The research in this report was supported by the Office of Naval Research under Contract No. N00014-67-A-0321-0002.

ON BASIC RESULTS OF POINT PROCESS THEORY

by

M. R. LEADBETTER

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 686

June, 1970
On Basic Results of Point Process Theory

M. R. Leadbetter
University of North Carolina at Chapel Hill

ERRATA

Page 6 10 ADD: provided \( \mu \) is \( \sigma \)-finite.

\(-7, "\sigma\text{-finite}" should be written before "measures" in each case.

8 Theorem 2.1 should include the condition that the principal measure be \( \sigma \)-finite.

9 -2 REPLACE: "Pr\{N(E) < \infty\}" by "\( \lambda(E) < \infty \)."

10 1 SHOULD READ: \( M(E) = \lim \inf \sum_{n} \chi_{nk}^{*} \)

\( \sum_{n} k \in I_{n} \)

4, "\( \lim \inf \)" should be inserted before each right-hand expression.

5 7 ADD: Hence \( E_{n}(T) = \lim_{n \to \infty} E_{n}(E_{n}) = 0 \)

7 -4 "\( \lambda(E) < \infty \)" should read "\( \lambda(E) > 0 \)."
ON BASIC RESULTS OF POINT PROCESS THEORY

by

M. R. LEADBETTER
Department of Statistics
University of North Carolina at Chapel Hill

Summary.

Basic results of point process theory (such as Khintchine's theorem, Korolyuk's theorem, Dobrushin's lemma) are discussed in a general setting. The treatment is based on an approach of Belayev, with emphasis on simplicity of proof. Palm distributions for stationary point processes on the real line are discussed (in essence along lines developed by Matthes) with particular reference to their expression as limits of conditional probabilities. A simple technique is used to prove known and obtain further results.

The research in this report was supported by the Office of Naval Research under Contract No. N00014-67-A-0321-0002.
1. INTRODUCTION. There are many existing approaches to the theory of point processes. Some of these -- following the original work of Khintchine [9] are "analytical" and others (e.g. [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 Belaev [2], while that for the section 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 $x(0) = 0$, and which increase

---

The research in this report was supported by the Office of Naval Research under Contract No. N00014-67-A-0321-0002.
by integer jumps, of which there are a finite number 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" \( \omega \) to be subsets of \( T \). This is the point of view taken in [18], where each \( \omega \) 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 \( (\Omega, \mathcal{F}, P) \) be a probability space and \( (T, T) \) a measurable space. \( T \) is the space "in which the events will occur". For each \( \omega \in \Omega \), let \( S_\omega \) be a subset of \( T \). If for each \( E \in T \)

\[
N(E) = N_\omega(E) = \text{card}(E \cap S_\omega)
\]

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

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 \( \text{card}(E \cap S_\omega) \) shows that \( N(E) \) is an integer-valued measure on \( T \) for each \( \omega \). As a measure, \( N_\omega(\cdot) \) has its mass confined to \( S_\omega \), and \( N_\omega(\{t\}) = 1 \) for each \( t \in S_\omega \) (we assume all one point sets are in \( T \)). To allow multiple events, we may simply redefine \( N_\omega(E) \) to be an integer-valued measure on \( T \) with all its mass confined to \( S_\omega \) and such
that $N_{\omega}(\{t\}) \geq 1$ for each $t \in S_{\omega}$. 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. In such a case, we shall write $N^*(E)$ for $\text{card}(S_{\omega} \cap E)$ and refer to $N^*(E)$ as the number of events in $E$ without regard to their multiplicities. ($N(E)$, of course 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 e.g. [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 e.g. [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 $S_{\omega}$ (e.g. translates of $S_{\omega}$). 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 Khintchine regarding the existence of the parameter $\lambda = \lim_{t \to 0} \Pr[N(0,t) \geq 1]/t$, 


(ii) Korolyuk's theorem which, in its sharpest form, says that for a stationary point process without multiple events, $\lambda$ is equal to the intensity $\mu = EN(0,1)$ (i.e., the mean number of events per unit time; $\lambda$ and $\mu$ may be infinite).

(iii) For the regular (orderly, ordinary) case (i.e., when $Pr(N(0,t) > 1) = o(t)$ as $t \rightarrow 0$) 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 non-stationary 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 Belayev [2]. Specifically in [2], generalizations of the two constants $\lambda, \mu$ 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 $T$ simply by $\mu(E) = EN(E)$ -- countable additivity of $N$ guaranteeing countable additivity of $\mu(\cdot)$. On the other hand, the parametric measure $\lambda(\cdot)$ is defined in [2] by

\[
\lambda(E) = \sup \left\{ \sum_{l=1}^{\infty} Pr(N(E_l) > 0) : E_l \in T, E_l \text{ disjoint}, \bigcup_{l=1}^{\infty} E_l = E \right\}
\]

$\lambda(\cdot)$ is easily shown to be a measure and it is clear that $\lambda(E) \leq \mu(E)$ for all $E \in T$. For a stationary point process on the real line $\lambda(E) = \lambda m(E)$, $\mu(E) = \mu m(E)$, 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 non-stationary, 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 Khintchine'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 Khintchine's theorem concerning the existence of the intensity, is contained in Belayev's definition of the parametric measure (1.1) 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, viz. $\lambda(E) = \mu(E)$ for all $E \in T$ (for a point process without multiple events), depends on the structure of $T$ rather than on any stationarity assumption. The proof given directly generalizes that of
for stationary processes on the real line. This is most clearly seen
for a non-stationary process on the real line ($T$ then being the Borel sets).
For then if $E$ is an interval $(a,b]$, we may divide $E$ into $n$ equal
subintervals $E_{n,i} (i = 1, \ldots n)$ and write $\chi_{ni} = 1$ if $N(E_{ni}) \geq 1$,
$\chi_{ni} = 0$ otherwise. Assuming there are no multiple events, it is easily
seen that $N = \sum_{i=1}^{n} \chi_{ni} \to N(E)$ with probability one, as $n \to \infty$, and hence
by Fatou's lemma,

$$\mu(E) \leq \liminf_{n} EN_{n} = \liminf_{n} \sum_{i=1}^{n} \text{Pr}(\chi_{ni} = 1)$$

$$\leq \lambda(E).$$

But $\lambda(E) \leq \mu(E)$ and hence $\lambda(E) = \mu(E)$ for all $E$ of the form $(a,b]$.
Thus $\lambda(E) = \mu(E)$ for all Borel sets $E$.

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 Belayer [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 $C = \{E_{nk} : n, k = 1, 2, \ldots\}$ of sets $E_{nk} \in T$ is a dissecting system for
$T$ if

(i) $C$ is a "determining class" (cf. [3]) for measures on $T$.
That is two measures equal on $C$ are equal on $T$ (e.g. $C$ may be a
semiring generating $T$).

(ii) For any given set $E \in C$

(a) Corresponding to each integer $n \geq 1$ there is a set
$I_{n}$ of integers such that $F_{nk}$ are disjoint for $k \in I_{n}$ and

$$E = \bigcup_{k \in I_{n}} E_{nk} \subset F_{n} \in T$$
where $F_n \neq \emptyset$, the empty set, as $n \to \infty$.

(b) Given any two points $t_1 \neq t_2$ of $E$, for all sufficiently large values of $n$ (i.e., all $n \geq n_0(t_1,t_2)$), 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 of length $1/n$. We note also that the requirement in (ii) (a) that $\lim F_n = \emptyset$ 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 as 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 process without multiple events 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

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

Then

$$N_n \Rightarrow N(E) \leq \infty \text{ with probability one, as } n \to \infty. \tag{2.1}$$

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

$$\sum_{k \in I_n} \chi_{nk}^* \to 0 \text{ with probability one, as } n \to \infty.$$
Proof. It is clear that \( N_n \leq N(\mathcal{E}) \). On the other hand, if \( \omega \geq N(\mathcal{E}) \geq m \), there are points \( t_1 \ldots t_m \) where events occur. For large \( n \), these are eventually contained in different sets \( E_{nk} \) and hence \( N_n \geq m \). (2.1) follows by combining these results.

The second part follows by noting that since (with probability one) only a finite number of events occur they are eventually contained in different \( E_{nk} \) sets when \( n \) is large and thus for such \( n \), \( N_n^\ast = 0 \).

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 (as contained in [2]),

**THEOREM 2.1:** (Generalized Korolyuk's theorem.)

For any point process, without multiple events, on a space \( T \) possessing a dissecting system, the parametric and principal measures coincide.

If \( \lambda(\cdot) \) or \( \mu(\cdot) \) is absolutely continuous with respect to some \( \sigma \)-finite measure \( \nu \) on \( T \), then under the above conditions the densities

\[
\frac{d\lambda}{d\nu}, \quad \frac{d\mu}{d\nu}
\]

coincide a.e. This reduces again to the usual statement of Korolyuk's theorem for stationary point processes on the line.

It was assumed in the above derivation that multiple events have zero probability. The statement is easily modified to apply to a point process where multiple events may occur as follows:

**COROLLARY to Theorem 2.1:** Consider a point process which may have multiple events, on a space \( T \) possessing a dissecting system. For \( \mathcal{E} \subseteq T \), let
\( N^*(E) \) denote the number of events in \( E \) counted without regard to multiplicity. Let \( \mu^*(E) = E N^*(E) \). Then the parametric measure \( \lambda(E) \) is equal to \( \mu^*(E) \) for all \( E \in \mathcal{T} \).

**Proof:** A new point process may be defined in which one event occurs at \( t \) if there is at least one event of the original process at \( t \) (i.e., multiple events are replaced by a single event). The parametric measure of the new process is the same as that of the old; viz. \( \lambda(\cdot) \). The principal measure of the new process is \( \mu^*(\cdot) \). Hence the result follows by applying Theorem 2.1 to the new process.

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

\[
\lim_{n \to \infty} \sup_k \frac{P(N(E_{nk}) > 1)}{P(N(E_{nk}) > 0)} = 0.
\]

(For simplicity, we shall always assume \( P(N(E_{nk}) > 0) \neq 0 \) 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 \( \mathcal{T} \).

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

**Proof.** Let \( E \in \mathcal{C} \) be such that \( \Pr(N(E) < \infty) = 1 \). Write \( N(E) \) for the number of multiple events in \( E \). Then
\[ M(E) \leq \sum_{k \in I_n} \chi_{nk}^* \]

(with the usual notation) where \( \chi_{nk}^* \) is one or zero according as \( N(E_{nk}) > 1 \) or not. Hence

\[
EM(E) \leq \sum_{k \in I_n} \Pr(N(E_{nk}) > 1) \leq \sum_{j \in I_n} \Pr(N(E_{nj}) > 0)
\]

which tends to zero as \( n \to \infty \) since by regularity the first term tends to zero, and the sum does not exceed \( \lambda(E) < \infty \).

A converse result of Theorem 2.2 is "Dobrushin's lemma". The general form of this given in [2] assumes a "homogeneous" point process -- for which \( T \) possesses a fundamental system \( C \) 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 \( E_{nk} \in C \). 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 \( C = \{E_{nk}\} \) for which there is a sequence \( \{\theta_n\} \) of non-negative real numbers, and a function \( \phi(\theta) \to 0 \) as \( \theta \to 0 \), such that for each \( n \)

\[
(2.2) \quad \theta_n \leq \Pr(N(E_{nk}) > 1)/\Pr(N(E_{nk}) > 0) \leq \phi(\theta_n), \quad \text{all } k.
\]
**Theorem 2.3:** (Generalized version of Dobrushin's Lemma.)

Consider a point process without multiple events, on a space $T$ possessing a dissecting system $C$ satisfying (2.2). Suppose $\lambda(E) < \infty$ for some $E \in C$. Then the point process is regular.

**Proof.** Using the notation of Lemma 2.1, we have

$$\sum_{k \in I_n} \chi_{nk} \to N(E), \quad \sum_{k \in I_n} \chi_{nk}^* \to 0$$

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

$$(2.3) \quad \sum_{k \in I_n} \Pr(N_{nk} > 0) = E \{ \sum_{k \in I_n} \chi_{nk} \} + \mu(E) = \lambda(E)$$

and similarly that

$$\sum_{k \in I_n} \Pr(N_{nk} > 1) \to 0.$$

Hence by (2.2)

$$\theta_n \sum_{k \in I_n} \Pr(N_{nk} > 0) \leq \sum_{k \in I_n} \Pr(N_{nk} > 1) \to 0$$

and thus by (2.3) $\theta_n \to 0$ (since $\lambda(E) < \infty$).

Finally from (2.2) again

$$\Delta \Phi \left[ \Pr(N_{nk} > 1) / \Pr(N_{nk} > 0) \right] \leq \phi(\theta_n) \to 0$$

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 $\mathbb{R}^n$ (see e.g. [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 $\sigma$-field $\mathcal{T}$ 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 $\sigma$-compact, and then $\mathcal{T}$ is also generated by the compact sets of $T$. (It is in fact sometimes (e.g. [15]) assumed that $T$ is second countable. 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 $N(tE_1)\ldots N(tE_n)$ for $t \in T$ where $n$ is any fixed positive integer, and $E_i$ are any fixed sets of $T$ ($tE = \{ts : s \in E\}$, $t$ denoting the group operation). If $T$ is not abelian this gives a concept of "left stationarity", "right stationarity" being correspondingly defined.

Under (e.g. left) stationarity, the principal and parametric measures $\mu(\cdot)$, $\lambda(\cdot)$ are (left) invariant Borel measures which are [7, Theorem 6.41] regular provided their values on compact sets are finite -- 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 $E \subseteq T$, where $\lambda$ and $\mu$ 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 above general line of argument is that of [2].
If in addition $T$ possesses a dissecting system $C = \{E_{nk}\}$ of, say bounded sets (i.e., 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, for $E \in C$

$$r_n P(N(E_{n0}) > 0) + \lambda(E) = \lambda m(E)$$

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 $C$,

$$E = \bigcup_{k \in I_n} E_{nk} \subset F_n + \phi$$

it follows that

$$r_n m(E_{n0}) = \sum_{k \in I_n} m(E_{nk}) + m(E)$$

and hence that

$$P(N(E_{n0}) > 0)/m(E_{n0}) \to \lambda \quad \text{as} \quad n \to \infty.$$ 

It is this latter property that the parameter satisfies in Khitchine's existence theorem. We summarize this as a theorem. For convenience of statement, we will here call a dissecting system $C$ 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 \( \sigma \)-compact. Then there exist constants \( \lambda, \mu \) such that \( \lambda(E) = \lambda m(E), \mu(E) = \mu m(E) \) for all \( E \in T \) where \( m(\cdot) \) is the Haar measure of \( T \) \((0 \leq \lambda \leq \mu < \infty)\).

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

\[
\lim_{n \to \infty} \frac{\Pr(N(E_{n0}) > 0)/m(E_{n0})}{m(E_{n0})} = \lambda
\]

where \( E_{n0} \) is any \( E_{nk} \).

COROLLARY: The stated results hold if the condition that \( C \) 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 \( T \) with \( \mu(E) < \infty \) (e.g., \( E \) compact), and if \( N_s(E) \) denotes the number of those events in \( E \) which have "multiplicity" \( s = 1, 2, \ldots \), then

\[
p_s = \frac{E N_s(E)}{\lambda(E)}
\]

is a probability distribution on the integers \( 1, 2, \ldots \) \( \{p_s\} \) may be interpreted as the "probability that an event has multiplicity \( s \)". If in addition, \( T \) has a homogeneous dissecting system \( C = \{E_{nk}\} \) and we choose \( E \in C \) with \( \mu(E) < \infty \), writing \( \chi_{nk}^s = 1 \) if \( N(E_{nk}) = s \), \( \chi_{nk}^s = 0 \) otherwise, then similarly to Lemma 2.1

\[
\sum_{k \in I_n} \chi_{nk}^s \rightarrow N_s(E) \quad \text{as} \quad n \rightarrow \infty,
\]
with probability one. The familiar argument of taking expectations and
using dominated convergence shows that \( \lim_{n \to \infty} \Pr(N(E_{n0}) = s) = \mathbb{E}_s N(E) \)
\((E_{n0} \) is any \( E_{nk} \) and \( r_u \) the number of points in \( I_n \)). Similarly
\( \lim_{n \to \infty} \Pr(N(E_{n0}) \geq 1) = \lambda(E) \). Thus
\[
\rho_s = \lim_{n \to \infty} \Pr(N(E_{n0}) = s | N(E_{n0}) \geq 1)
\]
giving intuitive justification to the description of \( \rho_s \) as the probability
that an event has multiplicity \( s \) (under these assumptions \( \rho_s \) does not
depend on \( E \)). Further questions of this type are considered in [16] when
\( T = \mathbb{R}^n \). 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" \( \rho_0 \) gives a
precise meaning to the intuitive notion of conditional probability "given
an event of the process occurred at some point (e.g. \( t = 0 \)). When \( T \) is
the real line, we may write (for certain sets \( E \in \mathcal{F} \))
\[
(4.1) \quad \rho_0(E) = \lim_{\delta \to 0} \Pr(E | N((-\delta, 0)) \geq 1).
\]
That is \( \rho_0(E) \) is then the limit of the conditional probability of \( E \)
given an event occurred in an interval near \( t = 0 \) as that interval shrinks.
For example, if \( E \) denotes the occurrence of at least one event in the in-
terval \((0, t] \) (i.e., \( \{N(0, t) \geq 1\} \)) then \( \rho_0(E) = F_1(t) \), the distribution
function for the time to the first event after time zero given an event oc-
curred "at" time zero.
This kind of procedure for particular sets $E$ was used by Khintchine [9] and is useful in providing an "analytical" approach to such conditional probabilities (see e.g. [12]). More sophisticated and general measure theoretic treatments involving the definition and properties of $P_0$ have been given by a variety of authors (e.g. [18], [19], [14], [15], [17]). 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 (cf. also [15]) but we consider just the real line case since the later results of this section apply to the real line.

Consider, then, a stationary point process (without multiple events, for simplicity) with finite parameter $\lambda$, on the real line. Again for simplicity we take the sample points $\omega$ to be themselves the subsets $S_\omega$ of $T = \mathbb{R}$, i.e. $\omega$ is a countable set of real numbers (without finite limit points, since $\lambda < \infty$). $T$ will denote the Borel sets of $T = \mathbb{R}$ and $F$ will be the smallest $\sigma$-field on $\Omega$ making $N(B) = N_\omega(B)$ measurable for each $B \in T$. Finally, we shall again write $N(s,t)$ for $N((s,t])$, the number of events ($\text{card}\{\omega \cap (s,t)\}$) 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 $\omega$ translated to the left by $t$; i.e. if $\omega = \{t_i\}$, $\omega_t = \{t_i - t\}$. If $F$
is any set of \( F \) 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 \( \omega t_i \in F \). That is, to form \( \omega^* \) we "thin" \( \omega \) by retaining only the points \( t_i \) such that \( \omega t_i \) (i.e. \( \omega \) translated to \( t_i \) as origin) is in \( F \). The \( \omega^* \) define a stationary point process formed from some of the events in the original point process. For example, if \( F = \{ \omega: N(0, t) \geq 1 \} \), the new process contains precisely those events \( t_i \) of the old process which are followed by a further event within a further time \( t \) (i.e. no later than \( t_i + t \)). Write \( N_F \) for the number of events of the thinned process in the interval \( (0, 1) \) i.e. 
\[ \text{card}(\omega^* \cap (0, 1)) \]. Then the thinned process has intensity \( \lambda_F = EN_F \).

Now this procedure may be carried out for any \( \epsilon F \) and, for fixed \( \omega \), \( N_F \) is countably additive as a function on \( F \). It follows at once that \( \lambda_F \) is a measure on \( F \) and hence that

\[ (4.2) \quad P_0(F) = \lambda_F / \lambda \]

is a probability measure on \( F \) (\( \lambda = \lambda_\Omega = EN(0, 1) \)). \( P_0 \) is the desired Palm distribution.

To give \( P_0 \) 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 \( h \), the distance to the first one after \( t = 0 \) being 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 \( (h - \eta, \eta) \). Clearly \( \Pr(F | N(-\delta, 0) \geq 1) = 1 \) when \( \delta < \eta \). But \( N_F(0, 1) = 0 \) 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 \( F \in F \) 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 \( \chi_F(\omega_s) + \chi_F(\omega_t) \) as \( s \to t \). Equivalently, this means that for any \( t \), if \( \omega \in F \) then \( \omega \in F \) when \( s \) is sufficiently close to \( t \) on the right, and conversely.

**Theorem 4.1:** Suppose \( \mathcal{F} \in F \) is a right-continuous set. Then

\[
\Pr(F|N(-\delta,0) \geq 1) \to P_0(F) \quad \text{as} \quad \delta \to 0.
\]

**Proof.** Let \( \delta_m \) be any sequence of non-negative numbers converging to zero as \( m \to \infty \). Write \( r_m \) for the integer part \( \lfloor \delta_m^{-1} \rfloor \) of \( \delta_m^{-1} \). Divide the interval \((0,1]\) into \( r_m \) intervals of length \( \delta_m \) (with perhaps an interval of length less than \( \delta_m \) left over). Write \( \chi_{mi} = 1 \) if \( N((i-1)\delta_m, i\delta_m) \geq 1 \), \( \chi_{mi} = 0 \) otherwise \( (i=0,1, \ldots r_m) \). Let

\[
N_m = \sum_{i=1}^{r_m} \chi_{mi} \chi_F(\omega_{i\delta_m}).
\]

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 \( N_m \leq N(0,1) \), and \( EN(0,1) < \infty \), it follows by dominated convergence that \( EN_m \to \lambda_F \) as \( m \to \infty \). That is

\[
\sum_{i=1}^{r_m} \Pr(\chi_{mi} = 1, \omega_{i\delta_m} \in F) \to \lambda_F
\]
or, by stationarity,

\[ r_m \Pr(X_{m0} = 1, \omega \in F) \to \lambda F. \]

But

\[ \Pr(X_{m0} = 1, \omega \in F) = \Pr(F|N(-\delta_m,0) \geq 1) \Pr(N(-\delta_m,0) \geq 1). \]

Hence, since \( r_m \sim \delta_m^{-1}, \Pr(N(-\delta_m,0) \geq 1) \sim \lambda \delta_m, \)

\[ \Pr(F|N(-\delta_m,0) \geq 1) \to \frac{\lambda F}{\lambda} = P_0(F), \text{ as required.} \]

As an example, consider the set \( F = \{ \omega: N(0,t) \geq r \} \quad (r = 1,2,\ldots). \)

This is easily seen to be right continuous and hence the theorem applies.

In this case,

\[ P_0(F) = \lim_{\delta \to 0} \Pr(N(0,t) \geq r|N(-\delta,0) \geq 1) \]

is interpreted as the distribution function for the time to the \( r \)-th 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 \( r \)-th event after time zero does not exceed \( t \).)

Similarly, if we take \( 0 < t_1 \leq t_2 \ldots \leq t_k, \quad 0 \leq r_1 \leq r_2 \ldots \leq r_k, \)

\[ F = \{ \omega: N(0,t_1) \geq r_1, \quad N(0,t_2) \geq r_2 \ldots N(0,t_k) \geq r_k \} \]

\( F \) is right continuous, leading to what could naturally be termed the joint distribution function for the time to the \( r_1 \)-st, \( r_2 \)-nd, \ldots, \( r_k \)-th events after the origin, given an event occurred at the origin.

The convergence in Theorem 4.1 does not occur for all \( F \in F \) in general. However, we may regard the probability space \( \Omega \) 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 $F \in F$. Specifically let $s = s(\omega)$ denote the time of the first event prior to the origin. Then instead of $P_\delta(F) = P(\omega \in F \mid N(-\delta,0) > 0)$ we may consider $P^*(F) = P(\omega \in F \mid N(-\delta,0) > 0)$.

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, Sec. 1(f)]) holds.

**Theorem 4.2:** For each $F \in F$

$$P^*_\delta(F) \rightarrow P_0(F) \quad \text{as} \quad \delta \rightarrow 0$$

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

**Proof:** This can be proved as in [10, Sec. 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

$$N^*_m = \sum_{i=1}^{r_m} \chi_{m_i} \chi_F(\omega_{s_{m_i}});$$

$s_{m_i}$ 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 $t_i \epsilon(0,1)$ for which $\omega_{t_i} \in F$. It follows as before by dominated convergence and stationarity that $r_m \Pr(\chi_{m0} = 1, \omega_0 \in F) + \lambda_F$, from which the desired result follows (for any sequence $\delta_m \rightarrow 0$) on using the fact that

$$\Pr(N(-\delta_m,0) \geq 1) \sim \lambda/r_m.$$
The fact that the limit in (4.2) holds for all \( P \in \mathcal{F} \) is, of course, more satisfying than that in (4.1) which requires "continuity sets". However, the definition of \( P^*_\delta \) is more complicated than that of \( P_\delta \) and the limit in (4.1) may be more useful in practice. The difference between \( P_\delta \) and \( P^*_\delta \) is, of course, slight (but we feel it worthy of exploration).

Theorems 4.1 and 4.2 concerned conditional expectations of the function \( \chi_F \), 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]) and which is useful in a number of contexts.

**Lemma:** If \( \phi \) is a measurable function on \( \Omega \) and \( \phi \) is either non-negative, or integrable with respect to \( P_0 \), then

\[
\lambda \int \phi dP_0 = E\{\sum_{j}(\omega_{t_j}) : t_j \in \omega \cap (0,1)\}.
\]

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

The following result generalizes Theorem 4.2.

**Theorem 4.3:** Let \( \phi \) be measurable (on \( \Omega \)), continuous to the right and such that \( |\phi(\omega_t)| < \psi(\omega) \) for all \( t \in (0,1) \) where \( E(\psi N(0,1)) < \infty \). Then

\[
E(\phi|N(-\delta,0) \geq 1) \to \int \phi dP_0 \quad \text{as} \quad \delta \to 0.
\]
Proof: The pattern of the proof of Theorem 4.1 applies, with $\phi$ written for $\chi_p$. Specifically

$$S_m = \sum_{i=1}^{r_m} \chi_{m_i} \phi(\omega_{1,\delta_m}) \rightarrow \sum_{j} \phi(\omega_{t_j}): t_j \in \omega \cap (0,1)$$

with probability one. But $|S_n| \leq \psi(\omega) \cdot N(0,1)$ which has finite expectation and thus, be dominated convergence, and the lemma,

$$\int \phi dP_0 = \lambda^{-1} \lim_{m \to \infty} E_{S_m}$$

$$= \lambda^{-1} \lim_{m \to \infty} r_m E(\chi_{m_0} \phi(\omega)) \text{ (by stationarity)}$$

$$= \lim_{m \to \infty} E(\phi | \chi_{m_0} = 1)$$

since $\Pr(\chi_{m_0} = 1) \sim \lambda \delta_m \sim \lambda r_m^{-1}$. This is the desired result (writing $N(-\delta_m,0) \geq 1$ for $\chi_{m_0} = 1$).

COROLLARIES: (i) If $\phi$ is a bounded, right continuous function, the result holds. For if $|\phi| \leq K$ we may take $\psi(\omega) = K$ and $E(\psi N(0,1)) = K \lambda < \infty$.

This corollary is similar to a theorem of Ryll-Nardzewski [18] (their two-sided continuity of $\phi(\omega_{t})$ is required and the condition $N(-\delta,0) \geq 1$ replaced by $N(-\delta,\delta) \geq 1$).

(ii) Suppose $E_N^{k+1}(0,\tau) < \infty$ for some positive integer $k$, $\tau > 0$. Then

$$\lim_{\delta \to 0} E_N^{k}(0,\tau | N(-\delta,0) \geq 1) = E_{p_0} N^{k}(0,\tau)$$

where $E_{p_0}$ denotes expectation with respect to the Palm distribution. That is, the $k$-th moment of $N(0,\tau)$ with respect to the Palm distribution, is
simply the k-th conditional moment (defined as a limit) given an event "at" the origin). The proof is immediate on noting that $\phi(\omega) = N^k(0, \tau)$ is continuous to the right and, for all $t \in [0, 1]$,

$$\phi(t) = N(0, t+1) \leq N(0, 1+\tau) \Rightarrow \psi(\omega)$$

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

Finally, we note the corresponding generalization of Theorem 4.2. For this, the condition required above that $\phi$ 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 $\phi$ be measurable and such that $|\phi(t)| < \psi(\omega)$ for all $t \in (0, 1)$ where $E(\psi N(0, 1)) < \infty$. Then

$$E(\phi(t) N(-\delta, 0) \geq 1) \to \int \phi dP_0 \quad \text{as} \quad \delta \to 0$$

where $s = s(\omega)$ denotes the position of the first event prior to $t = 0$. 
REFERENCES


**REPORT TITLE**

ON BASIC RESULTS OF POINT PROCESS THEORY

**AUTHOR(S)**

Malcolm R. Leadbetter

**REPORT DATE**

June, 1970

**DISTRIBUTION STATEMENT**

The distribution of this report is unlimited

**ABSTRACT**

Basic results of point process theory (such as Khinchine's theorem, Korolyuk's theorem, Dobrushin's lemma) are discussed in a general setting. The treatment is based on an approach of Belayev, with emphasis on simplicity of proof. Palm distributions for stationary point processes on the real line are discussed (in essence along lines developed by Matthes) with particular reference to their expression as limits of conditional probabilities. A simple technique is used to prove known and obtain further results.
<table>
<thead>
<tr>
<th>KEY WORDS</th>
<th>LINK A</th>
<th>LINK B</th>
<th>LINK C</th>
</tr>
</thead>
<tbody>
<tr>
<td>STOCHASTIC PROCESSES</td>
<td>ROLE</td>
<td>WT</td>
<td>ROLE</td>
</tr>
<tr>
<td>POINT PROCESSES</td>
<td>ROLE</td>
<td>WT</td>
<td>ROLE</td>
</tr>
<tr>
<td>RANDOM STREAMS</td>
<td>ROLE</td>
<td>WT</td>
<td>ROLE</td>
</tr>
</tbody>
</table>