)$.

By using the first part of this lemma, Korolyuk's theorem follows as for the real line by a simple application of Fatou's lemma. Stated specifically we have (see [2]):

THEOREM 2.1 (Generalized Korolyuk's theorem). *For a point process, with σ-finite principal measure, on a space T possessing a dissecting system, we have $\lambda(E) = \mathscr{E}N^*(E)$ for all $E \in \mathscr{T}$. In particular if there are no multiple events, then the principal and parametric measures coincide on $\mathscr{T}$.*

If $\lambda(\cdot)$ or $\mu(\cdot)$ is absolutely continuous with respect to some σ-finite measure v on T, then under the above conditions the densities $d\lambda/dv$, $d\mu/dv$ coincide a.e. This reduces again to the usual statement of Korolyuk's theorem for stationary point processes on the line.

A point process on T is called regular (orderly, ordinary—of [20], [5] and especially [2]) with respect to a dissecting system $\mathscr{C} = \{E_{nk}\}$ if

$$\lim_{n \to \infty} \sup_k \frac{Pr\{N(E_{nk}) > 1\}}{Pr\{N(E_{nk}) > 0\}} = 0. \tag{2.5}$$

(For simplicity, we shall always assume $Pr\{N(E_{nk}) > 0\} \neq 0$ for any n, k.) This definition applies to a point process which may conceivably have multiple events. However, the next result shows that in fact regularity precludes the occurrence of multiple events under simple conditions on T.

THEOREM 2.2. *Consider a point process (allowing multiple events) on a space T possessing a dissecting system $\mathscr{C}$. Suppose the process is regular and that there exist $E_n \in \mathscr{C}$, $E_n \uparrow T$ such that $\lambda(E_n) < \infty$. Then, with probability one, the process has no multiple events.*

PROOF. Let $E \in \mathscr{C}$ be such that $\lambda(E) < \infty$. Write again $M(E)$ for the number of *multiple* events in E. Then by Lemma 2.1, $M(E) = \lim_{n \to \infty} \Sigma_{k \in I_n} \chi^*_{nk}$ (with the usual notation), where χ^*_{nk} is one or zero according as $N(E_{nk}) > 1$ or not. Hence,

$$\begin{aligned} \mathscr{E}M(E) &\leqq \liminf_n \Big[\sum_{k \in I_n} Pr\{N(E_{nk}) > 1\}\Big] \\ &\leqq \liminf_n \left[\sup_k \frac{Pr\{N(E_{nk}) > 1\}}{Pr\{N(E_{nk}) > 0\}} \sum_{j \in I_n} Pr\{N(E_{nj}) > 0\}\right], \end{aligned} \tag{2.6}$$

which is zero since by regularity the first term in the braces tends to zero, and the sum does not exceed $\lambda(E) < \infty$. Since $\mathscr{E}M(E)$ is a measure on $\mathscr{T}$, $\mathscr{E}M(T) = \lim_{n\to\infty} \mathscr{E}M(E_n) = 0$, and hence, $M(T) = 0$ with probability one.

A converse result of Theorem 2.2 is "Dobrushin's lemma." A general form of this given in [2] assumes a "homogeneous" point process—for which T possesses a fundamental system $\mathscr{C}$ of dissecting sets such that $p_{nk}(0) = Pr\{N(E_{nk}) > 0\}$, $p_{nk}(1) = Pr\{N(E_{nk}) > 1\}$ are each dependent on n but not on k for $E_{nk} \in \mathscr{C}$. This assumption does not imply stationarity of the point process (indeed there may be no "translations" defined on T), but it may well be that the only interesting homogeneous processes are stationary ones. We give a less restricted result below. It may still be of greatest interest in the stationary case, but it does allow considerable variation in the quantities $p_{nk}(0)$, $p_{nk}(1)$ for fixed n.

Specifically to obtain Dobrushin's lemma, we shall assume the existence of a dissecting system $\mathscr{C} = \{E_{nk}\}$ for which there is a sequence $\{\theta_n\}$ of nonnegative real numbers, and a function $\phi(\theta) \to 0$ as $\theta \to 0$, such that for each n

$$\theta_n \leqq \frac{Pr\{N(E_{nk}) > 1\}}{Pr\{N(E_{nk}) > 0\}} \leqq \phi(\theta_n) \tag{2.7}$$

for all k.

THEOREM 2.3 (Generalized version of Dobrushin's lemma). *Consider a point process without multiple events, on a space T possessing a dissecting system $\mathscr{C}$ satisfying* (2.7). *Suppose $\lambda(E) < \infty$ for some $E \in \mathscr{C}$. Then the point process is regular.*

PROOF. Using the notation of Lemma 2.1, we have

$$\sum_{k\in I_n} \chi_{nk} \to N(E), \qquad \sum_{k\in I_n} \chi^*_{nk} \to 0 \tag{2.8}$$

with probability one, as $n \to \infty$. Since both sums are dominated by $N(E)$ and $\mathscr{E}N(E) = \mu(E) = \lambda(E) < \infty$, it follows by dominated convergence that

$$\sum_{k\in I_n} Pr\{N(E_{nk}) > 0\} = \mathscr{E}\{\sum_{k\in I_n} \chi_{nk}\} \to \mu(E) = \lambda(E), \tag{2.9}$$

and similarly that

$$\sum_{k\in I_n} Pr\{N(E_{nk}) > 1\} \to 0. \tag{2.10}$$

Hence by (2.7),

$$\theta_n \sum_{k\in I_n} Pr\{N(E_{nk}) > 0\} \leqq \sum_{k\in I_n} Pr\{N(E_{nk}) > 1\} \to 0, \tag{2.11}$$

and thus by (2.9), $\theta_n \to 0$ $\big(\lambda(E) \geqq Pr\{N(E) > 0\} > 0$ since $E \in \mathscr{C}\big)$.

Finally, from (2.7) again,

$$\sup_k \left[\frac{Pr\{N(E_{nk}) > 1\}}{Pr\{N(E_{nk}) > 0\}}\right] \leqq \phi(\theta_n) \to 0 \tag{2.12}$$

as $n \to \infty$.

3. Stationarity generalities

A very great deal of literature exists relative to stationary point processes on the real line (see Section 4). One expects to be able to say less about stationary point processes on the plane or in R^n (see, for example, [6]). However, there is quite a good deal that may be said even when T is just assumed to be a (locally compact) topological group. In this section, we comment briefly on a few aspects of such results.

If T is a locally compact (Hausdorff) group, the natural σ-field $\mathscr{T}$ is the class of Borel sets—generated by the open sets of T. It is usually convenient to assume (and we here do) that T is also σ-compact, and then $\mathscr{T}$ is also generated by the compact sets of T. (It is, in fact, sometimes assumed that T is second countable; for example, [15]. While this additional assumption may be necessary for some purposes it does, however, imply that the group is also metrizable.)

For a point process on such a group T, stationarity may be defined in terms of the invariance of the joint distributions of $N(tE_1), \cdots, N(tE_n)$ for $t \in T$, where n is any fixed positive integer and the E_i are any fixed sets of $\mathscr{T}$ ($tE = \{ts : s \in E\}$, ts denoting the group operation). If T is not abelian this gives a concept of "left stationarity," "right stationarity" being correspondingly defined.

Under (say, left) stationarity, the principal and parametric measures $\mu(\cdot)$, $\lambda(\cdot)$ are (left) invariant Borel measures which are regular provided their values on compact sets are finite ([7], Theorem 64I)—which we will assume. Thus, $\lambda(\cdot)$ and $\mu(\cdot)$ are just constant multiples of the Haar measure $m(\cdot)$ on T, $\lambda(E) = \lambda m(E)$, $\mu(E) = \mu m(E)$, say, for all $E \in \mathscr{T}$, where λ and μ are constants, the *parameter* and the *intensity* of the stationary point process, respectively. Questions concerning the parameter and intensity in such a setting have been discussed to some extent in [1]. The general line of argument above is that of [2].

If in addition T possesses a dissecting system $\mathscr{C} = \{E_{nk}\}$ of, say, bounded sets (that is, having compact closures) and if a stationary point process on T is without multiple events, then Theorem 2.1 shows that $\lambda = \mu < \infty$. This is Korolyuk's theorem in the stationary case. Further, in such a case it is not unreasonable to suppose that $P\{N(E_{nk}) > 0\}$ and $m(E_{nk})$ are independent of k (which will hold if, for example, for fixed n the E_{nk} are translates of each other). Then using the notation of Theorem 2.3 we have, from the proof of that theorem,

$$r_n Pr\,\{N(E_{n0}) > 0\} \to \lambda(E) = \lambda m(E) \tag{3.1}$$

for $E \in \mathscr{C}$, where r_n is the (necessarily finite) number of integers in the set I_n and E_{n0} is any given E_{nk} for $k \in I_n$.

But since by definition of $\mathscr{C}$.

$$E - \bigcup_{k \in I_n} E_{nk} \subset F_n \downarrow \varnothing, \tag{3.2}$$

it follows that

$$r_n m(E_{n0}) = \sum_{k \in I_n} m(E_{nk}) \to m(E), \tag{3.3}$$

and hence, that

$$\frac{Pr\{N(E_{n0}) > 0\}}{m(E_{n0})} \to \lambda \tag{3.4}$$

as $n \to \infty$. It is this latter property that the parameter satisfies in Khinchin's existence theorem. We summarize this as a theorem. For convenience of statement, we will *here* call a dissecting system $\mathscr{C}$ *homogeneous* if the distribution of $N(E_{nk})$ and $m(E_{nk})$ do not depend on k for each fixed n.

THEOREM 3.1. *Consider a stationary point process without multiple events on locally compact group T. Suppose T is also σ-compact. Then there exist constants λ, μ such that $\lambda(E) = \lambda m(E)$, $\mu(E) = \mu m(E)$ for all $E \in \mathscr{T}$, where $m(\cdot)$ is the Haar measure of T, $0 \leqq \lambda \leqq \mu < \infty$.*

Suppose, in addition, that T has a homogeneous dissecting system $\mathscr{C} = \{E_{nk}\}$ of bounded sets E_{nk}. Then the point process is regular, $\lambda = \mu$ and

$$\lim_{n \to \infty} \frac{Pr\{N(E_{n0}) > 0\}}{m(E_{n0})} = \lambda, \tag{3.5}$$

where E_{n0} is any E_{nk}.

COROLLARY 3.1. *The stated results hold if the condition that $\mathscr{C}$ be a homogeneous dissecting system is replaced by the requirement that for each fixed n, the sets E_{nk} are all translates of each other.*

The above remarks have been concerned with a stationary process without multiple events. When multiple events are allowed, the appropriate generalizations of the real line results occur. For example, if E is a set of $\mathscr{T}$ with $\mu(E) < \infty$ (for example, E compact), and if $N_s(E)$ denotes the number of those events in E which have "multiplicity" $s = 1, 2, \cdots$, then $p_s = \mathscr{E}N_s(E)/\lambda(E)$ is a probability distribution on the integers $1, 2, \cdots$. we may interpret $\{p_s\}$ as the "probability that an event has multiplicity s." If in addition T has a homogeneous dissecting system $\mathscr{C} = \{E_{nk}\}$ and we choose $E \in \mathscr{C}$ with $\mu(E) < \infty$ writing $\chi_{nk}^s = 1$ if $N(E_{nk}) = s$ and $\chi_{nk}^s = 0$ otherwise, then, similarly to Lemma 2.1,

$$\sum_{k \in I_n} \chi_{nk}^s \to N_s(E) \tag{3.6}$$

as $n \to \infty$, with probability one. The familiar argument of taking expectations and using dominated convergence shows that $r_n Pr\{N(E_{n0}) = s\} \to \mathscr{E}N_s(E)$, where E_{n0} is any E_{nk} and r_n is the number of points in I_n. Similarly, $r_n Pr\{N(E_{n0}) \geqq 1\} \to \lambda(E)$. Thus,

$$p_s = \lim_{n \to \infty} Pr\{N(E_{n0}) = s \,|\, N(E_{n0}) \geqq 1\}, \tag{3.7}$$

giving intuitive justification to the description of p_s as the probability that an event has multiplicity s (under these assumptions p_s does not depend on E). Further questions of this type are considered in [16] when $T = R^n$. We note that the above calculation may also be considered as a special case of that in the next section concerning Palm distributions.

4. Concerning Palm distributions

For a stationary point process, the Palm distribution P_0 gives a precise meaning to the intuitive notion of conditional probability "given an event of the process occurred at some point (for example, $t = 0$)." When T is the real line, we may write (for certain sets $F \in \mathscr{F}$)

$$P_0(F) = \lim_{\delta \downarrow 0} Pr\,\{F \,|\, N((-\delta, 0)) \geqq 1\}. \tag{4.1}$$

That is, $P_0(F)$ is then the limit of the conditional probability of F given an event occurred in an interval near $t = 0$ as that interval shrinks. For example, if F denotes the occurrence of at least one event in the interval $(0, t]$ (that is, $\{N(0, t) \geqq 1\}$), then $P_0(F) = F_1(t)$, the distribution function for the time to the first event after time zero given an event occurred "at" time zero.

This kind of procedure for particular sets F was used by Khinchin [9] and is useful in providing an "analytical" approach to such conditional probabilities (see, for example, [12]). More sophisticated and general measure theoretic treatments involving the definition and properties of P_0 have been given by a variety of authors (for example, [14], [15], [17], [18], [19]). In this section, we shall use a "middle of the road" approach to the definition of P_0 (based essentially on [14]) which is capable of considerable generality. Our main purpose will be to give simple proofs for formulae such as (4.1) and its generalizations to include "conditional expectations" of functions. Such results have application, for example, to the evaluation of the distributions of the times between events in terms of conditional moments [13].

We give the construction of the Palm distribution P_0 for stationary point process on the real line in the manner of [14], though, from a somewhat different viewpoint. The construction generalizes to apply to point processes on groups (see also [15]), but we consider just the real line case for simplicity relative to the later results.

Consider, then, a stationary point process (without multiple events for simplicity) with finite parameter λ on the real line. Again for simplicity, we take the sample points ω to be themselves the subsets S_ω of $T = R^1$, that is, ω is a countable set of real numbers (without finite limit points, since $\lambda < \infty$). Denote by $\mathscr{T}$ the Borel sets of $T = R^1$ and by $\mathscr{F}$ the smallest σ-field on Ω making $N(B) = N_\omega(B)$ measurable for each $B \in \mathscr{T}$. Finally, we shall again write $N(s, t)$ for $N\{(s, t]\}$, the number of events (card $\{\omega \cap (s, t]\}$) in the semiclosed interval $(s, t]$.

For any real t, and $\omega \in \Omega$, let $\omega_t \in \Omega$ denote the set of points of ω translated to the left by t; that is, if $\omega = \{t_i\}$, $\omega_t = \{t_i - t\}$. If F is any set of $\mathscr{F}$ and $\omega \in \Omega$, $\omega = \{t_i\}$, say, we define $\omega^* \in \Omega$ to consist of precisely those points $t_i \in \omega$ for which $\omega_{t_i} \in F$. In other words, to form ω^*, we "thin" ω by retaining only the points t_i such that ω_{t_i} (that is, ω translated to t_i as origin) is in F. The ω^* define a stationary point process formed from some of the events in the original point process. For example, if $F = \{\omega\colon N(0, t) \geqq 1\}$, the new process contains precisely those events t_i of the old process which are followed by a further event within a further time t (that is, no later than $t_i + t$). Write N_F for the number of events of the thinned process in the interval (0, 1), that is, card $\{\omega^* \cap (0, 1)\}$. Then the thinned process has intensity $\lambda_F = \mathscr{E} N_F$.

Now this procedure may be carried out for any $F \in \mathscr{F}$, and for fixed ω, N_F is countably additive as a function on $\mathscr{F}$. It follows at once that λ_F is a measure on $\mathscr{F}$, and hence, that

$$P_0(F) = \frac{\lambda_F}{\lambda} \tag{4.2}$$

is a probability measure on $\mathscr{F}$, $\lambda = \lambda_\Omega = \mathscr{E}N(0, 1)$. This P_0 is the desired *Palm distribution.*

To give P_0 an intuitive interpretation, one wishes to prove relations such as (4.1). Equation (4.1) is not universally true, however, as can be seen by considering a "periodic" stationary point process in which the events occur at a regular spacing h, where the distance to the first one after $t = 0$ is a uniform random variable on $(0, h)$. For this process take F to be the occurrence of at least one event in the open interval $(h - \eta, h)$. Clearly, $Pr\{F \mid N(-\delta, 0) \geqq 1\} = 1$ when $\delta < \eta$. But $N_F = 0$, and hence $P_0(F) = 0$.

We give now a class of sets for which (4.1) does hold. Specifically, we shall call a set $F \in \mathscr{F}$ *right continuous* if its characteristic function $\chi_F(\omega)$ is such that $\chi_F(\omega_t)$ is continuous to the right in t; that is $\chi_F(\omega_s) \to \chi_F(\omega_t)$ as $s \downarrow t$. Equivalently, this means that for any t, if $\omega_t \in F$ then $\omega_s \in F$ when s is sufficiently close to t on the right, and conversely.

THEOREM 4.1. *Suppose $F \in \mathscr{F}$ is a right continuous set. Then*

$$Pr\{F \mid N(-\delta, 0) \geqq 1\} \to P_0(F) \tag{4.3}$$

as $\delta \downarrow 0$.

PROOF. Let δ_m be any sequence of nonnegative numbers converging to zero as $m \to \infty$. Write r_m for the integer part $[\delta_m^{-1}]$ of δ_m^{-1}. Divide the interval (0, 1) into r_m intervals of length δ_m (with perhaps an interval of length less than δ_m left over). Write $\chi_{mi} = 1$ if $N((i - 1)\delta_m, i\delta_m) \geqq 1$, $\chi_{mi} = 0$ otherwise, $i = 0, 1 \cdots r_m$. Let

$$N_m = \sum_{i=1}^{r_m} \chi_{mi} \chi_F(\omega_{i\delta_m}). \tag{4.4}$$

Then N_m denotes the number of intervals $((i - 1)\delta_m, i\delta_m]$ containing an event and such that the translate $\omega_{i\delta_m}$ is in F. But by the right continuity assumption, if an event occurs at t_0, then $\omega_{t_0} \in F$ if and only if $\omega_{i\delta_m} \in F$ for that interval $((i - 1)\delta_m, i\delta_m]$ containing t_0 when m is sufficiently large. Further, with probability one, when m is sufficiently large the events all lie in different intervals and there is no event in the last short interval. Hence, with probability one,

$N_m \to N_F$ as $m \to \infty$. Since $N_m \leqq N(0, 1)$ and $\mathscr{E}N(0, 1) < \infty$, it follows by dominated convergence that $\mathscr{E}N_m \to \lambda_F$ as $m \to \infty$. That is,

$$\sum_{i=1}^{r_m} Pr\{\chi_{mi} = 1, \omega_{i\delta_m} \in F\} \to \lambda_F \tag{4.5}$$

or by stationarity, $r_m Pr\{\chi_{m0} = 1, \omega \in F\} \to \lambda_F$. But

$$Pr\{\chi_{m0} = 1, \omega \in F\} = Pr\{F \mid N(-\delta_m, 0) \geqq 1\}\, Pr\{N(-\delta_m, 0) \geqq 1\}. \tag{4.6}$$

Hence, since $r_m \sim \delta_m^{-1}$ and $Pr\{N(-\delta_m, 0) \geqq 1\} \sim \lambda\delta_m$, we have

$$Pr\{F \mid N(-\delta_m, 0) \geqq 1\} \to \frac{\lambda_F}{\lambda} = P_0(F), \tag{4.7}$$

as required.

As an example, consider the set $F = \{\omega : N(0, t) \geqq r\}$, $r = 1, 2, \cdots$. This is easily seen to be right continuous, and hence the theorem applies. In this case,

$$P_0(F) = \lim_{\delta \downarrow 0} Pr\{N(0, t) \geqq r \mid N(-\delta, 0) \geqq 1\} \tag{4.8}$$

is interpreted as the distribution function for the time to the rth event after time zero, given an event occurred "at" time zero. (Note that at least r events occur in $(0, t)$ if and only if the time to the rth event after time zero does not exceed t.)

Similarly, if we take $0 < t_1 \leqq t_2 \cdots \leqq t_k$, $0 \leqq r_1 \leqq r_2 \cdots \leqq r_k$, and

$$F = \{\omega : N(0, t_1) \geqq r_1, N(0, t_2) \geqq r_2, \cdots, N(0, t_k) \geqq r_k\}, \tag{4.9}$$

then F is right continuous, leading to what could naturally be termed the joint distribution function for the time to the r_1st, r_2nd, $\cdots$, r_kth events after the origin given an event occurred at the origin.

The convergence in Theorem 4.1 does not occur for all $F \in \mathscr{F}$ in general. However, we may regard the probability space Ω as consisting of real integer valued functions increasing by unit jumps where events occur, and consider it as a subspace of D, the space of functions with discontinuities of the first kind (see [3]) where D has the "Skorohod topology." Then Theorem 4.1 may be shown to imply weak convergence of $P_\delta = P(\cdot \mid N(-\delta, 0) > 0)$ to P_0.

A slightly different definition given by Matthes ([14]) does give convergence similar to Theorem 4.1 for all $F \in \mathscr{F}$. Specifically, let $s = s(\omega)$ denote the time of the first event prior to the origin. Then instead of $P_\delta(F) = P\{\omega : \omega \in F \mid N(-\delta, 0) > 0\}$, we may consider $P_\delta^*(F) = P\{\omega : \omega_s \in F \mid N(-\delta, 0) > 0\}$. That is, the "origin is moved" slightly to the point s of $(-\delta, 0)$, where an event occurs. Then the following theorem (which is virtually identical to that of [10], Section 1(f)) holds.

THEOREM 4.2. *For each* $F \in \mathscr{F}$,

$$P_\delta^*(F) \to P_0(F) \tag{4.10}$$

as $\delta \downarrow 0$ *(hence the total variation of* $P_\delta^* - P_0$ *tends to zero* $\delta \downarrow 0$*).*

PROOF. This can be proved as in [10], Section 1. However, a very easy proof follows by a simplification of the method of Theorem 4.1. In fact, using the notation of that proof, we consider N_m^* (instead of N_m), where

$$N_m^* = \sum_{i=1}^{r_m} \chi_{mi}\chi_F(\omega_{s_{mi}}) \tag{4.11}$$

with s_{mi} denoting the position of the last event prior to (or at) $i\delta_m$. (A contribution to the sum only occurs if this event is in $((i-1)\delta_m, i\delta_m]$.) Then, with probability one, N_m^* converges to the number of events $t_i \in (0, 1)$ for which $\omega_{t_i} \in F$. It follows as before by dominated convergence and stationarity that $r_m Pr\{\chi_{m0} = 1, \omega_s \in F\} \to \lambda_F$, from which the desired result follows (for any sequence $\delta_m \downarrow 0$) using the fact that $Pr\{N(-\delta_m, 0) \geqq 1\} \sim \lambda/r_m$.

The fact that the limit in (4.10) holds for all $F \in \mathscr{F}$ is, of course, more satisfying than that in (4.1) which requires "continuity sets." However, the definition of P_δ^* is more complicated than that of P_δ and the limit in (4.1) may be more useful in practice. The difference between P_δ and P_δ^* is, of course, slight (but we feel it worthy of exploration).

Theorems 4.1 and 4.2 concerned conditional expectations of the function χ_F, given an event near the origin. One may ask whether similar results hold for other functions. To answer this in relation to Theorem 4.1, we will call a measurable function $\phi(\omega)$ *continuous to the right* if $\phi(\omega_t)$ is continuous to the right in t.

Before stating the generalization of Theorem 4.1, we give a lemma (the result of which is contained in [14]) which is useful in a number of contexts.

LEMMA 4.1. *If ϕ is a measurable function on Ω and ϕ is either nonnegative, or integrable with respect to P_0, then*

$$\lambda \int \phi dP_0 = \mathscr{E}\{\Sigma\, \phi(\omega_{t_j}) : t_j \in \omega \cap (0, 1)\}. \tag{4.12}$$

The statement of this lemma, when $\phi = \chi_F$, $F \in \mathscr{F}$, is just the definition of $P_0(F)$. Its truth for nonnegative measurable or P_0 integrable ϕ follows at once by the standard approximation technique.

The following result generalizes Theorem 4.2.

THEOREM 4.3. *Let ϕ be measurable (on Ω), continuous to the right and such that $|\phi(\omega_t)| < \psi(\omega)$ for all $t \in (0, 1)$, where $\mathscr{E}\{\psi N(0, 1)\} < \infty$. Then*

$$\mathscr{E}\{\phi \mid N(-\delta, 0) \geqq 1\} \to \int \phi dP_0 \tag{4.13}$$

as $\delta \downarrow 0$.

PROOF. The pattern of the proof of Theorem 4.1 applies, with ϕ written for χ_F Specifically,

$$S_m = \sum_{i=1}^{r_m} \chi_{mi}\phi(\omega_{i\delta_m}) \to \sum \{\phi(\omega_{t_j}) : t_j \in \omega \cap (0, 1)\} \tag{4.14}$$

with probability one. But $|S_n| \leqq \psi(\omega)\cdot N(0, 1)$ which has finite expectation, and thus, by dominated convergence and Lemma 4.1,

$$(4.15)\qquad \int \phi dP_0 = \lambda^{-1} \lim_{m\to\infty} \mathscr{E}S_m = \lambda^{-1} \lim_{m\to\infty} r_m \mathscr{E}\{\chi_{m0}\phi(\omega)\} = \lim_{m\to\infty} \mathscr{E}\{\phi \mid \chi_{m0} = 1\}$$

since $Pr(\chi_{m0} = 1) \sim \lambda\delta_m \lambda r_m^{-1}$. This is the desired result (writing $N(-\delta_m, 0) \geqq 1$ for $\chi_{m0} = 1$).

COROLLARY 4.1. *If ϕ is a bounded, right continuous function, the result holds. For if $|\phi| \leqq K$ we may take $\psi(\omega) = K$ and $\mathscr{E}\{\psi N(0, 1)\} = K\lambda < \infty$.*

This corollary is similar to a theorem of Ryll-Nardzewski [18] (there two sided continuity of $\phi(\omega_t)$ is required and the condition $N(-\delta, 0) \geqq 1$ replaced by $N(-\delta, \delta) \geqq 1$).

COROLLARY 4.2. *Suppose $\mathscr{E}N^{k+1}(0, \tau) < \infty$ for some positive integer k, $\tau > 0$. Then*

$$(4.16)\qquad \lim_{\delta \downarrow 0} \mathscr{E}\{N^k(0, \tau) \mid N(-\delta, 0) \geqq 1\} = \mathscr{E}_{P_0}N^k(0, \tau),$$

where $\mathscr{E}_{P_0}$ denotes expectation with respect to the Palm distribution. That is, the kth moment of $N(0, \tau)$ with respect to the Palm distribution is simply the kth conditional moment (defined as a limit) given an event "at" the origin.

The proof is immediate on noting that $\phi(\omega) = N_\omega^k(0, \tau)$ is continuous to the right, and for all $t \in [0, 1]$,

$$(4.17)\qquad \phi(\omega_t) = N_\omega(t, t + \tau) \leqq N(0, 1 + \tau) = \psi(\omega),$$

where $\mathscr{E}\{\psi(\omega)N(0, 1)\} \leqq \mathscr{E}N^{k+1}(0, 1 + \tau)$. This latter quantity is finite since it is easily seen by Minkowski's inequality and stationarity that $\mathscr{E}N^{k+1}(0, s) < \infty$ for all $s > 0$.

Finally, we note the corresponding generalization of Theorem 4.2. For this, the condition required above that ϕ be continuous to the right can be omitted, but the origin must "be moved" to measure from the time s of the first event prior to zero. We state this formally.

THEOREM 4.4. *Let ϕ be measurable and such that $|\phi(\omega_t)| < \psi(\omega)$ for all $t \in (0, 1)$, where $\mathscr{E}\{\psi N(0, 1)\} < \infty$. Then*

$$(4.18)\qquad \mathscr{E}\{\phi(\omega_s) \mid N(-\delta, 0) \geqq 1\} \to \int \phi dP_0$$

as $\delta \downarrow 0$, where $s = s(\omega)$ denotes the position of the first event prior to $t = 0$.

REFERENCES

[1] R. A. AGNEW, "Transformations of uniform and stationary point processes," Ph.D. thesis, Northwestern University, 1968.

[2] YU. K. BELYAYEV, "Elements of the general theory of random streams," Appendix to Russian edition of *Stationary and Related Stochastic Processes* by H. Cramér and M. R. Leadbetter, MIR, Moscow, 1969. (English translation under preparation for University of North Carolina Statistics Mimeograph Series, No. 703, 1970.)

[3] P. BILLINGSLEY, *Convergence of Probability Measures*, New York, Wiley, 1968.

[4] W. FIEGER, "Eine für beliebige Call-Prozesse geltende Verallgemeinerung der Palmschen Formeln," *Math. Scand.*, Vol. 16 (1965), pp. 121–147.

[5] ———, "Zwei Verallgemeinerungen der Palmschen Formeln," *Transactions of the Third Prague Conference on Information Theory*, Prague,—Czechoslovak Academy of Sciences, 1964, pp. 107–122.

[6] J. R. GOLDMAN, "Infinitely divisible point processes in R^n," *J. Math. Anal. Appl.*, Vol. 17 (1967), pp. 133–146.

[7] P. R. HALMOS, *Measure Theory*, Princeton, Van Nostrand, 1950.

[8] T. E. HARRIS, "Random measures and motions of point processes," *Z. Wahrscheinlichkeitstheorie und verw. Gebiete*, Vol. 18 (1971), pp. 85–115.

[9] Y. A. KHINCHIN, *Mathematical Methods in the Theory of Queueing*, London, Griffin, 1960.

[10] D. KÖNIG and K. MATTHES, "Verallgemeinerungen der Erlangschen Formeln, I," *Math. Nachr.*, Vol. 26 (1963), pp. 45–56.

[11] M. R. LEADBETTER, "On three basic results in the theory of stationary point processes," *Proc. Amer. Math. Soc.*, Vol. 19 (1968), pp. 115–117.

[12] ———, "On streams of events and mixtures of streams," *J. Roy. Statist. Soc. Ser. B*, Vol. 28 (1966), pp. 218–227.

[13] ———, "On certain results for stationary point processes and their application," *Bull. Inst. Internat. Statist.*, Vol. 43 (1969) pp. 309–319.

[14] K. MATTHES, "Stationäre zufällige Punktfolgen I," *Jber. Deutsch. Math.-Verein.*, Vol. 66 (1963), pp. 66–79.

[15] J. MECKE, "Stationäre zufällige Masse auf lokalkompakten Abelschen Gruppen," *Z. Wahrscheinlichkeitstheorie und verw. Gebiete*, Vol. 9 (1967), pp. 36–58.

[16] R. K. MILNE, "Simple proofs of some theorems on point processes," *Ann. Math. Statist.*, Vol. 42 (1971), pp. 368–372.

[17] J. NEVEU, "Sur la structure des processus ponctuels stationnaires," *C. R. Acad. Sci. Paris, Ser. A*, Vol. 267 (1968), pp. 561–564.

[18] C. RYLL-NARDZEWSKI, "Remarks on processes of calls," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1961, Vol. 2, pp. 455–466.

[19] I. M. SLIVNYAK, "Stationary streams of homogeneous random events," *Vestnik Harkov. Gos. Univ.*, Vol. 32 (1966), pp. 73–116. (In Russian.)

[20] F. ZITEK, "On the theory of ordinary streams," *Czechoslovak Math. J.*, Vol. 8 (1958), pp. 448–458. (In Russian.)