AN ABSTRACT OF THE THESIS OF

Marie M. Franzosa for the degree of Doctor of Philosophy in Mathematics presented on January 28, 1988

Title: Densities and Dependence for Point Processes

Abstract approved: Signature redacted for privacy.

Robert M. Burton

Product densities have been widely used in the literature to give a concrete description of the distribution of a point process. A rigorous description of properties of product densities is presented with examples to show that in some sense these results are the best possible. Product densities are then used to discuss positive dependence properties of point processes.

There are many ways of describing positive dependence. Two well known notions for Bernoulli random variables are the strong FKG inequalities and association, the strong FKG inequalities being much stronger. It is known, for example, from van den Berg and Burton, that the strong FKG inequalities are equivalent to all conditional distributions being associated, which is equivalent to all conditional distributions being positively correlated. In the case of point processes for which product densities exist, analogs of such positive dependence properties are given. Examples are presented to show that unlike the Bernoulli case none of these conditions are equivalent, although some are shown to be implied by others.

Densities and Dependence
for Point Processes

by

Marie M. Franzosa

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed January 28, 1988

Commencement June 1988

APPROVED:

Signature redacted for privacy.

Associate Professor of Mathematics in charge of major

Signature redacted for privacy.

Chair of department of Mathematics

Signature redacted for privacy.

Dean of Graduate School

Date thesis is presented January 28, 1988

ACKNOWLEDGEMENTS

This work would not have been possible without the support and guidance of my major professor, Robert M. Burton. I am grateful to him for his encouragement and friendship. I also wish extend my thanks John Lee, Bent Petersen and Ed Waymire for their help during my years as a graduate student, and to the many friends and fellow graduate students who have shared this experience with me.

Special thanks go to the members of my family, especially to my brother Bob for encouraging my early interest in mathematics and to my parents Alfred and Dorothea, who encouraged growth and learning for all of their children. Lastly I wish to thank my husband, Dennis Garity, for his love, support and patience.

This thesis is dedicated to the memory of my father, who would have been very proud to see me get this far.

TABLE OF CONTENTS

Chapter		page
I.	INTRODUCTION	1
	1.1 Introduction	1
	1.2 Positive Dependence for 0-1 Valued Random Variables	2
II.	POINT PROCESSES	9
	2.1 Preliminaries for the Point Process Case	9
	2.2 Product Densities	11
	2.3 Conditional Densities and Cumulant Densities	31
	2.4 Generating Functionals	32
	2.5 Examples of Point Processes	35
III.	POSITIVE DEPENDENCE FOR POINT PROCESSES	40
	3.1 Definitions of Positive Dependence	40
	3.2 The Relationships Between Positive Dependence Definitions	48
	3.3 Proofs and Examples	49
	BIBLIOGRAPHY	71
	APPENDICES	75
	A.1 Weak Convergence (Convergence in Distribution) of Random Elements in a Metric Space	75
	A.2 Weak Convergence and convergence in Distribution for Random Measures	77

DENSITIES AND DEPENDENCE FOR POINT PROCESSES

I. INTRODUCTION

1.1 Introduction

Product densities have been widely used in the literature to give a concrete description of the distribution of a point process (See for example, Moyal [23], Srinivasan [32], or Karr [17]). The first attempt to rigorously discuss the properties of product densities was made by Macchi [21]. In chapter II we expand the results given by Macchi to a more general setting and provide more rigorous arguments for some of the claims made in her paper. Examples are then constructed to show that these results are in some sense the best possible. Examples of many commonly occurring point processes are also discussed and the densities are computed for some of these processes.

Chapters I and III deal with the concept of positive dependence. Roughly put, a collection of random variables satisfies a positive dependence property if knowledge that the variables are "large" (or small) in some sense does not decrease the likelihood that they are also "large" (resp. small) in some other sense. In the case of point processes, several notions of positive dependence can be described in terms of the product densities and absolute product densities.

Positive dependence notions arose independently in several fields. In reliability theory one can give estimates on the reliability of a system based on the assumption that the breakdown of certain components can not decrease the probability that other components will also fail (Barlow and Proschan [1]). Another field in which positive dependence has played a role is statistical physics. In an attractive statistical mechanical system the occurrence of a large number of "plus" sites makes it more likely that there will be "plus" sites elsewhere (see Kinderman and Snell [19]). Newman's Central Limit Theorem [25] also depends on positive dependence properties. This theorem was used by Newman in [24] to analyze Ising model magnetization fluctuations and by Newman and Schulman [26] to analyze density fluctuations of infinite clusters in percolation models.

The relationships between various positive dependence properties for random variables have been explored by Barlow and Proschan [1] and also, for the specific case of Bernoulli random variables, by van den Berg and Burton [2]. These relationships are presented in chapter I. In chapter III we extend this exploration to the case of point processes and show by examples that the analogous theorem does not hold.

1.2 Positive Dependence for 0 - 1 Valued Random Variables

We begin by considering positive dependence properties of Bernoulli random variables. Let P be a probability measure on $\Omega=\{0,1\}^n$ with the σ-algebra generated by points. Ω is a distributive lattice with the coordinate-wise ordering given by $\alpha\wedge\beta=(\min(\alpha_1,\beta_1),\ldots,\min(\alpha_n,\beta_n))$ and $\alpha\vee\beta=(\max(\alpha_1,\beta_1),\ldots,\max(\alpha_n,\beta_n))$ for $\alpha,\beta\in\Omega$. Let $\underline{X}$ = $(X_1, \cdots X_N)$ where $X_i\in\{0,1\}$ for each i = 1, 2, $\cdots$ N. Positive dependence of such random variables can be described in many ways. The following definitions give a few such descriptions.

Definition 1.1 $\underline{X}$ satisfies the strong FKG inequalities (named after Fortuin, Kastelyn, and Ginibre) if

$$P(\underline{X} = \alpha \wedge \beta)\, P(\underline{X} = \alpha \vee \beta) \geq P(\underline{X} = \alpha)\, P(\underline{X} = \beta) \text{ for all } \alpha,\beta \in \{0,1\}^N .$$

Intuitively this condition means that the string $(X_1,\cdots,X_N)$ is more likely to contain a majority of zeroes or of ones than to contain an even mixture of each. Such a string is "positively dependent" in that knowing one part of the string contains many ones (for example) increases the likelihood that there are ones elsewhere in the string. The next three definitions can be found in Barlow and Proschan [1].

Definition 1.2 (a) A random variable Y is stochastically increasing in the random variables $X_1,\cdots,X_N$ if $P(Y > y \mid X_1 = x_1,\cdots,X_N = x_N)$ is increasing in the variables $x_1,\cdots,x_N$. We will denote this by $Y \uparrow st$ in $X_1,\cdots,X_N$.

Note that if Y is a Bernoulli random variable the above condition is equivalent to requiring that $P(Y=1 \mid X_1 = x_1,\cdots,X_N = x_N)$ be increasing in the variables $x_1,\cdots,x_N$. Intuitively this condition says that if some of the random variables $X_1,\cdots,X_N$ are ones then Y is more likely to also be one.

Definition 1.2 (b) $\underline{X}$ is conditionally increasing in sequence ($\underline{X}$ is CIS) if $X_n \uparrow st$ in $X_1,\cdots,X_{n-1}$ for $n = 1,\cdots,N$.

Definition 1.3 $\underline{X}$ has positively correlated increasing cylinder sets ($\underline{X}$ is PCIC) if when I, K are disjoint subsets of $\{1,\cdots,N\}$ and $A_I = \{X_i = 1 \ \forall\ i \in I\}$ then $P(A_I \cap A_K) \geq P(A_I)P(A_K)$.

Like strong FKG, definition 1.3 implies that $\underline{X}$ is more likely to contain a lot of ones. It does not say, however, what is likely to be seen at locations with indices outside of the designated sets.

Definition 1.4 $\underline{X}$ is positively correlated ($\underline{X}$ is PC) if

$$P(X_i=1,\ X_j=1) \geq P(X_i = 1)P(X_j = 1) \text{ for all } i,\ j \in \{ 1, \cdots N \}.$$

Clearly if $\underline{X}$ has PCIC, then $\underline{X}$ is positively correlated.

Definition 1.5 $\underline{X}$ is associated (or, $X_1,\cdots,X_N$ is an associated collection of random variables) if for all pairs of nondecreasing, real valued functions f and g, $\mathrm{Cov}(f(\underline{X}),g(\underline{X})) \geq 0$, as long as this covariance exists.

The intuitive interpretation of association is more clear if we first consider the following theorem giving an equivalent definition, and a lemma giving an alternative formula for calculating covariances.

Theorem 1.6 If $\mathrm{Cov}(f(\underline{X}),g(\underline{X})) \geq 0$ for all pairs of nondecreasing binary functions f and g then $\underline{X}$ is associated.

Proof omitted (see Esary, Proschan and Walkup [9]).

The following lemma is due to Hoeffding [14]. The proof given here appears in Lehmann [20].

Lemma 1.7 $\mathrm{Cov}(f(\underline{X}),g(\underline{X})) =$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(f(\underline{X}) > u, g(\underline{X}) > v) - P(f(\underline{X}) > u)P(g(\underline{X}) > v)\,du\,dv, \quad (1.1)$$

provided the covariance exists.

Proof Let $U = f(\underline{X})$ and $V = g(\underline{X})$. Let (U_1,V_1) and (U_2,V_2) be independent and each distributed as (U,V), then

$$2\,\mathrm{Cov}(U,V) = 2\,\mathrm{Cov}(U_1,V_1) = 2(E(U_1V_1) - E(U_1)E(V_1)) = E((U_1 - U_2)(V_1 - V_2))$$

$$= E\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [1_{\{u \le U_1\}} - 1_{\{u \le U_2\}}][1_{\{v \le V_1\}} - 1_{\{v \le V_2\}}]\,du\,dv \tag{1.2}$$

where $1_{\{u \le U\}}$ is the indicator function of the event $\{u \le U\}$. That is,

$$1_{\{u \le U\}} = \begin{cases} 1 \text{ if } U > u \\ 0 \text{ if } U \le u \end{cases} \tag{1.3}$$

Since Cov(U,V) exists, $E|UV|$, $E|U|$ and $E|V|$ are finite, so we may take the expectation inside of the integral signs to get that (1.2) is

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\,[1_{\{u \le U_1\}} - 1_{\{u \le U_2\}}][1_{\{v \le V_1\}} - 1_{\{v \le V_2\}}]\,du\,dv$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\,[1_{\{u \le U_1\}}1_{\{v \le V_1\}}] - E[1_{\{v \le V_1\}}1_{\{u \le U_2\}}]$$

$$- E[1_{\{v \le V_2\}}1_{\{u \le U_1\}}] + E[1_{\{u \le U_2\}}1_{\{v \le V_2\}}]\,du\,dv$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(U_1 \ge u, V_1 \ge v) - P(U_2 \ge u, V_1 \ge v)$$

$$- P(U_1 \ge u, V_2 \ge v) + P(U_2 \ge u, V_2 \ge v)\,du\,dv$$

$$= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(U \ge u, V \ge v) - P(U \ge u)P(V \ge v)\,du\,dv \qquad \square$$

Note that Theorem 1.6 and Lemma 1.7 imply that $\underline{X}$ is associated if and only if

$$P(f(\underline{X}) > u \mid g(\underline{X}) > v) \ge P(f(\underline{X}) > u) \tag{1.4}$$

for all nondecreasing binary f and g and all real numbers u and v. The lemma implies $P(f(\underline{X}) > u,\ g(\underline{X}) > v) \ge P(f(\underline{X}) > u)P(g(\underline{X}) > v)$ which implies that the integral in (1.2) is nonnegative. Conversely, if $\mathrm{Cov}(f(\underline{X}), g(\underline{X})) \ge 0$ for all nondecreasing binary f and g, then Cov ($1_{\{f(\underline{X})>u\}}, 1_{\{g(\underline{X})>v\}}$) ≥ 0 where $1_{\{f(\underline{X})>u\}}$ is the indicator function of the event $\{f(\underline{X})>u\}$. Thus, $E(1_{\{f(\underline{X})>u\}}\ 1_{\{g(\underline{X})>v\}}) \geq E(1_{\{f(\underline{X})>u\}})E(1_{\{g(\underline{X})>v\}})$ which is equivalent to (1.4).

From (1.4) the intuitive interpretation of association is that $\underline{X}$ is associated if whenever we know that one type of nondecreasing measurement of $\underline{X}$ is large the likelihood that another type of nondecreasing measurement is also large is increased.

It is well known that for two binary random variables definitions 1.1 —1.5 are equivalent. For the case of n binary random variables, where $n>2$ definitions 1.1 through 1.5 are related by the following theorem, the proof of which can be found in Barlow and Proschan [1].

Theorem 1.8. $\underline{X}$ satisfies the strong FKG inequalities $\Rightarrow$ $\underline{X}$ is CIS $\Rightarrow$ $\underline{X}$ is associated $\Rightarrow$ $\underline{X}$ has PCIC $\Rightarrow$ $\underline{X}$ is PC.

Theorem 1.8 is actually true for more general random variables, i.e. not only for binary random variables. That strong FKG implies association is a consequence of the FKG paper [12]. That strong FKG implies CIS is automatic from the definition and that CIS implies associated appears in Barlow and Proschan [1]. The other implications are automatic.

None of the reverse implications in Theorem 1.8 hold. Van den Berg and Burton [2] have shown, however, that the above definitions can be modified so that they become equivalent.

Definition 1.10 $\underline{X}$ is conditionally associated if for each $J \subseteq \{1,\cdots,N\}$ and $\alpha_0 \in \{0,1\}^N$ dist($\underline{X} \mid X_j = (\alpha_0)_j \ \forall\ j \in J$) is associated.

Definition 1.11 $\underline{X}$ has conditionally positively correlated increasing cylinder sets ($\underline{X}$ has CPCIC) when if I, J, K are all subsets of $\{1,\cdots,N\}$, $\alpha_0 \in \{0,1\}^N$ and if A_I is the event that $X_i = 1$ for all $i \in I$ then

$$P(A_I \cap A_K \mid X_j = (\alpha_0)_j \ \forall\ j \in J) \geq$$
$$P(A_I \mid X_j = (\alpha_0)_j \ \forall\ j \in J) P(A_K \mid X_j = (\alpha_0)_j \ \forall\ j \in J)$$

Similarly, we can define conditionally positively correlated (CPC)

Theorem 1.12 (van den Berg and Burton) If P assigns positive probability to each outcome in $\{0,1\}^n$ then the following are equivalent

a. $\underline{X}$ satisfies the strong FKG inequalities.

b. $\underline{X}$ is conditionally increasing in sequence (CIS) in all orderings. That is, if σ is any permutation of $\{1, 2, \cdots N\}$ and $Y_i = X_{\sigma(i)}$ then $\underline{Y} = (Y_1, \cdots Y_N)$ is CIS.

c. $\underline{X}$ is conditionally associated.

d. $\underline{X}$ has conditionally positively correlated increasing cylinder sets.

e. $\underline{X}$ is conditionally positively correlated (CPC).

Theorem1.12 was essentially proved by J. Kemperman [18] in a different context. See also Perlman and Olkin [29].

That the conditions given in Theorem 1.8 are not equivalent is easily shown by the following example:

Let $\underline{X} = (X_1, X_2, X_3)$ where X_1, X_2, X_3 are distributed so that the outcomes of $\underline{X}$ have the following probabilities:

outcome	probability	outcome	probability
(1,1,0)	$\frac{1}{12}$	(0,1,1)	$\frac{1}{6}$
(1,1,1)	$\frac{1}{6}$	(0,0,0)	$\frac{3}{16}$
(1,0,0)	$\frac{1}{8}$	(0,0,1)	$\frac{1}{16}$
(1,0,1)	$\frac{1}{8}$		
(0,1,0)	$\frac{1}{12}$		

(1.5)

Then, $P(X_2=1 \mid X_1=0)=\frac{1}{2}$, $P(X_2=1 \mid X_1=1)=\frac{1}{2}$, $P(X_3=1 \mid X_2=0, X_1=0)=\frac{1}{4}$, $P(X_3=1 \mid X_2=1, X_1=0)=\frac{2}{3}$, $P(X_3=1 \mid X_2=0, X_1=1)=\frac{1}{2}$, $P(X_3=1 \mid X_2=1, X_1=1)=\frac{2}{3}$. So $\underline{X}$ is CIS and thus also associated. Now reverse the order, that is let $\underline{Y} = (Y_1,Y_2,Y_3)=(X_1,X_2,X_3)$. $\underline{Y}$ is also associated, since the property of association does not depend on order, but

$$P(Y_3=1 \mid Y_2=0, Y_1=1) = P(X_1=1 \mid X_2=0, X_3=1) = \frac{P(X_1=1, X_2=0, X_3=1)}{P(X_2=0, X_3=1)} = \frac{\frac{1}{8}}{\frac{1}{8}+\frac{1}{16}} = \frac{2}{3}$$

whereas,

$$P(Y_3=1 \mid Y_2=1, Y_1=1) = P(X_1=1 \mid X_2=1, X_3=1) = \frac{P(X_1=1, X_2=0, X_3=1)}{P(X_2=0, X_3=1)} = \frac{\frac{1}{8}}{\frac{1}{8}+\frac{1}{16}} = \frac{1}{2}$$

so that $\underline{Y}$ is not CIS even though it is associated. $\underline{Y}$ is also not FKG since it is not CIS in all orderings.

Van den Berg and Burton also showed that even if all configurations do not have positive probability we still have most of Theorem 1.12.

Theorem 1.13 Suppose that $\Omega = \{0,1\}^n$, as above. Then the following are equivalent:

a. $\underline{X}$ satisfies the strong FKG inequalities.

b. $\underline{X}$ is conditionally associated.

c. $\underline{X}$ has conditionally positively correlated increasing cylinder sets (CPCIC).

II. POINT PROCESSES

2.1 Preliminaries For the Point Process Case

Let $\mathbb{R}^d$ be d- dimensional Euclidean space and $D \subseteq \mathbb{R}^d$ a fixed, possibly infinite, subrectangle. Let $\mathfrak{B}^d$ be the collection of Borel subsets of D. Denote the subset of $\mathfrak{B}^d$ consisting of bounded sets (i.e. sets with compact closures) by $\hat{\mathfrak{B}}^d$. A measure μ on $(D,\mathfrak{B}^d)$ is called Radon if $\mu(B) < \infty$ for all sets $B \in \hat{\mathfrak{B}}^d$. Let M denote the set of all Radon measures on $(D,\mathfrak{B}^d)$ and N the subset of M consisting of counting measures. Thus, $\mu \in N$ if and only if $\mu(B) \in \mathbb{Z}^+ = \{0,1,2,\cdots\}$ and $\mu(B) < \infty$ for all $B \in \hat{\mathfrak{B}}^d$. N is naturally identified with the set of all finite or infinite configurations of points (including multiplicities) in D without limit points.

Let $\mathcal{M}$ be the σ-algebra on M generated by sets of the form $\{\mu \in M \mid \mu(A) < k\}$ for all $A \in \hat{\mathfrak{B}}^d$ and $0 \leq k < \infty$. Likewise $\mathcal{N}$ is the σ-algebra generated by such sets of measures in N. Note that $\mathcal{N} \subset \mathcal{M}$ and so $\mathcal{N}$ is the restriction of $\mathcal{M}$ to N (see Kallenberg [15] for details). $\mathcal{N}$ is the σ-algebra on N which allows us to count the points in bounded regions of D.

The vague topology on M (or N) is the topology generated by the class of all finite intersections of subsets of M (resp. N) of the form $\{\mu \in M \mid s \leq \int f\, d\mu \leq t\}$ for all functions $f \in \mathfrak{F}_C$ (where $\mathfrak{F}_C = \{f: \mathbb{R}^d \to \mathbb{R}^+ = [0,\infty] \mid f$ is continuous and has compact support$\}$) and $s, t \in \mathbb{R}$.

Both M and N are metrizable as complete separable metric spaces (i.e. M and N are Polish spaces) in the vague topology and $\mathcal{M}$ and $\mathcal{N}$ are the Borel sets generated by these topologies [15].

Convergence in the vague topology is defined as follows. μ_n converges to μ in the vague topology, written $\mu_n \xrightarrow{v} \mu$ if $\int f\, d\mu_n \to \int f\, d\mu$ for every $f \in \mathcal{F}_C$.

Further details on the structure of these spaces and on convergence of random measures are given in the appendix.

Definition 2.1 A point process is a measurable mapping X from a probability space $(\Omega,\mathcal{F},P)$ into $(N,\mathcal{N})$. The distribution of X is the induced measure on $(N,\mathcal{N})$ given by $P_X = PX^{-1}$.

Thus if $A \in \mathcal{B}^d$ we set X(A) equal to the (random) number of occurrences in A.

Definition 2.2 Define the translation operator $T_x: N \to N$ for $x \in \mathbb{R}^d$, $w = \{\delta_{x_i}\} \in N$ by $T_x w = \{\delta_{x_i + x}\}$. For $A \in \mathcal{N}$ let $T_x(A)$ denote the set $\{T_x w \mid w \in A\}$. X (or its distribution P_X) is called stationary if for every $x \in \mathbb{R}^d$, T_x is P_X invariant. That is, $P_X(T_x(A)) = P_X(A)$ for every $x \in \mathbb{R}^d$ and $A \in \mathcal{N}$.

Example 2.3 The most fundamental point process is the Poisson point process. Given a Radon measure Λ on $\mathbb{R}^d$, a Poisson point process X with intensity Λ is a point process such that

$$P(X(B_1) = k_1, \ldots, X(B_n) = k_n) = \frac{(\Lambda(B_1))^{k_1}}{k_1!}\, e^{-\Lambda(B_1)} \cdots \frac{(\Lambda(B_n))^{k_n}}{k_n!}\, e^{-\Lambda(B_n)} \qquad (2.1)$$

for all $k_1, \ldots, k_n \in \mathbb{Z}^+$ and $B_1, \ldots, B_n$ disjoint elements of $\mathcal{B}^d$. That is, a Poisson process is defined for the measure Λ on $\mathbb{R}^d$ by the following two conditions (see, for example, Daley and Vere-Jones [8])

(1) For every $B \in \mathfrak{B}^d$, $X(B)$ has a Poisson distribution with mean $\Lambda(B)$.

(2) For any finite family of mutually disjoint bounded sets $B_i \in \mathfrak{B}^d$ the random variables $X(B_i)$ are mutually independent.

When the measure Λ is taken to be a multiple of Lebesgue measure, i.e. $\Lambda(A) = \lambda|A|$ for some $0 < \lambda < \infty$, we obtain a Stationary Poisson Process.

An equivalent characterization of the stationary Poisson process is the Law of Rare Events (See Billingsley [4]):

Theorem 2.4 Suppose that X is a point process satisfying the following conditions:

(1) (completely random) $X(B_1),\ldots,X(B_n)$ are independent for disjoint Borel sets $B_1,\ldots,B_n$.

(2) (without multiple occurrences) $P(X(A) \geq 2) = o(|A|)$ as $|A| \to 0$.

(3) (homogeneous) $P(X(A) = 1) = \lambda|A| + o(|A|)$ as $|A| \to 0$, where $0 < \lambda \leq \infty$ and $|A|$ is the Lebesgue measure of A.

Then X is the stationary Poisson point process with intensity $\Lambda(A) = \lambda|A|$.

2.2 Product Densities

We now wish to define densities for point processes. That is, for each n, $n = 1,2,\ldots$ we want functions $p(x_1 \ldots x_n)$ such that $p(x_1 \ldots x_n)|\Delta x_1| \ldots |\Delta x_n|$ approximates the probability of points occurring in the intervals $\Delta x_1,\ldots,\Delta x_n$ about $x_1,\ldots,x_n$ when the Lebesgue measure $|\Delta x_i|$ is sufficiently small.

The description of such densities will depend on certain regularity conditions. Such conditions rule out the possibility of multiple occurrences in X. Much of what follows is based on ideas presented in Macchi [21] and Fisher [11].

Let $\mathcal{N}_0 = \{ \mu \in \mathcal{N} : \mu(A) = \sum_{i \in I} \delta_{x_i}(A) \}$ for each set in D where I is a subset of $\mathbb{Z}^+$ (I may be either finite or countable) , $x_i \in \mathbb{R}^d$ and

$$\delta_{x_i} = \begin{cases} 1 \text{ if } x_i \in A \\ 0 \text{ if } x_i \notin A \end{cases} \tag{2.2}$$

Thus $\mathcal{N}_0$ consists of all outcomes μ which have no multiple occurrences.

Definition 2.5 X is called almost surely orderly (a.s. orderly) if $P(X \in \mathcal{N}_0) = 1$.

Definition 2.6 X is called analytically orderly if for each $x \in D$ $P(X(\Delta x) > 1) = o(|\Delta x|)$ as the d-dimensional rectangles Δx, members of a fixed substantial family, decrease in Lebesgue measure to x.

Note that it is not necessary for the rectangles Δx in our definition to be centered at x. By "substantial family" we are following the terminology of Rudin [31]. That is, a collection $\mathcal{S}$ of open sets in $\mathbb{R}^d$ is called a substantial family if

(1) there is a constant $\infty > \beta > 0$ such that each $E \in \mathcal{S}$ lies in an open ball B with $|B| < \beta |E|$ and

(2) for each $x \in \mathbb{R}^d$ and $\delta > 0$ there is an $E \in \mathcal{S}$ with $x \in E$ and diameter of $E = \sup\{|x-y| : x \in E, y \in E\}$ less than δ.

Example 2.7 An example of a substantial family is $\mathcal{S}_\alpha$ ={rectangular boxes with side lengths $s_1, \ldots, s_d$ such that there exists $0 < \alpha < 1$ with $\frac{s_i}{s_j} > \alpha$

for each $i \neq j\}$. The Lebesgue measure of a set $E \in \mathcal{S}$ is given by $|E| = \prod_{i=1}^{d} s_i$.

Let $s_{i_0} = \max_{i=1,\ldots,d} s_i$. Then $|E| = \prod_{i=1}^{d} s_i = (s_{i_0})^d \cdot \prod_{i=1}^{d} \frac{s_i}{s_j} > \alpha^d (s_{i_0})^d = \alpha^d \cdot c\,|B_0|$ where c is a constant (depending only on d) and B_0 is the ball circumscribed about the cube with side lengths s_{i_0}. If B is the smallest ball enclosing E then $B \subset B_0$, so $\alpha^d \cdot c\,|B_0| \geq \alpha^d \cdot c\,|B|$. Thus $\mathcal{S}_\alpha$ satisfies condition (1) with $\beta = \alpha^d c$. Since we are considering rectangles of arbitrarily small size it is clear that $\mathcal{S}_\alpha$ satisfies condition (2).

It is easy to see that analytically orderly implies a.s. orderly, although the converse is not true.

Example 2.8 Let Λ be a random variable on $(0,\infty)$ with density function given by $f(\lambda) = \frac{1}{\lambda^2}$ for $\lambda \geq 1$ and $f(\lambda) = 0$ for $0 \leq \lambda < 1$. Given that $\Lambda = \lambda$, let X be a Poisson process with intensity λ. X is an example of a mixed Poisson process. Such processes are considered again in Example 2.32.

Since X is Poisson, it is clearly a.s. orderly however,

$$P(X(\Delta x)>1) \geq P(X(\Delta x)=2) = \int_0^\infty \frac{\lambda^2(|\Delta x|)^2}{2} e^{-\lambda(|\Delta x|)} f(\lambda)\, d\lambda$$

$$= \frac{(|\Delta x|)^2}{2} \int_1^\infty e^{-\lambda(|\Delta x|)}\, d\lambda$$

$$= \frac{(|\Delta x|)^2}{2} \left\{ \frac{-1}{|\Delta x|} e^{-\lambda(|\Delta x|)} \Big|_1^\infty \right\} = \frac{|\Delta x|}{2} e^{-|\Delta x|}.$$

Thus as $|\Delta x|$ approaches 0, $\frac{1}{|\Delta x|} P(X(\Delta x)>1) \geq e^{-|\Delta x|}$, which approaches 1. That is, $P(X(\Delta x)>1)$ is not $o(|\Delta x|)$ as Δx approaches 0, and so X is not analytically orderly. This example is not so nice in the sense that $E(\Lambda) = E(X(\Delta x)) = \infty$. A similar example, for which X turns out to be stationary, follows.

Example 2.9 This example was presented by E. Waymire in seminar notes. Let X_1 be distributed on (0,1) with distribution function given by $F(x) = \sqrt{x}$. Given $X_1 = x_1$, let X_2 be uniformly distributed on $(0,x_1)$. Then let X be the point process given by $X = \sum_{-\infty}^{\infty} \delta_{X_2 + nX_1}$ where δ is the dirac delta function, $\delta_X(A) = 1$ if $x \in A$, 0 if $x \notin A$. X is then a periodic point process with phase X_2 and random period X_1. In fact X is a mixture of stationary periodic point processes and so is itself stationary (although we still have that $E(X(\Delta x)) = \infty$).

X is a.s. orderly but $P(X(0,x] \geq 2) \geq P(X_1$ is in the interval $(0,\frac{1}{2}x])$ since if $X_1 \in (0,\frac{1}{2}x]$ then $X_2 \in (0,\frac{1}{2}]$ and $X_2 + X_1 \leq \frac{1}{2}x + \frac{1}{2}x = x$ so $X_2 + X_1 \in (0,\frac{1}{2}]$ but $P(X_1 \in (0,\frac{1}{2}x]) = P(X_1 \leq \frac{1}{2}x) = F(\frac{1}{2}x) = \frac{\sqrt{2}}{2}\sqrt{x}$. If we let $(0,x] = \Delta x$ then as $|\Delta x|$ approaches 0, $\frac{1}{|\Delta x|}P(X(0,x] \geq 2) \geq \frac{1}{|\Delta x|}\frac{\sqrt{2}}{2}\sqrt{x} = \frac{\sqrt{2}}{2\sqrt{x}}$ which approaches ∞ as x approaches 0. Thus $P(X(\Delta x)>1)$ is not $o(|\Delta x|)$ as Δx approaches 0, and so, again, X is not analytically orderly.

Given a compact subset A of D and disjoint Borel subsets $A_1,\ldots,A_n$ of A let $R_A^n(A_1,\ldots,A_n)$ be $\frac{1}{n!}$ times the probability of exactly one point occurring in each of the sets $A_1,\ldots,A_n$ and no other points occurring in A.

Definition 2.10 A point process is called semi-regular if all R_A^n are absolutely continuous with respect to Lebesgue measure on $A^n = A \times \ldots \times A$

Thus, by the Radon-Nikodym Theorem, for a semi-regular point process we can write

$$n!\ R_A^n(x_1,\ldots,x_n) = \int_A \cdots \int_{A^n} r_A^n(x_1,\ldots,x_n)dx_1 \ldots dx_n. \tag{2.3}$$

So that $r_A^n(x_1,\ldots,x_n)|\Delta x_1|\ldots|\Delta x_n|$ has the interpretation as an approximation to the probability that X has exactly n occurrences in A, exactly one in each of the regions Δx_i. That is, the probability that there are n occurrences in A, one in each region Δx_i is $r_A^n(x_1,\dots,x_n)|\Delta x_1|\dots|\Delta x_n| + o(\max_i |\Delta x_i|)$. Note that the functions r_A^n are unique up to a.e. equivalence classes.

Definition 2.11 We will refer to the Radon-Nikodym derivatives r_A^n as the absolute product densities of X.

Definition 2.12 Assuming that the expectations given below are finite for bounded A_i,

(a) The n^{th} order moment measure is the measure on the product space $A \times \dots \times A$ (n-fold product) given by
$M_n(A_1 \times \dots \times A_n) = E(X(A_1)\dots X(A_n))$

(b) The n^{th} order factorial moment measure is given by
$M_{[n]}(A_1^{t_1} \times \dots \times A_k^{t_k}) = E(X(A_1)^{[t_1]}\dots X(A_k)^{[t_k]})$ where $t_1 + \dots + t_k = n$ and $s^{[t]} = s(s-1)\dots(s-t+1)$.

That M_n and $M_{[n]}$ really are measures follows from the fact that X is non-negative and σ-additive, thus M_1 is a measure. Extending this to measures on product spaces gives that M_n is a measure.

Note that $M_n(A_1 \times \dots \times A_n) = M_{[n]}(A_1 \times \dots \times A_n)$ for disjoint $A_1,\dots,A_n$. We will drop the subscript n or [n] whenever it is clear which measure is meant. The following theorem was stated by Macchi in [21] without proof.

Theorem 2.13 If X is a semi-regular, almost surely orderly point process with absolute product densities $r_A^n(x_1,\dots,x_n)$ then for $A_1,\dots,A_n$ disjoint

$$M(A_1 \times \dots \times A_n) = \int_{A_1} \dots \int_{A_n} p_n(x_1,\dots,x_n)\,dx_1\dots dx_n \qquad (2.4)$$

where

$$p_n(x_1,\ldots,x_n) = \sum_{j=0}^{\infty} \frac{1}{j!} \int_{A^j} r_A^{n+j}(x_1,\ldots,x_n,\theta_1,\ldots,\theta_j)\,d\theta_1\ldots d\theta_j \tag{2.5}$$

Proof: Let $A_1,\ldots,A_n$ be disjoint d-dimensional rectangles. By definition, $M(A_1\times\ldots\times A_n) = E(X(A_1)\ \ldots\ X(A_n)) =$ $\sum_{k_0,k_1,\ldots,k_n=0}^{\infty} k_1\ldots k_n\, P(A_0,k_0;\ldots;A_n,k_n)$ where $P(A_0,k_0;\ldots;A_n,k_n)$ represents the probability that there are k_0 points in A_0, k_1 in $A_1,\ldots,k_n$ in A_n and $A_0 = A\setminus\bigcup_{i=1}^{n} A_i$.

Fix $\epsilon > 0$ and much smaller than $P(A_0,k_0;\ldots;A_n,k_n)$. Given $X\in\mathcal{N}$, let $S(X)$ be the minimum distance between the point occurrences of X. Since X is a.s. orderly $S(X) > 0$ a.s. Set $E_\nu=\{X\in\mathcal{N}:\ S(X) > \frac{1}{\nu}\}$. Then $E_1\subset E_2\subset\ldots$ and $\bigcup_{\nu=1}^{\infty} E_\nu = \mathcal{N}$ almost surely, so $\lim_{\nu\to\infty} P(E_\nu) = 1$. Thus, there exists a ν_0 so that for $\nu\geq\nu_0$, $P(E_\nu) > 1-\epsilon$. That is, all points (with the exception of an event of probability less than ϵ) are farther apart than $\frac{1}{\nu}$ units.

For each $\nu\geq\nu_0$ partition each set A_j by a partition $\mathcal{P}_j(\nu)$ consisting of $T_j(\nu)$ subsets $\Delta_j^{(i)}(\nu)$ $i=1,\ldots,T_j(\nu)$ where each subset has diameter less than $\frac{1}{\nu}$ and belongs to a fixed substantial family $\mathcal{S}$. (In fact we may take the partition elements to be elements of a substantial family of the form $\mathcal{S}_\alpha$ as described in Example 2.7, so that each $\Delta_j^{(i)}(\nu)$ is actually a rectangle). Then $\mathcal{P}(\nu) = \bigcup_{j=0}^{n} \mathcal{P}_j(\nu)$, the collection of all subrectangles $\Delta_j^{(i)}(\nu)$, forms a partition of the set $\bigcup_{j=0}^{n} A_j$. For the remainder of the proof will omit the (ν) in order to simplify the notation. It is to be understood however that each partition and partition element depends on ν.

For $\nu\geq\nu_0$ let B be the event that there are k_0 points in A_0, k_1 in

$A_1,\ldots,k_n$ in A_n and C the event that there are k_0 points in A_0, k_1 in $A_1,\ldots,k_n$ in A_n and each point occurs in a different subrectangle $\Delta_j^{(i)}$. Then $C \subset B$ and $B\backslash C$ is the event that there are k_0 points in A_0, k_1 in $A_1,\ldots,k_n$ in A_n and at least one subrectangle contains more than one point. $B\backslash C \subset E_\nu^{\,c}$, the complement of E_ν, and so $P(B\backslash C) < \epsilon$. So

$P(A_0,k_0;\ldots;A_n,k_n) =$

$$\sum P_k(\Delta_0^{(1)},\ldots,\Delta_0^{(k_0)},\Delta_1^{(1)},\ldots,\Delta_1^{(k_1)},\ldots,\Delta_n^{(1)},\ldots,\Delta_n^{(k_n)}) + P(B\backslash C) \qquad (2.6)$$

where the sum is taken over the $\prod_{j=0}^{n}\binom{T_j}{k_j}$ possible ways of choosing k_0 of the rectangles $\Delta_0^{(i)}$ $i=1,\ldots,T_0$, k_1 of the $\Delta_1^{(i)}$ $i=1,\ldots,T_1$, etc. and $P_k(\Delta_0^{(1)}\ldots,\Delta_n^{(k_n)})$ where $k=k_0+\ldots+k_n$ denotes the probability that exactly k points occur, one in each of the listed sets.

By symmetry (2.6) becomes

$$\sum_{\substack{i_1^n,\ldots,i_{k_n}^n=1 \\ i_j^n \neq i_l^n}}^{T_n} \cdots \sum_{\substack{i_1^0,\ldots,i_{k_0}^0=1 \\ i_j^0 \neq i_l^0}}^{T_n} \frac{1}{k_0!}\cdots\frac{1}{k_n!}\,P_k^* \;+\; P(B\backslash C) \qquad (2.7)$$

Where $P_k^* = P_k(\Delta_0^{(i_1^0)},\ldots,\ \Delta_0^{(i_{k_0}^0)}\ldots\ \Delta_n^{(i_1^n)},\ldots\Delta_n^{(i_{k_n}^n)})$. By absolute continuity of P_k with respect to Lebesgue measure, (2.7) becomes

$$\sum\cdots\sum \frac{1}{k_0!}\cdots\frac{1}{k_n!} \int_{\Delta_0^{(i_1^0)}}\cdots\int_{\Delta_n^{(i_{k_n}^n)}} r_A^k(\underline{x})\,d\underline{x} \quad + P(B\backslash C) \qquad (2.8)$$

$$= \frac{1}{k_0!}\cdots\frac{1}{k_n!} \int_{A_0^{k_0}\backslash\tilde{A}_0}\cdots\int_{A_n^{k_n}\backslash\tilde{A}_n} r_A^k(\underline{x})\,d\underline{x} + P(B\backslash C) \qquad (2.9)$$

Where $\underline{x} = (x_1^0,\ldots,x_{k_0}^0,x_1^1,\ldots,x_{k_1}^1,\ldots,x_1^n,\ldots,x_{k_n}^n)$ and $\tilde{A}_i$ = diagonal elements of the partition in $A_i^{k_i}$, i.e. elements of the partition which contain points of the diagonal of $A_i^{k_i}$.

By Lebesgue's Dominated Convergence Theorem, as we let $\epsilon \to 0$ (thus causing the diameters of the elements of $\mathcal{P}$ to approach 0, $A_i^{k_i} \setminus \tilde{A}_i \to A_i^{k_i}$, and $P(B \setminus C) \to 0$), (2.9) becomes

$$= \frac{1}{k_0!} \cdots \frac{1}{k_n!} \int_{A_0^{k_0}} \cdots \int_{A_n^{k_n}} r_A^k(\underline{x})\, d\underline{x} \tag{2.10}$$

Thus, $M(A_1 \times \ldots \times A_n) = \displaystyle\sum_{k_0=0,k_1,\ldots,k_n=1}^{\infty} k_1 \ldots k_n \frac{1}{k_0!} \cdots \frac{1}{k_n!} \int_{A_0^{k_0}} \cdots \int_{A_n^{k_n}} r_A^k(\underline{x})\, d\underline{x}$

$$= \sum_{k_0=0,k_1,\ldots,k_n=1}^{\infty} \frac{1}{k_0!} \frac{1}{(k_1-1)!} \cdots \frac{1}{(k_n-1)!} \int_{A_0^{k_0}} \cdots \int_{A_n^{k_n}} r_A^k(\underline{x})\, d\underline{x} \tag{2.11}$$

On the other hand,

$$\int_{A_1} \cdots \int_{A_n} \sum_{j=0}^{\infty} \int_{A^j} \frac{1}{j!}\, r_A^n(\, x_1,\ldots,x_n,\theta_1,\ldots,\theta_j\,) dx_1 \ldots dx_n d\theta_1 \ldots d\theta_j$$

$$= \int_{A_1} \cdots \int_{A_n} \sum_{j=0}^{\infty} \int_{A} \cdots \int_{A} \frac{1}{j!}\, r_A^n(\, x_1,\ldots,x_n,\theta_1,\ldots,\theta_j\,) dx_1 \ldots dx_n d\theta_1 \ldots d\theta_j \tag{2.12}$$

Let $j = k_0 + k_1 + \ldots + k_n - n$. Partition A into disjoint pieces $A_1,\ldots,A_n$ and let $A_0 = A \setminus \bigcup_{i=1}^{n} A_i$. Then rewriting the integrals over A in (2.12) as integrals over the subsets A_i we get

$$\sum_{k_0=0}^{\infty} \sum_{k_1=1}^{\infty} \cdots \sum_{k_n=1}^{\infty} \frac{j!}{k_0!(k_1-1)!\ldots(k_n-1)!} \int_{A_0^{k_0}} \cdots \int_{A_n^{k_n}} \frac{r_A^n(\,\tilde{x},\tilde{\theta})}{(\sum_{i=1}^{\infty} k_i) - n}\, d\tilde{x} d\tilde{\theta} \tag{2.13}$$

$$= \sum_{k_0=0,k_1=1,\ldots,k_n=1}^{\infty} \frac{1}{k_0!(k_1-1)!\ldots(k_n-1)!} \int_{A_0^{k_0}} \cdots \int_{A_n^{k_n}} r_A^n(\,\tilde{x},\tilde{\theta})\, d\tilde{x} d\tilde{\theta} \tag{2.14}$$

where $(\tilde{x},\tilde{\theta}) = (x_1,\ldots,x_n,\theta_1,\ldots,\theta_J)$. Since (2.14) matches (2.11) the theorem is proved in the case where $A_0,A_1,\ldots,A_n$ are rectangles. Since rectangles form a generating class for the σ-algebra of Borel sets on $\mathbb{R}^d$ the theorem holds for all Borel sets. □

Corollary 2.14 If X is a semi-regular, almost surely orderly point process with absolute product densities $r^n_A(x_1,\ldots,x_n)$ then

$$M_{[n]}(A^n) = \int_A \cdots \int_A p_n(x_1,\ldots,x_n)dx_1 \ldots dx_n \tag{2.15}$$

Proof This proof follows the above quite closely, so many details will be omitted. By definition

$$\begin{aligned} M_{[n]}(A^n) &= E[(X(A))(X(A)-1)\ldots(X(A)-n+1)] \\ &= \sum_{k=n}^{\infty} k(k-1)\ldots(k-n+1)P(X(A)=k) = \sum_{k=n}^{\infty}\frac{k!}{(k-n)!}P(X(A)=k) \end{aligned} \tag{2.16}$$

As in the above proof, partition A by a partition $\mathcal{P}(\nu)$ consisting of $T(\nu)$ subsets $\Delta^{(i)}(\nu)$, $i=1,\ldots,T(\nu)$ where each subset has diameter less than $\frac{1}{\nu}$ and belongs to a fixed substantial family $\mathcal{S}$. For fixed ν let B be the event that there are k points in A and C the event that there are k points in A, each in a different subset $\Delta^{(i)}(\nu)$. Then, as before $P(B\setminus C) < \epsilon$. Now,

$$P(X(A)=k) = \sum P_k(\Delta^1,\ldots,\Delta^k) + P(B\setminus C) \tag{2.17}$$

where the sum is taken over the $\binom{T(\nu)}{k}$ possible ways to choose k of the subsets from $\mathcal{P}(\nu)$ and $P_k(\Delta^1,\ldots,\Delta^k)$ is the probability that k points occur, one in each of the listed sets. By symmetry (2.17) is the same as

$$\sum_{\substack{i_1,\ldots,i_k=1 \\ i_j\neq i_l}}^{T(\nu)} \frac{1}{k!} P_k(\Delta^{i_1},\ldots,\Delta^{i_k}) + P(B\setminus C) \tag{2.19}$$

By absolute continuity of P_k with respect to Lebesgue measure this is

$$\sum \frac{1}{k!} \int_{\Delta^{i_1}} \cdots \int_{\Delta^{i_k}} r_A^k(\underline{x})\, d\underline{x} + P(B\backslash C) = \frac{1}{k!} \int_{A^n\backslash\tilde{A}} r_A^k(\underline{x})\, d\underline{x} + P(B\backslash C) \qquad (2.20)$$

Where $\underline{x} = (x_1,\ldots,x_k)$ and $\tilde{A}$ = diagonal elements of the partition in A^n, i.e. elements of the partition which contain points of the diagonal of A^n.

By Lebesgue's Dominated Convergence Theorem, as we let $\epsilon \to 0$, (2.20) becomes

$$= \frac{1}{k!} \int_{A^n} r_A^k(\underline{x})\, d\underline{x} \qquad (2.21)$$

Thus
$$\begin{aligned} M_{[n]}(A^n) &= \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} P(X(A)=k) \\ &= \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} \cdot \frac{1}{k!} \int_{A^n} r_A^k(\underline{x})\, d\underline{x} = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{A^n} r_A^{k+n}(\underline{x})\, d\underline{x} \qquad \square \end{aligned}$$

Piecing together the above theorem and corollary we get

Corollary 2.16 If X is a semi-regular, almost surely orderly point process with absolute product densities $r_A^n(x_1,\ldots,x_n)$ then
$$M_{[n]}(A_1^{t_1} \times \ldots \times A_k^{t_k}) = \int_{A_1^{t_1}} \cdots \int_{A_1^{t_k}} p_n(x_1,\ldots,x_n)\, dx_1 \ldots dx_n$$
where $t_1 + \ldots + t_k = n$.

By the above corollary, in the case when X is a.s. orderly and semi-regular the factorial moment measures $M_{[n]}(A_1^{t_1} \times \ldots \times A_k^{t_k})$ are also absolutely continuous with respect to Lebesgue measure. Conversely if we know that the moment measures are absolutely continuous and that the process X is a.s. orderly then a similar argument shows that X is semi-regular.

Definition 2.17 The functions $p(x_1,\ldots,x_n)$ will be referred to as the product densities of X.

The following theorem (Fisher [11]) shows that the product densities are well defined in the sense that the definition does not depend on the initial set A. That is, if we define p starting with a different set, say B, which also contains the points $x_1,\ldots,x_n$ and let p_A be the product density based on the set A, p_B the product density defined based on the set B, then $p_A(x_1,\ldots,x_n) = p_B(x_1,\ldots,x_n)$ almost surely.

Theorem 2.18 If X is a semi-regular, a.s. orderly point process then its product densities $p(x_1,\ldots,x_n)$ are independent of the set A.

Proof Let S be a set disjoint from A. If $x_1,\ldots,x_n$ are points in A then

$$r_A^n(x_1,\ldots,x_n) = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{S^k} r_{A\cup S}^{n+k}(x_1,\ldots,x_n,\theta_1,\ldots,\theta_k)\,d\theta_1\ldots d\theta_k \tag{2.22}$$

and

$$p_{A\cup S}(x_1,\ldots,x_n) = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{(A\cup S)^k} r_{A\cup S}^{n+k}(x_1,\ldots,x_n,\theta_1,\ldots,\theta_k)\,d\theta_1\ldots d\theta_k$$

$$= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \binom{k}{j} \int_{A^j} \int_{S^{k-j}} \frac{1}{k!} r_{A\cup S}^{n+k}(x_1,\ldots,x_n,\theta_1,\ldots,\theta_k)\,d\theta_1\ldots d\theta_k \tag{2.23}$$

Using Fubini's theorem to rearrange the sums we get

$$= \sum_{j=0}^{\infty} \frac{1}{j!} \int_{A^j} \left\{ \sum_{k-j=0}^{\infty} \int_{S^{k-j}} \frac{1}{(k-j)!} r_{A\cup S}^{(n+j)+(k-j)}(\underline{x},\underline{\theta})\,d\underline{\theta}^{(k-j)} \right\} d\underline{\theta}^{(j)} \tag{2.24}$$

where $\underline{x} = (x_1,\ldots,x_n)$, $\underline{\theta} = (\theta_1,\ldots,\theta_k)$, $\underline{\theta}^{(k-j)} = (\theta_1,\ldots,\theta_{k-j})$, and $\underline{\theta}^{(j)} = (\theta_{k-j},\ldots,\theta_k)$. By (2.22) this is

$$= \sum_{j=0}^{\infty} \frac{1}{j!} \int_{A^j} r_A^{n+j}(x_1,\ldots,x_n,\theta_1,\ldots,\theta_j)\,d\theta_1\ldots d\theta_j$$

$$= p_A(x_1,\ldots,x_n) \qquad \square$$

Lemma 2.19 If $p(x,y) \geq 0$ for all (x,y) and $\int_0^1 \int_0^1 p(x,y)\,dx\,dy < \infty$ then if we define $T_\Delta(x)$ by $T_\Delta(x)=\int_{\Delta x}\int_{\Delta x} p(u,v)\,du\,dv$, $T_\Delta(x)=o(|\Delta x|)$ for almost all x, where $\Delta x=[x-\frac{1}{2}\Delta, x+\frac{1}{2}\Delta]$ for $0<\Delta\frac{1}{2}(1-x)$.

<u>Proof</u> Case (1): Assume $p(x,y)>0$ for all (x,y). Let $g(x)=\int_0^1 p(x,u)\,du$. g is integrable for almost all x and $\frac{1}{|\Delta x|}\int_{\Delta x} g(u)\,du \to g(x)$ for almost all x. i.e. $S_\Delta(x)=\int_{\Delta x} g(u)\,du=O(|\Delta x|)$ for almost all x. Thus

$$T_\Delta(x)=\frac{T_\Delta(x)}{S_\Delta(x)}S_\Delta(x)=\frac{T_\Delta(x)}{S_\Delta(x)}O(|\Delta x|) \quad \text{for almost all x.} \tag{2.25}$$

It suffices then to show that $\frac{T_\Delta(x)}{S_\Delta(x)}\to 0$ for almost all x as $|\Delta x|\to 0$. Assume not, then there exists a measurable set $F\subseteq[0,1]$ with $|F|>0$ and an $\alpha>0$ so that for every $x\in F$, $\limsup_{|\Delta x|\to 0}\frac{T_\Delta(x)}{S_\Delta(x)}\geq\alpha>0$. Let $\mathcal{G}$ be the class of sets $\mathcal{G}=\left\{[x-\frac{1}{2}\Delta, x+\frac{1}{2}\Delta]=\Delta x \mid \frac{T_\Delta(x)}{S_\Delta(x)}>\frac{\alpha}{2}\right\}$. $\mathcal{G}$ is a Vitali cover of F. That is, for each $\epsilon>0$ and $x\in F$ there is a $\Delta x\in\mathcal{G}$ so that $x\in\Delta x$ and $|\Delta x|<\epsilon$ (see Royden [30]).

Now, let $\epsilon=\frac{\alpha}{4}\int_F g(u)\,du$ and $\epsilon'=\frac{1}{2}\int_F g(u)\,du$. Recall that for g nonnegative and integrable with respect to a measure μ over a set F, given $\epsilon>0$ there is a $\delta>0$ so that for every set $E\subseteq F$ with $|E|<\delta$ we have $\int_E g\,d\mu<\epsilon$ ([30]). We now choose $\delta>0$ so that $|E|<\delta\Rightarrow\int_E p(u,v)\,du\,dv<\epsilon$ and $\delta'>0$ so that $|E|<\epsilon\Rightarrow\int_E g(u)\,du<\epsilon'$.

Let $\Delta_i<\delta$ and by the Vitali covering lemma find disjoint intervals $\Delta x_i=[x_i-\frac{1}{2}\Delta_i, x_i+\frac{1}{2}\Delta_i]$, $i=1,\ldots,N$ so that $|F\setminus\bigcup_{i=1}^N\Delta x_i|<\delta'$. Note that

$|\bigcup_{i=1}^{N}[x_i - \frac{1}{2}\Delta_i, x_i + \frac{1}{2}\Delta_i]^2| = \sum_{i=1}^{N} \Delta_i^2 < \delta \sum_{i=1}^{N} \Delta_i < \delta$ since Δx_i, $i=1,\ldots,N$ are disjoint subsets of [0,1]. Thus

$$\frac{T_{\Delta_i}(x)}{S_{\Delta_i}(x)} > \frac{\alpha}{2} \Rightarrow \sum_{i=1}^{N} T_{\Delta_i}(x) > \frac{\alpha}{2} \sum_{i=1}^{N} S_{\Delta_i}(x) \text{ so that}$$

$$\epsilon > \int_{\bigcup_{i=1}^{N}(\Delta x_i)^2} p(u,v)\,du\,dv > \frac{\alpha}{2} \int_{\bigcup_{i=1}^{N}(\Delta x_i)} g(u)\,du$$

$$= \frac{\alpha}{2}\left(\int_F g(u)\,du - \int_{F \setminus \bigcup_{i=1}^{N}(\Delta x_i)} g(u)\,du \right) > \frac{\alpha}{2}\left\{ \int_F g(u)\,du - \epsilon' \right\}$$

$$= \frac{\alpha}{4} \int_F g(u)\,du = \epsilon$$

which is a contradiction.

Case (2): If there exist (x,y) such that $p(x,y) = 0$ we may apply case 1 to the function $h(x,y) = p(x,y) + 1$, since $h(x,y) > 0$. By case 1 for almost all x, $\frac{1}{|\Delta x|} \int_{\Delta x} \int_{\Delta x} h(u,v)\,du\,dv \to 0$ as $|\Delta x| \to 0$. Thus, $\frac{1}{|\Delta x|} \int_{\Delta x} \int_{\Delta x} 1 + p(u,v)\,du\,dv$ $= |\Delta x| + \frac{1}{|\Delta x|} \int_{\Delta x} \int_{\Delta x} p(u,v)\,du\,dv \to 0$ for almost all x as $|\Delta x| \to 0$. Thus $\frac{1}{|\Delta x|} \int_{\Delta x} \int_{\Delta x} p(u,v)\,du\,dv \to 0$ for almost all x. □

<u>Lemma</u> <u>2.20</u> If $p(x_1,\ldots,x_n) \geq 0$ for all $(x_1,\ldots,x_n)$ and $\int_A \cdots \int_A p(x_1,\ldots,x_n)dx_1\ldots dx_n < \infty$ for d-dimensional rectangles A then if we define $T_\Delta(x_1,\ldots,x_n)$ by $T_\Delta(x_1,\ldots,x_n) = \int_{\Delta x_1} \int_{\Delta x_1} \int_{\Delta x_2} \cdots \int_{\Delta x_n} p(u_1,\ldots,u_{n+1})du_1\ldots du_{n+1}$, $T_\Delta(x_1,\ldots,x_n) = o(\max_i |\Delta x_i|)$ for almost all $(x_1,\ldots,x_n)$ as $\max_i |\Delta x_i| \to 0$, where Δx_i is

the d-dimensional cube centered at x_i, having volume $|\Delta x_i| > 0$.

Proof As above we first assume that $p(x_1,\ldots,x_n) > 0$ for all $(x_1,\ldots,x_n)$. Let $h(x_1,\ldots,x_n) = \int_A p(x_1,u,x_2,\ldots,x_n)\,du$. h is integrable for almost all $(x_1,\ldots,x_n)$ and $\frac{1}{|\Delta x_1|\ldots|\Delta x_n|} \int_{\Delta x_1} \cdots \int_{\Delta x_n} h(u_1,\ldots,u_n) \to h(x_1,\ldots,x_n)$ for almost all $(x_1,\ldots,x_n)$, i.e. $S_\Delta(x_1,\ldots,x_n) = \int_{\Delta x_1} \cdots \int_{\Delta x_n} h(u_1,\ldots,u_n) = O(\max_i |\Delta x_i|)$ for almost all x. Thus, as in, Lemma 2.19 we need only show that $\frac{T_\Delta(x_1,\ldots,x_n)}{S_\Delta(x_1,\ldots,x_n)} \to 0$ for almost all $(x_1,\ldots,x_n)$ as $\max_i |\Delta x_i| \to 0$. Assume not, then there exists a measurable set $F \subseteq A^n$ with $|F| > 0$ and an $\alpha > 0$ so that for every $(x_1,\ldots,x_n) \in F$, $\limsup \frac{T_\Delta(x_1,\ldots,x_n)}{S_\Delta(x_1,\ldots,x_n)} \geq \alpha > 0$. To complete the proof, proceed now as in the proof of Lemma 2.19 applying the multidimensional version of the Vitali covering lemma (see Cohn [7]). □

Note that the above lemma continues to hold if each Δx_i is, instead of a cube, a member of some fixed substantial family $\mathcal{S}$.

For X semi-regular and almost surely orderly let $H(A_1,\ldots,A_n)$ be the probability that exactly one point occurs in each of the sets $A_1,\ldots,A_n$ and other points may or may not occur in A. Then

$$H(A_1,\ldots,A_n) = \sum_{j=0}^{\infty} \frac{(j+n)!}{j!} R_A^{j+n}(A_1,\ldots,A_n,A_0^j) \tag{2.26}$$

Where $A_0 = A \setminus \bigcup_{i=1}^{n} A_i$. By semi- regularity this becomes

$$H(A_1,\ldots,A_n) = \sum_{j=0}^{\infty} \frac{1}{j!} \int_{A_1} \cdots \int_{A_n} \int_{A_0^j} r_A^{j+n}(x_1,\ldots,x_{j+n})\,dx_1 \ldots dx_{j+n} \tag{2.27}$$

The following theorem shows that $p(x_1,\ldots,x_n)$ can be interpreted as being the functions for which $p(x_1,\ldots,x_n)|\Delta x_1|\ldots|\Delta x_n|$ approximates $H(\Delta x_1,\ldots,\Delta x_n)$ for $\Delta x_1,\ldots,\Delta x_n$ sufficiently small. That is, the probability that there is an occurrence in each Δx_i is $p(x_1,\ldots,x_n)|\Delta x_1|\ldots|\Delta x_n| + o(\max_i |\Delta x_i|)$.

<u>Theorem</u> <u>2.21</u> If X is semi-regular and a.s. orderly with product densities $p_n(x_1,\ldots,x_n)$, then H is absolutely continuous with respect to Lebesgue measure and has density

$$p_n(x_1,\ldots,x_n) = \sum_{j=0}^{\infty} \frac{1}{j!} \int_{A^j} r_A^{n+j}(x_1,\ldots,x_n,\theta_1,\ldots,\theta_j)\,d\theta_1\ldots d\theta_j \tag{2.28}$$

<u>Proof</u> The proof begins by following Macchi's work [21] and concludes by applying Lemma 2.20. Let $A_1,\ldots,A_n$ be disjoint subsets of A and members of a fixed substantial family. We know that p is the density for the factorial moment measure $M(A_1\times\ldots\times A_n)$. i.e.,

$$M(A_1\times\ldots\times A_n) = \int_{A_1}\cdots\int_{A_n} \sum_{j=0}^{\infty} \frac{1}{j!} \int_{A^j} r_A^{n+j}(x_1,\ldots,x_n,\theta_1,\ldots,\theta_j)\,d\theta_1\ldots d\theta_j\,dx_1\ldots dx_n$$

We compare this with (2.27) to get

$$M(A_1\times\ldots\times A_n) - H(A_1,\ldots,A_n) =$$

$$\int_{A_1}\cdots\int_{A_n} \sum_{j=0}^{\infty} \frac{1}{j!} \int_{A^j\setminus A_0^j} r_A^{n+j}(x_1,\ldots,x_n,\theta_1,\ldots,\theta_j)\,d\theta_1\ldots d\theta_j\,dx_1\ldots dx_n \tag{2.29}$$

If we let each A_i decrease in size $|\Delta x_i|$ towards the set consisting of the single point x_i^0, with each x_i^0 distinct, then A_0 increases to A.

Define Q by

$$Q = \frac{M(A_1\times\ldots\times A_n) - H(A_1,\ldots,A_n)}{|\Delta x_1|\ldots|\Delta x_n|} \tag{2.30}$$

Note that $A^j \setminus A_0^j = \bigcup_{i=0}^{j-1} X^i\,(A\setminus A_0)\,X^{j-i-1}$. Then due to symmetry of the absolute product densities we may consider only the set where $i=0$ to obtain

$$\int_{A^j\setminus A_0^j} r_A^{n+j}(\,x_1,\ldots,x_n,\theta_1,\ldots,\theta_j\,)d\theta_1\ldots d\theta_j$$

$$= \int_{\bigcup_{i=0}^{j-1} X^i\,(A\setminus A_0)\,X^{j-i-1}} r_A^{n+j}(\,x_1,\ldots,x_n,\theta_1,\ldots,\theta_j\,)d\theta_1\ldots d\theta_j$$

$$= j\cdot \int_{(A\setminus A_0)\,A^{j-1}} r_A^{n+j}(\,x_1,\ldots,x_n,\theta_1,\ldots,\theta_j\,)d\theta_1\ldots d\theta_j \tag{2.31}$$

So that $Q|\Delta x_1|\ldots|\Delta x_n|$

$$= \int_{A_1}\cdots\int_{A_n}\sum_{j=0}^{\infty}\frac{j}{j!}\int_{(A\setminus A_0)A^{j-1}} r_A^{n+j}(\,x_1,\ldots,x_n,\theta_1,\ldots,\theta_j\,)d\theta_1\ldots d\theta_j dx_1\ldots dx_n$$

$$= \int_{A_1}\cdots\int_{A_n}\int_{(A\setminus A_0)}\sum_{j=1}^{\infty}\frac{1}{(j-1)!}\int_{A^{j-1}} r_A^{n+j}(\,x_1,\ldots,x_n,\theta_1,\ldots,\theta_j\,)d\theta_1\ldots d\theta_j dx_1\ldots dx_n$$

$$= \int_{A_1\ldots A_n(A\setminus A_0)}\cdots\int\sum_{j=0}^{\infty}\int_{A^j}\frac{1}{j!}\, r_A^{n+j+1}(\,x_1,\ldots,x_{n+1},\theta_1,\ldots,\theta_j\,)d\theta_1\ldots d\theta_j dx_1\ldots dx_{n+1} \tag{2.32}$$

Thus
$$Q = \frac{M_{[n+1]}(A_1\times\ldots\times A_n\times(A\setminus A_0))}{|\Delta x_1|\ldots|\Delta x_n|}. \tag{2.33}$$

but by Lemma 2.20

$$M_{[n+1]}(A_1\times\ldots\times A_n\times(A\setminus A_0)) = M_{[n+1]}\Big(\bigcup_{i=1}^{n} A_1\times\ldots\times A_{i-1}\times A_i^2\times A_{i+1}\times\ldots\times A_n\Big)$$

$$= \sum_{i=1}^{n} M_{[n+1]}(A_1\times\ldots\times A_{i-1}\times A_i^2\times A_{i+1}\times\ldots\times A_n)$$

$$= \sum_{i=1}^{n} \int_{A_1} \cdots \int_{A_i^2} \cdots \int_{A_n} p(x_1,\ldots,x_{n+1})\, dx_1 \ldots dx_{n+1}$$

$$= \sum_{i=1}^{n} o(\max_i |\Delta x_i|) \tag{2.34}$$

for a.a. $(x_1,\ldots,x_n)$ so that $Q|\Delta x_1|,\ldots,|\Delta x_n| = o(\max_i |\Delta x_i|)$. i.e.,

$$\lim_{|\Delta x_1|,\ldots,|\Delta x_n| \to 0} \frac{M(A_1 \times \ldots \times A_n) - H(A_1,\ldots,A_n)}{|\Delta x_1| \ldots |\Delta x_n|} =$$

$$\lim_{|\Delta x_1|,\ldots,|\Delta x_n| \to 0} \frac{M(A_1 \times \ldots \times A_n)}{|\Delta x_1|,\ldots,|\Delta x_n|} - \lim_{|\Delta x_1|,\ldots,|\Delta x_n| \to 0} \frac{H(A_1,\ldots,A_n)}{|\Delta x_1|,\ldots,|\Delta x_n|} = 0 \tag{2.35}$$

So that

$$\lim_{|\Delta x_1|,\ldots,|\Delta x_n| \to 0} \frac{H(A_1,\ldots,A_n)}{|\Delta x_1|,\ldots,|\Delta x_n|} = \lim_{|\Delta x_1|,\ldots,|\Delta x_n| \to 0} \frac{M(A_1 \times \ldots \times A_n)}{|\Delta x_1|,\ldots,|\Delta x_n|}$$

$$= p(x_1,\ldots,x_n) \tag{2.36}$$

Thus, as each Δx_i approaches 0, $\frac{H(A_1,\ldots,A_n)}{|\Delta x_1|,\ldots,|\Delta x_n|}$ approaches $p(x_1,\ldots,x_n)$, that is, $H(A_1,\ldots,A_n)$ has density $p(x_1,\ldots,x_n)$. □

The first part of the following theorem was proved by Macchi in [21]. The extension of the theorem follows from Lemma 2.20.

Theorem 2.23 If X is a.s orderly and $p(x,y)$ is bounded on compact sets then X is analytically orderly. If X is a.s. orderly and the functions $p(x,y)$ exist then X is analytically orderly almost everywhere.

Proof Let Δx be a neighborhood of x and assume that $p(x,y) < N$ for each x, y $\in \Delta x$. Then

$$E[X(\Delta x)(X(\Delta x)-1)] = \int_{\Delta x}\int_{\Delta x} p(x,y)\,dxdy < N\,|\Delta x|^2 \tag{2.37}$$

which implies that

$$|\Delta x|^{-2}\,E[X(\Delta x)(X(\Delta x)-1)] = |\Delta x|^{-2}\int_{\Delta x}\int_{\Delta x} p(x,y)\,dxdy < N \tag{2.38}$$

But, since $P(X(\Delta x) \geq 2) = \sum_{k=2}^{\infty} P(X(\Delta x)=k)$ and $E[X(\Delta x)(X(\Delta x)-1)]$

$= \sum_{k=1}^{\infty} k(k-1)P(X(\Delta x)=k)$ we have that $E[X(\Delta x)(X(\Delta x)-1)] \geq P(X(\Delta x) \geq 2)$

So that $\dfrac{P(X(\Delta x) \geq 2)}{|\Delta x|^2} < N$, which implies $\dfrac{P(X(\Delta x) \geq 2)}{|\Delta x|} < N\,|\Delta x|$, which in

turn implies $\lim_{|\Delta x|\to 0} \dfrac{P(X(\Delta x) \geq 2)}{|\Delta x|} = 0$. i.e. $P(X(\Delta x) \geq 2) = o(|\Delta x|)$.

Now assume that p(x,y) exists for every pair x, y, but is not necessarily bounded. As above we have

$$P(X(\Delta x) \geq 2) \leq E[X(\Delta x)(X(\Delta x)-1)] = \int_{\Delta x}\int_{\Delta x} p(x,y)\,dxdy \tag{2.39}$$

but by Lemma 2.20 $\int_{\Delta x}\int_{\Delta x} p(x,y)\,dxdy = o(|\Delta x|)$ for almost all x as $|\Delta x| \to 0$ □

Note that a.s. orderly is a necessary condition in the above theorem. For example, if we take X_1 to be uniform on [0,1] and let X be the point process $X = 2\delta_{X_1}$. Then $p(x)=2$, and $p(x_1,\ldots,x_n)=0$ for each $n>1$, so that the product densities exist and are even bounded, but X is clearly not analytically orderly.

<u>Corollary</u> <u>2.24</u> If X is a stationary, a.s. orderly point process for which p(x,y) exists almost everywhere, then X is analytically orderly.

Proof The set S of points where X is not analytically orderly has measure 0 by the above theorem. By stationarity of X, S is translation invariant, thus S must be empty. □

Example 2.25 It is possible, as this example will show, to have a point process X with product densities which are finite almost everywhere but X is not analytically orderly. The process X is determined as follows. Let X_1 be a random variable distributed on [0,1] with density $f_1(x)=\frac{1}{2}x^{-\frac{1}{2}}$. Then, choose X_2 distributed uniformly on $[0,x_1]$. X_2 thus has density given by

$$f_2(x)=\begin{cases}\frac{1}{x_1} \text{ for } 0\leq x_2\leq x_1\\ 0 \text{ otherwise}\end{cases}. \tag{2.40}$$

Thus, $P(X(0,\Delta t)\geq 2)=P(X_1\in[0,\Delta t])=\int_0^{\Delta t}\frac{1}{2}x^{-\frac{1}{2}}dx=x^{\frac{1}{2}}\Big|_0^{\Delta t}=(\Delta t)^{\frac{1}{2}}$, which is clearly not $o(|\Delta t|)$ as $\Delta t\to 0$.

On the other hand we can easily calculate the product densities of X. If we let $0\leq x<y\leq 1$ we find $p(x,y)=\frac{1}{2}y^{-\frac{1}{2}}\cdot\frac{1}{y}=\frac{1}{2}y^{-\frac{3}{2}}=r(x,y)$. Note that $r(x)$ and $r(x_1,\ldots,x_n)$ are 0 for $n>2$ since we always have exactly two points. Thus, $p(x)=\sum_{j=0}^{\infty}\frac{1}{j!}\int_{[0,1]^j} r_{[0,1]}^{1+j}(x,\theta_1,\ldots,\theta_j)d\theta_1\ldots d\theta_j=\int_0^1 r(x,y)dy$

$$=\int_0^x r(x,y)dy+\int_x^1 r(x,y)dy=\int_0^x\frac{1}{2}x^{-\frac{3}{2}}dy+\int_x^1\frac{1}{2}y^{-\frac{3}{2}}dy$$

$=\frac{1}{2}x^{-\frac{1}{2}}+\frac{1}{2}(-2y^{-\frac{1}{2}})\Big|_x^1=\frac{1}{2}x^{-\frac{1}{2}}-1+x^{-\frac{1}{2}}=\frac{3}{2}x^{-\frac{1}{2}}-1$, so that the product densities are finite except at 0.

Corollary 2.23 indicates that Examples 2.8, 2.9, and 2.24 are in some sense the best possible.

<u>Theorem</u> <u>2.26</u> Let R be a bounded d-dimensional rectangle. X has absolute product densities $r_A^n(x_1,\ldots,x_n)$ if and only if it is absolutely continuous with respect to a Poisson point process.

<u>Proof</u> Let P_X be the distribution of X and ρ the distribution of a Poisson point process with parameter $\lambda=1$. Assume that $P_X << \rho$. By the Radon-Nikodym Theorem there exists a measurable function ϕ on $(N,\mathcal{N})$ such that $P_X(d\mu) = \phi(\mu)\rho(d\mu)$.

Recall that a measure $\mu \in \mathcal{N}$ corresponds to an unordered set of points $\{x_1,\ldots,x_n\}$ in $\mathbb{R}^d$. Let $N_n = \{\mu \mid \mu(R) = n\}$. N_n corresponds to all n point configurations in $\mathbb{R}^d$ and $N = \bigcup_{n=0}^{\infty} N_n$. For any subset A of N_n define $\tilde{A}$ by $\tilde{A} = \{(x_1,\ldots,x_n) \mid$ there exists $\mu \in A$ with μ corresponding to the measure $\delta_{x_1} + \ldots + \delta_{x_n}\}$. Then

$$P(X \in A) = P_X(A) = \int_A \phi(\mu)\rho(d\mu) = \frac{1}{n!}\int_{\tilde{A}} \tilde{\phi}(x_1,\ldots,x_n)e^{-|R|}dx_1\ldots dx_n \qquad (2.41)$$

Given disjoint subsets $B_1,\ldots,B_n$ of R let $A = \{\mu \in N_n \mid \mu(B_i) = 1,\ i=1,\ldots,n\}$ then

$$\frac{1}{n!}\int_{\tilde{A}} \tilde{\phi}(x_1,\ldots,x_n)e^{-|R|}dx_1\ldots dx_n = \frac{1}{n!}\int_{B_1}\cdots\int_{B_n} \tilde{\phi}(x_1,\ldots,x_n)e^{-|R|}dx_1\ldots dx_n \qquad (2.42)$$

but $P(X \in A) = R_A^n(x_1,\ldots,x_n)$, so we have found that $n!R_A^n$ has density $\tilde{\phi}(x_1,\ldots,x_n)e^{-|R|}$. By uniqueness of the absolute product density, X then has absolute product density $r_A^n(x_1,\ldots,x_n) = \tilde{\phi}(x_1,\ldots,x_n)e^{-|R|}$ a.e.

Conversely, if X has absolute product density $r_A^n(x_1,\ldots,x_n)$ we let $\tilde{\phi}(x_1,\ldots,x_n) = r_A^n(x_1,\ldots,x_n)e^{|R|}$. Then following the above argument in reverse $\phi(\mu)$ is the density of P_X with respect to ρ. i.e. $P_X << \rho$ with Radon-Nikodym derivative $\phi(\mu)$. □

From here on we will take X to be semi-regular and a.s. orderly.

Definition 2.27 A point process X is called completely regular if its absolute product densities exist and can be computed from its product densities using the following inversion formula

$$r_A^n(x_1,\ldots,x_n)=\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\int_{A^j} p(x_1,\ldots,x_n,\theta_1,\ldots,\theta_j)d\theta_1\ldots d\theta_j \tag{2.43}$$

Note that if the sum in (2.43) is absolutely convergent we can rearrange the sums and integrals in the expression

$$\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\int_{A^j}\sum_{k=0}^{\infty}\frac{1}{k!}\int_{A^k} r_A^{(n+j)+k}(x_1,\ldots,x_{n+j+k})dx_{n+j+1}\ldots dx_{n+j+k} \tag{2.44}$$

and by the binomial theorem obtain $r_A^n(x_1,\ldots,x_n)$. i.e. if (2.43) is absolutely convergent, the inversion formula is valid, so X is completely regular. For example, as long as the product densities p are bounded by an exponential function, the process is completely regular.

2.3 Conditional Densities and Cumulant Densities

In order to consider conditional distributions, that is the probability of events occurring given that we already have some information about which points have occurred, we need conditional densities. The conditional product densities are defined in a natural way.

Definition 2.28 A point process X with product densities $p(x_1,\ldots,x_n)$ has conditional product densities given by

$$p(x_1,\ldots,x_n|y_1,\ldots,y_m)=\frac{p(x_1,\ldots,x_n,y_1,\ldots,y_m)}{p(y_1,\ldots,y_m)} \tag{2.45}$$

and <u>conditional</u> <u>absolute</u> <u>product</u> <u>densities</u> given by

$$r_A(x_1,\ldots,x_n|y_1,\ldots,y_m) = \frac{r_A(x_1,\ldots,x_n,y_1,\ldots,y_m)}{p(y_1,\ldots,y_m)} \tag{2.46}$$

for $x_1,\ldots,x_n,y_1,\ldots,y_m \in A$.

If X has product densities one can also define <u>cumulant</u> <u>densities</u> $q(x_1,\ldots,x_n)$ (also called the correlation functions) corresponding to X inductively by the following relationship with $p(x_1,\ldots,x_n)$

$$\begin{aligned} p(x_1) &= q(x_1) \\ p(x_1,x_2) &= q(x_1,x_2) + q(x_1)q(x_2) \\ p(x_1,x_2,x_3) &= q(x_1,x_2,x_3) + q(x_1,x_2)q(x_3) + q(x_2,x_3)q(x_1) \\ &\quad + q(x_1,x_3)q(x_2) + q(x_1)q(x_2)q(x_3) \end{aligned} \tag{2.47}$$

and so on, so that $p(x_1,\ldots,x_n)$ is written in terms of q by subdividing $(x_1,\ldots,x_n)$ into all possible configurations of disjoint subsets and adding the corresponding product of q's.

2.4 Generating Functionals

We seek here to extend the notion of multivariate probability generating functions and characteristic functions to the more general setting of point processes. This development follows that given by Gupta and Waymire [13] or Fisher [11]. More on the theory of probability generating functionals for point processes can be found in Westcott [33].

We begin by considering the finite dimensional structure of the process X in order to discover what the natural extension ought to be. Let $A_1,\ldots,A_n$ be subsets of $\mathbb{R}^d$. The joint distribution of the random vector

$(X(A_1),\ldots,X(A_n))$ is uniquely determined by the probability generating functional

$$g(t_1,\ldots,t_n) = E\left\{t_1^{X(A_1)} \cdot \ldots \cdot t_n^{X(A_n)}\right\}, \qquad 0 \leq t_i \leq 1 \tag{2.48}$$

We can rewrite (2.48) as

$$g(t_1,\ldots,t_n) = E\left\{\exp\left[\log\left(t_1^{X(A_1)} \cdot \ldots \cdot t_n^{X(A_n)}\right)\right]\right\}$$

$$= E\left\{\exp\left[\sum_{i=1}^{n} \log(t_i)X(A_i)\right]\right\} \tag{2.49}$$

Since X can be thought of as corresponding directly to a sequence of points $\omega = \{x_i\}$ in $\mathbb{R}^d$, it induces a counting measure $N(A)(\omega) = \#\{i\colon x_i \in A_i\}$ on $\mathbb{R}^d$. Thus, given a real-valued measurable function f on $\mathbb{R}^d$ we may define for each ω the integral of f with respect to the process X as follows

$$\int_{\mathbb{R}^d} f(x)\, dX(x) = \sum_i f(x_i). \tag{2.50}$$

If we define the function ξ on $\mathbb{R}^d$ by

$$\xi(x) = \begin{cases} t_i \text{ for } x \in A_i,\ 1 \leq i \leq n \\ 1 \text{ for } x \notin \bigcup_{i=1}^{n} A_i \end{cases} \tag{2.51}$$

Then

$$\sum_{i=1}^{n} \log(t_i)X(A_i) = \int_{\mathbb{R}^d} \log \xi(x)\, dX(x) \tag{2.52}$$

So that,

$$g(t_1,\ldots,t_n) = E\left\{\exp\left[\int_{\mathbb{R}^d} \log \xi(x)\, dX(x)\right]\right\} \tag{2.53}$$

The above discussion leads quite naturally to the following definition:

<u>Definition</u> <u>2.29</u> Let S be the set of all real valued, measurable

functions ξ on $\mathbb{R}^d$ satisfying

(i) $0 \leq \xi(x) \leq 1$ for all $x \in \mathbb{R}^d$

(ii) $\xi(x) = 1$ on the complement of a bounded subset of $\mathbb{R}^d$ (2.54)

then the Probability Generating Functional corresponding to X is defined by

$$G(\xi) = E\Big\{\exp\Big[\int_{\mathbb{R}^d} \log \xi(x)\, dX(x)\Big]\Big\}, \quad \xi \in S \tag{2.55}$$

Note that the finite dimensional distributions of X can be recovered by taking ξ to be of the form (2.51).

Theorem 2.30 The probability generating functional G of a point process X can be expressed as

$$G(\xi) = E\Big\{ \prod_{x_i \in \omega} \xi(x_i) \Big\}$$

where ω is the point configuration $\{x_i\}$ corresponding to X.

Proof $G(\xi) = E\Big\{\exp\Big[\int_{\mathbb{R}^d} \log \xi(x)\, dX(x)\Big]\Big\} = E\Big\{\exp\Big[\sum_{i=1}^{n} \log(t_i)\Big]\Big\}$ (by (2.49))

$$= E\Big\{ \prod_{x_i \in \omega} \exp(\log \xi(x_i)) \Big\} = E\Big\{ \prod_{x_i \in \omega} \xi(x_i) \Big\}. \qquad \square$$

Definition 2.31 The Characteristic functional Φ of a point process X is defined by

$$\Phi(\phi) = E\Big\{\exp\Big[i \int_{\mathbb{R}^d} \phi(x)\, dX(x)\Big]\Big\} \tag{2.56}$$

for ϕ bounded, measurable, and having compact support.

The characteristic functional is actually just a special case of the probability generating functional. If we allow complex valued functions ξ and take ξ to be of the form $\xi(x) = e^{i\phi(x)}$, then $G(\xi) = \Phi(\phi)$.

The factorial moment measures (and thus the product densities) can be computed using the probability generating functional according to the following formula due to Moyal [23]:

$M_{x_1,\ldots,x_n} \equiv M([0,x_1],\ldots,[0,x_n]) =$

$$\lim_{\eta\uparrow 1}\left\{\frac{\partial}{\partial\lambda_1\ldots\partial\lambda_n} G\left[\eta+\sum_{i=1}^{n}\lambda_i 1_{[0,x_i]}\right]\right\}_{\lambda_1=\ldots=\lambda_n=0} \tag{2.57}$$

Where $1_{[0,x_i]}$ denotes the indicator function of the interval $[0,x_i]$, that is

$$1_{[0,x_i]}(x) = \begin{cases} 1 \text{ if } x \in [0,x_i] \\ 0 \text{ if } x \notin [0,x_i] \end{cases} \tag{2.58}$$

To calculate the product densities from this formula we differentiate $M_{x_1,\ldots,x_n}$. That is,

$$p(x_1,\ldots,x_n) = \frac{\partial^n}{\partial x_1\ldots\partial x_n}(M_{x_1,\ldots,x_n}) \tag{2.59}$$

2.5 Examples of Point Processes

Example 2.32 The absolute product densities of a stationary Poisson point process (defined in Example 2.3) are easily calculated:

$$\begin{aligned} r_A^n(x_1,\ldots x_n)|\Delta x_1|\ldots|\Delta x_n| + o(|\Delta x_1|\ldots|\Delta x_n|) &= P(X(\Delta x_1)=1,\ldots,X(\Delta x_n)=1,X(A)=n) \\ &= (\lambda^1|\Delta x_1|^1 e^{-\lambda|\Delta x_1|}\ldots\lambda^1|\Delta x_n|^1 e^{-\lambda|\Delta x_n|})\lambda^0|A'|^0 e^{-\lambda|A'|} \\ &= \lambda^n e^{-\lambda|A|}|\Delta x_1|\ldots|\Delta x_n|, \text{ where } A' = A\setminus\bigcup_{i=1}^{n}\Delta x_i. \end{aligned}$$

Thus $r_A^n(x_1,\ldots x_n) + \dfrac{o(|\Delta x_1|\ldots|\Delta x_n|)}{|\Delta x_1|\ldots|\Delta x_n|} = \lambda^n e^{-\lambda|A|}$. Taking the limit as the sets Δx_i decrease to the single points x_i we obtain

$$r_A^n(x_1,\ldots x_n) = \lambda^n e^{-\lambda|A|} \tag{2.60}$$

The probability generating functional of a Poisson point process X with intensity Λ has the form:

$$G(\xi) = \exp\{\int [\xi(x)-1]\,d\Lambda(x)\} \tag{2.61}$$

and its characteristic functional is:

$$\Phi(\phi) = \exp\{\int [e^{i\phi(x)}-1]\,d\Lambda(x)\} \tag{2.62}$$

For a Poisson point process with intensity Λ which is absolutely continuous with respect to Lebesgue measure with density f(x), Moyal's formula (2.57) can be used to calculate the product densities as follows:

Using the probability generating functional given above (2.61),

$$M_{x_1,\ldots,x_n} = \lim_{\eta\uparrow 1}\left\{\frac{\partial^n}{\partial\lambda_1\ldots\partial\lambda_n} G\Big(\eta+\sum_{i=1}^{n}\lambda_i 1_{[0,x_i]}\Big)\right\}_{\lambda_1=\ldots=\lambda_n=0}$$

$$=\left\{\frac{\partial}{\partial\lambda_1\ldots\partial\lambda_n}\exp\Big(\int\sum_{i=1}^{n}\lambda_i 1_{[0,x_i]}f(x)dx\Big)\right\}_{\lambda_1=\ldots=\lambda_n=0}$$

$$=\left\{\Big(\prod_{i=1}^{n}\int f(x)1_{[0,x_i]}dx\Big)\exp\Big(\int\sum_{i=1}^{n}\lambda_i 1_{[0,x_i]}f(x)dx\Big)\right\}_{\lambda_1=\ldots=\lambda_n=0}$$

$$=\prod_{i=1}^{n}\int f(x)1_{[0,x_i]}dx \tag{2.63}$$

So that $p(x_1,\ldots,x_n) = \dfrac{\partial^n}{\partial x_1\ldots\partial x_n}(M_{x_1,\ldots,x_n})$

$$=\frac{\partial^n}{\partial x_1\ldots\partial x_n}\prod_{i=1}^{n}\int f(x)1_{[0,x_i]}dx = f(x_1)\cdot\ldots\cdot f(x_n) \tag{2.64}$$

Using the inversion formula (2.43) we can now easily calculate the absolute product densities:

$$r_A(x_1,\ldots,x_n) = \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\int_{A^j} p(x_1,\ldots,x_n,\theta_1,\ldots,\theta_j)\,d\theta_1\ldots d\theta_j$$

$$= \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \int_{A^J} f(x_1)\ldots f(x_n) f(\theta_1)\ldots f(\theta_J) d\theta_1 \ldots d\theta_J$$

$$= f(x_1)\cdot\ldots\cdot f(x_n) \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \int_{A^J} f(\theta_1)\ldots f(\theta_J) d\theta_1 \ldots d\theta_J$$

$$= f(x_1)\cdot\ldots\cdot f(x_n) \sum_{j=0}^{\infty} \frac{(-1)^J}{j!} \left(\int_A f(\theta)\, d\theta \right)^J$$

$$= f(x_1)\cdot\ldots\cdot f(x_n) \exp\left(-\int_A f(\theta)\, d\theta \right). \tag{2.65}$$

Example 2.33 To obtain a Mixed Poisson Process we start with a random variable I. We then take X to be a stationary Poisson process with (random) intensity I. That is, one first observes the outcome of I and then forms a Poisson process with that outcome for its intensity. A mixed Poisson process X has probability generating functional given by

$$G(\xi) = E_I[\exp(\int [\xi(x) - 1]\, dI(x))] \tag{2.66}$$

By our calculations for the stationary Poisson process (2.60) the mixed Poisson process has absolute product densities

$$r_A(x_1,\ldots,x_n) = E[I^n e^{-I|A|}]. \tag{2.67}$$

The corresponding product densities become

$$p(x_1,\ldots,x_n) = \sum_{j=0}^{\infty} \frac{1}{j!} \int_{A^J} r_A^n(x_1,\ldots,x_n,\theta_1,\ldots,\theta_J) d\theta_1 \ldots d\theta_J$$

$$= \sum_{j=0}^{\infty} \frac{1}{j!} \int_{A^J} E[I^{n+J} e^{-I|A|}]\, d\theta_1 \ldots d\theta_J$$

$$= E\left[\sum_{j=0}^{\infty} \frac{1}{j!} \int_{A^J} I^{n+J} e^{-I|A|}\, d\theta_1 \ldots d\theta_J \right]$$

$$= E\left[I^n \sum_{j=0}^{\infty} \frac{1}{j!} |A|^j I^j e^{-I|A|} \right] = E[I^n] \tag{2.68}$$

Example 2.34 To define a Mixed Sample Process we first take Y to be a random variable with values in $\{0,1,\ldots\}$. Conditioned on $Y=k$ let $Z_1,\ldots,Z_k$ be uniformly and independently distributed on $[0,b]$ as occurrences of the mixed sample process X. Such processes are characterized by the fact that their distributions are invariant under measure preserving transformations. The absolute product densities are thus independent of the location of the points $x_1,\ldots,x_n$. That is they are of the form $r_A(x_1,\ldots,x_n)=f_A(n)$. These models are analyzed in Kallenberg [16].

Example 2.35 Cluster Processes A major class of point processes is the cluster point processes. Such processes have been used to model a wide variety of things from populations (Moyal [23]) to clustering of galaxies (Neyman and Scott [27]). A nice overview of examples of cluster process models is given in Neyman and Scott [28].

In general, a cluster process can be thought of as being generated in two steps as follows. First a point process X_1 is observed. This process generates the cluster centers. Next, for each cluster center x_i a new process $X_2(\cdot\,|\,x_i)$ is observed, giving the cluster members. The cluster process X is then the process $X(A) = \sum_{x_i \in X_1} X_2(A\,|\,x_i)$.

The probability generating functional corresponding to X is (by Theorem 2.29)

$$G(\xi) = E\Big\{ \prod_{y_i \in X} \xi(y_i) \Big\} \tag{2.69}$$

but $y_i \in X$ implies $y_i \in X_2(A\,|\,x_j)$ for some j, so

$$G(\xi) = E\Big\{ \prod_{x_j \in X_1} \Big(\prod_{y_i \in X_2(A\,|\,x_j)} \xi(y_i) \Big\} = G_1(G_2(\xi\,|\,x)). \tag{2.70}$$

Next we will consider some specific examples of cluster processes.

Example 2.36 a Let U be a stationary Poisson process on $\mathbb{R}^d$ with intensity λ and V be a point process satisfying $E[V(\mathbb{R}^d)] < \infty$. As above let u_i denote the random occurrences of U and take these as the cluster centers. Let $V_1, V_2, \ldots$ representing the cluster members, be point processes which are independent and identically distributed as V. The resulting cluster process X is then a Poisson Center Cluster Process with centers U and clusters V. X can be written as $X = \sum_{u_i \in U} V(A - u_i)$ and the probability generating functional (2.70) takes the form

$$G(\xi) = \exp\left\{\int [G_V(T_X\xi) - 1]\, \lambda\, dx\right\} \tag{2.71}$$

where T_X represents the translation operator $(T_X\xi)(y) = \xi(x+y)$.

Example 2.37 b This special case of the Poisson center cluster process was used by Neyman and Scott [27] to model clustering of galaxies. Here we consider a Poisson center cluster process where about each of the cluster centers a random number of cluster members are independently distributed according to a common distance distribution F. The resulting point process X is referred to as a Neyman Scott Cluster Process.

If we let G_1 be the probability generating function of the cluster size then the probability generating functional of X is

$$G(\xi) = \exp\left\{\int \left\{G_1\left[\int (T_X\xi)(r)\, dF(r)\right] - 1\right\} \lambda\, dx\right\} \tag{2.72}$$

III POSITIVE DEPENDENCE FOR POINT PROCESSES

3.1 Definitions of Positive Dependence

The definitions of positive dependence for Bernoulli random variables are easily extended to the point process case. The first definition we give was first stated by Burton and Waymire [6].

Definition 3.1 X satisfies the strong FKG inequalities if for all sets $A \in \hat{\mathfrak{B}}^d$ there exists a version of the absolute product densities such that

$$r_A^n(x_1, \ldots, x_n) r_A^{j-i+1}(x_i, \ldots, x_j) \geq r_A^j(x_1, \ldots, x_j) r_A^{n-i+1}(x_i, \ldots, x_n) \tag{3.1}$$

for all $x_1, \ldots, x_n \in A$, $1 \leq i \leq j \leq n$.

Knowing that X satisfies the strong FKG inequalities gives us some information about the structure of the point process X. For example the following two theorems tell us something about types of configurations we can expect for point processes satisfying the strong FKG inequalities along with other conditions. From now on, whenever it is clear what is intended, we will omit the superscript indicating the number of arguments for the absolute product densities.

Theorem 3.2 If X satisfies the strong FKG inequalities then $r_B(\varnothing) > 0$.

Proof: The proof is by induction. Let $\underline{x} = x_1$, $\underline{y} = y_1$, $x_1 \neq y_1$ and let $r_B^{(0)}$ be a version of the absolute product densities for which the strong FKG inequalities hold and $r_B^{(0)}(\varnothing) = 0$. Then

$$0 = r_B^{(0)}(x_1, y_1)\, r_B^{(0)}(\varnothing) \geq r_B^{(0)}(x_1)\, r_B^{(0)}(y_1) \tag{3.2}$$

which implies that either $r_B^{(0)}(x_1) = 0$ or $r_B^{(0)}(y_1) = 0$. Thus, there is at most one point x_1^1 for which $r_B^{(0)}(x_1^1) > 0$. It then follows that

$$P(X(B) = 1) = \int_B r_B^{(0)}(x)\,dx = 0. \tag{3.3}$$

Before proceeding, we choose a slightly different version of the absolute product densities, $r_B^{(1)}$, differing from $r_B^{(0)}$ on a set of measure 0. $r_B^{(1)}$ is the product density that agrees with $r_B^{(0)}$ except that $r_B^{(1)}$ is 0 whenever it is evaluated at any configuration which includes the point x_1^1. That is, $r_B^{(1)}(\underline{x}) = 0$ for any $\underline{x}$ of the form $\underline{x} = (x_1,\ldots,x_n)$ where $x_i = x_1^1$ for some i. Note that $r_B^{(1)}(\underline{x})$ still satisfies the strong FKG inequalities and that $r_B^{(1)}(\underline{x}) = 0$ for any singleton $\underline{x} = x_i$.

Now suppose that we have $\underline{x} = (x_1,x_2)$ and $\underline{y} = (y_1,y_2)$ where x_1,x_2,y_1 and y_2 are all distinct. Then by the strong FKG inequalities

$$0 = r_B^{(1)}(\underline{x},\underline{y})\,r_B^{(1)}(\emptyset) \geq r_B^{(1)}(\underline{x})\,r_B^{(1)}(\underline{y}) \tag{3.4}$$

which implies that either $r_B^{(1)}(\underline{x}) = 0$ or $r_B^{(1)}(\underline{y}) = 0$. That is, no two disjoint pairs can both have positive absolute product densities $r_B^{(1)}$. If x_1,x_2,y_1,y_2 are not all distinct, say $x_2 = y_1$, then

$$0 = r_B^{(1)}(x_1,x_2,y_2)\,r_B^{(1)}(x_2) \geq r_B^{(1)}(x_1,x_2)\,r_B^{(1)}(y_1,y_2) \tag{3.5}$$

so that either $r_B^{(1)}(x_1,x_2) = 0$ or $r_B^{(1)}(y_1,y_2) = 0$. Thus no two pairs with one component in common can both have positive absolute product densities $r_B^{(1)}$. Putting together (3.4) and (3.5) we find that there is at most one pair (x_1^2,x_2^2) for which $r_B^{(1)}(x_1^2,x_2^2) \neq 0$. Thus

$$P(X(B) = 2) = \frac{1}{2}\int_B\int_B r_B^{(1)}(x,y)\,dx\,dy = 0. \tag{3.6}$$

We again choose a new version of the absolute product densities, $r_B^{(2)}$, where $r_B^{(2)}$ agrees with $r_B^{(1)}$ except for on configurations containing (x_1^2,x_2^2). We will now have $r_B^{(2)}(\underline{x}) = 0$ for any $\underline{x} = (x_1,\ldots,x_n)$ where $x_i = x_1^2$ and $x_j = x_2^2$ for some i,j. Thus $r_B^{(2)}(\underline{x})=0$ for any configuration $\underline{x}$ containing one component or two components and $r_B^{(2)}$ still satisfies the strong FKG inequalities.

Inductively assume that we have chosen $r_B^{(k)}$ as above so that it agrees with $r_B^{(k-1)}$ except for on configurations containing $(x_1^{k-1},\ldots,x_{k-1}^{k-1})$. Thus $r_B^{(k-1)}$ satisfies the strong FKG inequalities and $r_B^{(k-1)}(\underline{x})=0$ for all configurations $\underline{x}=(x_1,\ldots,x_j)$ where $j\leq k$.

Now let $\underline{x}=(x_1,\ldots,x_{k+1})$ and let $\underline{y}=(y_1,\ldots,y_{k+1})$, $\underline{x}$ not identically equal to $\underline{y}$. Then, since $\underline{x}\wedge\underline{y}$ has k or fewer components,

$$0=r_B^{(k)}(\underline{x}\wedge\underline{y})\,r_B^{(k)}(\underline{x}\vee\underline{y})\geq r_B^{(k)}(\underline{x})\,r_B^{(k)}(\underline{y}) \tag{3.7}$$

Thus, as before there is at most one $\underline{x}=(x_1^{k+1},\ldots,x_{k+1}^{k+1})$ so that $r_B^{(k)}(x_1^{k+1},\ldots,x_{k+1}^{k+1})=0$ and we have

$$P(X(B)=k+1)=\int_B\cdots\int_B r_B^{(k)}(x_1,\ldots,x_k)\,dx_1\ldots dx_k=0. \tag{3.8}$$

As above, we choose a new version of the product densities, $r_B^{(k+1)}$ agreeing with $r_B^{(k)}$ except on the set of measure zero consisting of the configurations containing $(x_1^{k+1},\ldots,x_{k+1}^{k+1})$. Thus $r_B^{(k+1)}$ satisfies the strong FKG inequalities and $r_B^{(k+1)}(\underline{x})=0$ for all configurations $\underline{x}$ containing $k+1$ or fewer components.

We have found by induction that $P(X(B)=n)=0$ for any n. This implies $r_B(\varnothing)=1$, which is a contradiction. Thus $r_B(\varnothing)>0$. □

The following theorem guarantees that for a point process X on $\mathbb{R}^d$ which satisfies the strong FKG inequalities, if any individual points have positive probability of occurring (positive absolute product densities) then any configuration of those points has positive absolute product densities.

i.e. all configurations are possible.

Theorem 3.3 Suppose that X satisfies the strong FKG inequalities then there is a subset $A \subset D$ so that $P[X(A) = 0] = 1$ and all configurations are possible on $D\backslash A$, in the sense that if $B \subset D\backslash A$ is a bounded Borel set, then there is a version of $r_B(x_1,\cdots,x_n)$ that is strictly positive for all distinct $x_1,\cdots,x_n$ in B.

Proof We may assume that all the product densities and absolute product densities are Borel measurable. The relation

$$p(x) = \sum_{n=0}^{\infty}\frac{1}{n!}\int_{B^n} r_B(x,\underline{y})\, d\underline{y} \tag{3.9}$$

of course holds only a.e. on B, since r_B is unique up to a.e. equivalence classes. If (3.9) does not hold for x, redefine $r_E(x,\underline{y}) = 0$ for all $\underline{y} \in B^n$ and all n. The strong FKG inequalities still hold and we may redefine $p(x) = 0$ so that (3.9) holds everywhere. Furthermore this will change p only on a set of measure 0 (even as we vary B). This is because the expression (3.9) is independent of B by Theorem 2.18.

Now let $A = \{ x \in D \mid p(x) = 0 \}$ and let B be a bounded Borel subset of $D\backslash A$. We show that if $(x_1,\cdots,x_m) \in B^m$ has distinct coordinates then $r_B(x_1,\cdots,x_m) > 0$. Suppose otherwise. Repeated applications of the strong FKG inequalities gives

$$0 = r_B(x_1,\cdots,x_m)r_B(\emptyset)^{m-1} \geq r_B(x_1)r_B(x_2)\cdots r_B(x_m) \tag{3.10}$$

This means that there is an x_i so that $r_B(x_i) = 0$. We rename $x_i = x$. If $\underline{y}$, $\underline{z} \in B^n$ have distinct coordinates and if for each $i = 1,\cdots,n$ $x \neq y_i \neq z_i \neq x$ then

$$0 \;=\; r_B(x,\, \underline{y},\, \underline{z})r_B(x) \;\geq\; r_B(x,\, \underline{y})r_B(x,\, \underline{z}). \tag{3.11}$$

Thus at most one of $r_B(x, \underline{y})$ and $r_B(x, \underline{z})$ can be strictly positive. This implies $r_B(x, \underline{y}) = 0$ for $\underline{y}$ - a.e. on B^n. But in view of (3.9) this means that $p(x) = 0$ so $x \in A$, a contradiction. □

<u>Definition</u> <u>3.4</u> X has <u>Positively</u> <u>Correlated</u> <u>Increasing</u> <u>Cylinder</u> <u>Sets</u> (X has PCIC) if

$$p(x_1,\ldots,x_n) \geq p(x_1,\ldots,x_i)\, p(x_{i+1},\ldots,x_n). \tag{3.12}$$

<u>Definition</u> <u>3.5</u> X has <u>Conditionally</u> <u>Positively</u> <u>Correlated</u> <u>Increasing</u> <u>Cylinder</u> <u>Sets</u> (X has CPCIC) if

$$p(x_1,\ldots,x_n)\, p(x_i,\ldots,x_j) \;\geq\; p(x_1,\ldots,x_j)\, p(x_i,\ldots,x_n) \tag{3.13}$$

for $1 \leq i \leq j \leq n$.

<u>Definition</u> <u>3.6</u> X is <u>Associated</u> if $\mathrm{Cov}(F(X),G(X)) \geq 0$ for all pairs of functions $F,G:N \rightarrow \mathbb{R}$ that are increasing, measurable and bounded (where increasing means increasing with respect to the ordering on N given by $\mu < \nu$ if $\mu(B) < \nu(B)$ for all $B \in \mathcal{B}^d$)

Burton and Waymire [5] showed that Definition 3.6 is equivalent to the family of random variables $\{X(B) \mid B \in \mathcal{B}^d\}$ being associated. That is, all finite subsets of $\{X(B) \mid B \in \mathcal{B}^d\}$ are associated in the sense of Definition 1.5.

To see how these definitions are a natural extension of the definitions of positive dependence for Bernoulli random variables consider the following method of approximating a point process X with well defined densities on the interval [0,1]. For each n and $k = 1,\ldots,n$ define

$$X_k^{(n)} = \begin{cases} 1 \text{ if there is an occurrence of X in } [\,\frac{k-1}{n}, \frac{k}{n}\,] \\ 0 \text{ if there is no occurrence of X in } [\,\frac{k-1}{n}, \frac{k}{n}\,] \end{cases} \tag{3.14}$$

That is, $X_k^{(n)} = \min(1, X[\frac{k-1}{n}, \frac{k}{n}])$ and each $X_k^{(n)}$ is a Bernoulli random variable. $\underline{X}^{(n)} = (X_1^{(n)}, \ldots, X_k^{(n)})$ approximates X and $\underline{X}^{(n)}$ converges to X in distribution. Thus, for example, FKG for a point process X can be thought of as a limiting condition of FKG for the random vectors $\underline{X}^{(n)}$. In the limit we indicate just where the ones are, ones indicating the occurrence of points. Burton and Waymire used a similar approximation technique for point processes on $\mathbb{R}^d$ to prove the following theorem [5].

<u>Theorem</u> <u>3.7</u> If a point process X is completely regular and satisfies the strong FKG inequalities for all cubes $A \in \mathbb{R}^d$ then X is associated.

<u>Remark</u> In their original version of the above theorem Burton and Waymire required that the absolute product densities be piecewise continuous. This condition was used to create an appropriate partition in order to approximate the integral of the absolute product density by a Riemann sum. The following lemma, however, shows that the piecewise continuous condition is not necessary.

<u>Lemma</u> <u>3.8</u> If $A \subseteq \mathbb{R}^m$ is a bounded rectangular box and f is a non-negative real valued function on A for which the Lebesgue integral $\int_A f(\underline{x})\,d\underline{x}$ is finite then there exists a sequence $N_k \to \infty$ so that if P_k is the even partition of A into $(N_k)^m$ rectangles of equal measure, $P_k = \{\Delta_i^{(k)}\}$ $i = 1, \ldots, (N_k)^m$ and an $x_i^{(k)} \in \Delta_i^{(k)}$ so that $\lim_{k \to \infty} \sum f(x_i^{(k)})\,|\Delta_i^{(k)}| = \int_A f(\underline{x})\,d\underline{x} < \infty$.

<u>Proof</u> Given $\epsilon > 0$ choose k so that $\frac{1}{k} < \epsilon$. Let $B_M = \{\underline{x} \mid f(\underline{x}) > M\}$ where M is chosen large enough so that $|B_M| < \frac{1}{6k}$ and $\int_{B_M} f(\underline{x})\,d\underline{x} < \frac{1}{30k}$. Let $f_M(\underline{x}) = \min(f(\underline{x}), M)$ and note that $\int_{B_M} f(\underline{x})\,d\underline{x} = \int_A (f(\underline{x}) - f_M(\underline{x}))\,d\underline{x}$. Choose a

continuous function $g:A \to [0,\infty)$ such that $|g(\underline{x})| \leq M$, $\int_A |g(\underline{x}) - f_M(\underline{x})|\,d\underline{x} < \frac{1}{6k}$ and if $U = \{\underline{x} \notin B_M \mid |g(\underline{x}) - f(\underline{x})| > \frac{1}{6|A|k}\}$ then $|U| < \frac{1}{12Mk}$. Choose N_k so that all Riemann sums based on $P_k = \{\Delta_i^{(k)}\}$ are closer to $\int_A g(\underline{x})\,d\underline{x}$ than $\frac{1}{6k}$.

Let G be the set of all $\Delta_i^{(k)} \in P_k$ such that there exists an $\underline{x} \in \Delta_i^{(k)} \setminus (B_M \cup U)$. Let this $\underline{x}$ be the choice for $x_i^{(k)}$. Let L be the set of all $\Delta_i^{(k)} \in P_k$ such that $\Delta_i^{(k)} \notin G$ but there is an $\underline{x}$ in $\Delta_i^{(k)} \setminus B_M$. Let this $\underline{x}$ be the choice for $x_i^{(k)}$. Let B be the set of all $\Delta_i^{(k)}$ which are not in G or L. i.e. $B = \{\Delta_i^{(k)} \mid \Delta_i^{(k)} \subseteq B_M\}$. In this case choose $x_i^{(k)}$ so that $f(x_i^{(k)}) < \inf_{x \in \Delta_i^{(k)}} f(\underline{x}) + 1$.

For the remainder of the proof we will omit the superscript k. Now we have

$$
\begin{aligned}
\Big| \int_A f(\underline{x})\,d\underline{x} - \sum f(x_i)|\Delta_i| \Big| &< \Big| \int_A f(\underline{x})\,d\underline{x} - \int_A f_M(\underline{x})\,d\underline{x} \Big| \\
&\quad + \Big| \int_A f_M(\underline{x})\,d\underline{x} - \int_A g(\underline{x})\,d\underline{x} \Big| + \Big| \int_A g(\underline{x})\,d\underline{x} - \sum g(x_i)|\Delta_i| \Big| \\
&\quad + \Big| \sum g(x_i)|\Delta_i| - \sum f(x_i)|\Delta_i| \Big| \\
&< \frac{1}{30k} + \frac{1}{6k} + \frac{1}{6k} + \sum_{\Delta_i \in G} |g(x_i) - f_M(x_i)|\,|\Delta_i| + \sum_{\Delta_i \in L} |g(x_i) - f_M(x_i)|\,|\Delta_i| \\
&\quad + \sum_{\Delta_i \in B} |g(x_i)|\,|\Delta_i| + \sum_{\Delta_i \in B} |f(x_i)|\,|\Delta_i| \\
&< \frac{1}{30k} + \frac{1}{3k} + \sum_{\Delta_i \in G} \frac{1}{6|A|k}|\Delta_i| + \sum_{\Delta_i \in L} 2M|\Delta_i| + M|B_M| \\
&\quad + \sum_{\Delta_i \in B} (1 + \inf_{x \in \Delta_i} f(x))|\Delta_i| \\
&< \frac{11}{30k} + \frac{1}{6|A|k}|A| + 2M(|U| + |B_M|) + M|B_M| + \sum_{\Delta_i \in B} (1 + \inf_{x \in \Delta_i} f(x))|\Delta_i| \\
&< \frac{11}{30k} + \frac{1}{6k} + 2M\Big(\frac{1}{12Mk} + \frac{1}{6k}\Big) + M\frac{1}{6k} + \sum_{\Delta_i \in B} (1 + \inf_{x \in \Delta_i} f(x))|\Delta_i| \\
&< \frac{21}{30k} + \frac{M}{2k} + \sum_{\Delta_i \in B} (1 + \inf_{x \in \Delta_i} f(x))|\Delta_i| \qquad (3.15)
\end{aligned}
$$

Note: (1) $\frac{M}{2k} = 3M\frac{1}{6k} < 3\int_{B_M} f(\underline{x})\,d\underline{x} < 3\frac{1}{30k} = \frac{3}{30k}$ and

(2) $\sum_{\Delta_i \in B} (1 + \inf_{x\in\Delta_i} f(x))|\Delta_i| = \sum_{\Delta_i \in B} |\Delta_i| + \sum_{\Delta_i \in B} (\inf_{x\in\Delta_i} f(x))|\Delta_i|$

$$< |B_M| + \int_{B_M} f(\underline{x})\,d\underline{x} < \frac{1}{6k} + \frac{1}{30k}$$

So that (3.15) is

$$< \frac{21}{30k} + \frac{3}{30k} + \frac{1}{6k} + \frac{1}{30k} = \frac{1}{k} < \epsilon. \qquad \square$$

Example 3.9 The Poisson point process is easily seen to satisfy the strong FKG inequalities. We earlier calculated the absolute product densities (2.65) and found that $r_A(x_1,\ldots,x_n) = f(x_1)\cdot\ldots\cdot f(x_n)\exp\left[-\int_A f(\theta)\,d\theta\right]$ so that

$$r_A^n(x_1, \ldots, x_n)r_A^{j-i}(x_i, \ldots, x_j)$$

$$= \left[f(x_1)\cdot\ldots\cdot f(x_n)\exp\left[-\int_A f(\theta)\,d\theta\right]\right]\left[f(x_i)\cdot\ldots\cdot f(x_j)\exp\left[-\int_A f(\theta)\,d\theta\right]\right]$$

$$= \left[f(x_1)\cdot\ldots\cdot f(x_j)\right]\left[f(x_{j+1})\cdot\ldots\cdot f(x_n)\right]\left[f(x_i)\cdot\ldots\cdot f(x_j)\right]\left[\exp\left[-\int_A f(\theta)\,d\theta\right]\right]^2$$

$$= \left[f(x_1)\cdot\ldots\cdot f(x_j)\right]\left[f(x_i)\cdot\ldots\cdot f(x_n)\right]\left[\exp\left[-\int_A f(\theta)\,d\theta\right]\right]^2$$

$$= r_A^j(x_1, \ldots, x_j)r_A^{n-i}(x_i, \ldots, x_n).$$

In fact, we have found that for the Poisson Process, equality holds.

Example 3.10 An application of Schwarz's inequality (see Feller [10]) shows that the mixed Poisson process also satisfies the strong FKG inequalities. Recall that we found $r_A(x_1,\ldots,x_n) = E[I^n e^{-I|A|}]$ where I is a nonnegative random variable. Let $Y^k = I^k e^{-I|A|}$, $f(k) = E[Y^k]$ and $g(k) = \log(f(k))$. Let $b > a$. By Schwarz's inequality $E[Y^{\frac{a+b}{2}}] \leq (E[Y^a])^2(E[Y^b])^2$.

i.e. $f(\frac{a+b}{2}) \leq (f(a)f(b))^{\frac{1}{2}}$. Thus, $g(\frac{a+b}{2}) \leq \frac{1}{2}(g(a)+g(b))$. i.e. the moments of Y are log convex.

An alternative expression for convexity says that (for $c>0$) $g(b+c)-g(a+c) \geq g(b)-g(a)$ so that $g(b)+g(a+c) \leq g(b+c)+g(a)$. Equivalently, in terms of f we have $f(b)f(a+c) \leq f(b+c)f(a)$. Now let $a=j-i$, $b=n-i$, and $c=i$ to get the strong FKG inequalities.

Example 3.11 Burton and Waymire [5] showed that Poisson center cluster processes are associated. As we will see later, these need not satisfy the strong FKG inequalities.

3.2 The Relationships Between Positive Dependence Definitions

We must be careful with approximations such as the one suggested in (3.14). Based on the comparisons of the definitions one might expect a theorem for point processes analogous to Theorem 1.12. In particular it seems reasonable to expect that strong FKG and CPCIC are equivalent. They are not however, and the relationship between the positive dependence definitions is a bit more complicated.

Theorem 3.12 If X is completely regular then the following implications, and no others, hold

$$\begin{tikzcd}
 & \text{X satisfies the strong FKG inequalities} \arrow[dl, Rightarrow] \arrow[dr, Rightarrow] & \\
\text{X is CPCIC} \arrow[dr, Rightarrow] & & \text{X is associated} \arrow[dl, Rightarrow] \\
 & \text{X has PCIC} &
\end{tikzcd}$$

Note that if the product densities are strictly positive for all configurations, CPCIC and conditionally positively correlated (CPC) are equivalent since we could write CPC as $p(x, z| \underline{y}) \geq p(x| \underline{y})p(z| \underline{y})$ where $\underline{y} = y_1, \ldots, y_n$. Thus $\frac{p(x, z, \underline{y})}{p(\underline{y})} \geq \frac{p(x, \underline{y})}{p(\underline{y})} \frac{p(z, \underline{y})}{p(\underline{y})}$, or equivalently, $p(x, \underline{y}, z)p(\underline{y}) \geq p(x, \underline{y})p(\underline{y}, z)$. If $\underline{x} = x_1, \ldots, x_{i-1}$, $\underline{y} = y_i, \ldots, y_j$, and $\underline{z} = z_{j+1}, \ldots, z_n$ a simple induction argument extends this to $p(\underline{x}, \underline{y}, \underline{z})p(\underline{y}) \geq p(\underline{x}, \underline{y})p(\underline{z}, \underline{y})$ which is CPCIC. Thus adding this further restriction on PCIC in the point process case still does not give a definition equivalent to strong FKG.

Note also that if X satisfies the strong FKG inequalities, then the conditional absolute product densities will also satisfy (3.1). Thus all conditional distributions of X will also be associated by Theorem 3.12.

By the following theorem, the cumulant densities defined in section 2.3 also play a role in describing positive dependence properties.

Theorem 3.13 If a point process X has cumulant densities which are always non-negative then X has PCIC, but not conversely.

3.3 Proofs and Examples

In this section we give the proofs of Theorems 3.12 and 3.13 and examples which show the negative implications.

Proof of Theorem 3.12 (1) X satisfies the strong FKG inequalities implies X is associated by Theorem 3.7. Furthermore, by the above note, if X satisfies the strong FKG inequalities then all conditional distributions of X are also associated.

(2) X has CPCIC implies X has PCIC is immediate.

(3) We show that X satisfies the strong FKG inequalities implies X has CPCIC. Let $\underline{x} = (x_1,\ldots,x_n)$ and suppose (with no loss of generality) that each x_i is in D\A (as defined in Theorem 3.3). That is, $p(x_i) > 0$. In fact all of the points discussed in this part of the proof are assumed to be in D\A. Let $r(\underline{x}|\underline{y})$ be the conditional absolute product density, conditioned on the fact that there are known to be point occurrences at $\underline{y} = (y_1,\ldots,y_k)$.

Define $\Phi_{\underline{z}}$ by $\Phi_{\underline{z}}(\underline{x}|\underline{y}) = \frac{r(\underline{x},\underline{z}|\underline{y})}{r(\underline{x}|\underline{y})}$ where $\underline{z} = (z_1,\ldots,z_j)$ (so that $(\underline{x},\underline{z}) = (x_1,\ldots x_n,z_1,\ldots,z_j)$). Note that $\Phi_{\underline{z}}$ is an increasing function of $\underline{x}$ for fixed $\underline{y}$ since X satisfies the strong FKG inequalities implies $r(\underline{x},\underline{y},\underline{z})r(\underline{y}) \geq r(\underline{x},\underline{y})r(\underline{y},\underline{z}) \Rightarrow r(\underline{x},\underline{z}|\underline{y})r(\underline{y}|\underline{y}) \geq r(\underline{x}|\underline{y})r(\underline{z}|\underline{y}) \Rightarrow \frac{r(\underline{x},\underline{z}|\underline{y})}{r(\underline{x}|\underline{y})} \geq \frac{r(\underline{z}|\underline{y})}{r(\underline{y}|\underline{y})}$. Then,

$$E[\Phi_{\underline{z}}] = \sum_{n=0}^{\infty} \frac{1}{n!} \int \cdots \int \Phi_{\underline{z}}(\underline{x}|\underline{y}) r(\underline{x}|\underline{y}) d\underline{x}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int \cdots \int r(\underline{x},\underline{z}|\underline{y}) d\underline{x} = p(\underline{z}|\underline{y}) \tag{3.16}$$

$$E[\Phi_{\underline{z}}\Phi_{\underline{w}}] = \sum_{n=0}^{\infty} \frac{1}{n!} \int \cdots \int \Phi_{\underline{z}}(\underline{x}|\underline{y}) \Phi_{\underline{w}}(\underline{x}|\underline{y}) r(\underline{x}|\underline{y}) d\underline{x}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int \cdots \int \frac{r(\underline{x},\underline{z}|\underline{y}) r(\underline{x},\underline{w}|\underline{y})}{r(\underline{x}|\underline{y})} d\underline{x}$$

$$\leq \sum_{n=0}^{\infty} \frac{1}{n!} \int \cdots \int r(\underline{x},\underline{z},\underline{w}|\underline{y}) d\underline{x} \quad \text{(by the strong FKG inequalities)}$$

$$= p(\underline{z},\underline{w}|\underline{y}) \tag{3.17}$$

Since X satisfies the strong FKG inequalities its conditional distributions are associated, so $\mathrm{Cov}(\Phi_{\underline{z}},\Phi_{\underline{w}}) \geq 0$. i.e. $E[\Phi_{\underline{z}}\Phi_{\underline{w}}] \geq E[\Phi_{\underline{z}}]E[\Phi_{\underline{w}}]$ or $p(\underline{z},\underline{w}|\underline{y}) \geq p(\underline{z}|\underline{y})p(\underline{w}|\underline{y})$ which implies $p(\underline{y})^2 p(\underline{z},\underline{w}|\underline{y}) \geq p(\underline{y})^2 p(\underline{z}|\underline{y})p(\underline{w}|\underline{y})$, so that $p(\underline{y})p(\underline{w},\underline{z}) \geq p(\underline{z},\underline{y})p(\underline{w},\underline{y})$. Thus X has CPCIC.

(4) Lastly we show that if X is associated then X has PCIC. Let A_J^k represent the event that the interval of length Δx about x_i is occupied for $i=j,\ldots,k$. Let $f(X) = 1_{A_1^k}(X)$ and $g(X) = 1_{A_{k+1}^{n+k}}(X)$. Since f and g are increasing functions of X and X is associated, $0 \leq E[f(X)g(X)] - E[f(X)]E[g(X)]$ $= p(x_1,\ldots,x_{n+k})(\Delta x)^{n+k} - p(x_1,\ldots,x_k)(\Delta x)^k p(x_{k+1},\ldots,x_{k+n})(\Delta x)^n + o(|\Delta x|^{n+k})$.

Thus, $p(x_1,\ldots,x_{n+k}) \geq p(x_1,\ldots,x_k)p(x_{k+1},\ldots,x_{k+n})$, that is, X has PCIC. To complete the proof we will find examples of to point processes, one of which has CPCIC but is not associated and one which is associated but does not have CPCIC. □

Example 3.14 In this first example, we will show that X having CPCIC does not imply X is associated. The idea is that we know the Poisson point process satisfies the strong FKG inequalities and so is associated, CPCIC, has PCIC, etc. It's product densities satisfy a nice convexity condition which implies CPCIC. Thus, we adjust our densities in such a way as to preserve this convexity condition while at the same time the adjustment changes the process enough so that it is not associated. The process we end up with turns out to be a mixed sample process X on a bounded interval $B = [0,b]$.

For the actual construction of the densities, first note that if X is a point process with product densities p and absolute product densities $r_B^k(x_1,\ldots,x_k) = f_B(k)$, that is the absolute product density depends only on the number of occurrences and not on their locations (e.g. Poisson), then

$$g(k) = p(x_1,\ldots,x_k) = \sum_{n=0}^{\infty} \frac{1}{n!} \int_B \cdots \int_B f_B(k+n)\, dy_1\ldots dy_n = \sum_{n=0}^{\infty} \frac{b^n}{n!} f_B(k+n) \quad (3.18)$$

Conversely, if $p(x_1,\ldots,x_n) = g(k)$ we find that

$$f_B(k) = r_B^k(x_1,\ldots,x_k) = \sum_{n=0}^{\infty}(-1)^n \frac{b^n}{n!} g(k+n) \tag{3.19}$$

We will use these formulas to determine from a given choice of function g representing product densities, the corresponding absolute product densities. In order for given functions $g(k)$ to be product densities with corresponding absolute product densities given by (3.19) the following three conditions must be satisfied:

$$\begin{aligned} &(1)\ f_B(k) \geq 0 \text{ for all } k. \\ &(2)\ \sum_{n=0}^{\infty} \frac{b^n}{n!} f_B(n) = 1. \\ &(3)\ g(k) = \sum_{n=0}^{\infty} \frac{b^n}{n!} f_B(k+n). \end{aligned} \tag{3.20}$$

Condition 2 says that the total probability of all configurations on $[0,b]$ must be one, i.e. $\sum_{n=0}^{\infty} P[\text{there are exactly } n \text{ points in } [0,b]] = 1$. Condition 3 is the condition that allows us to use the inversion formula to get $g(k)$ back from the functions $f(k)$. Condition 1 guarantees that any number of points has a non-negative probability of occurring in $[0,b]$.

<u>Lemma</u> <u>3.15</u> If g satisfies

(a) $\sum_{n=0}^{\infty}(-1)^n \frac{b^n}{n!} g(k+n) \geq 0$ for $k = 0,1,2,\ldots$

(b) $g(0) = 1$

(c) $g(k) \geq 0$ for $k = 0,1,2,\ldots$

(d) $\sum_{L=0}^{\infty} \frac{(2b)^L}{L!} g(k+L) < \infty$ for $k = 0,1,2,\ldots$

then g determines product densities with corresponding absolute product densities given by (3.19).

<u>Proof</u> (a) gives us condition 1 of (3.20), and given condition 3, (b)

gives condition 2. Thus we must only show that condition 3 is satisfied.

$$\sum_{n=0}^{\infty} \frac{b^n}{n!} f_B(k+n) = \sum_{n=0}^{\infty} \frac{b^n}{n!} \sum_{j=0}^{\infty} (-1)^j \frac{b^j}{j!} g(j+k+n)$$

$$= \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} (-1)^j \frac{b^n}{n!} \frac{b^j}{j!} g(j+k+n). \qquad (3.21)$$

The sum (3.21) can be rearranged as long as we have absolute convergence of the corresponding double sum i.e. as long as

$$\sum_{n,j} \frac{b^{n+j}}{n!\, j!} g(j+k+n) < \infty. \qquad (3.22)$$

Since each term of (3.22) is non-negative, it converges if and only if (summing over diagonals)

$$\sum_{L=0}^{\infty} \frac{g(L+k)}{L!} \sum_{n=0}^{L} \frac{L!}{n!(L-n)!} b^L < \infty, \qquad (3.23)$$

or, equivalently,

$$\sum_{L=0}^{\infty} \frac{g(L+k)}{L!} \sum_{n=0}^{L} \binom{L}{n} b^L < \infty. \qquad (3.24)$$

By the binomial theorem $\sum_{n=0}^{L} \binom{L}{n} b^L = (b+b)^L$, so (3.24) becomes

$$\sum_{L=0}^{\infty} \frac{(2b)^L}{L!} g(k+L) < \infty \qquad (3.25)$$

which is given by (d). Thus we may rearrange (3.21), again summing over the diagonals

$$\sum_{n=0}^{\infty} \sum_{j=0}^{\infty} (-1)^j \frac{b^n}{n!} \frac{b^j}{j!} g(j+k+n) = \sum_{L=0}^{\infty} \sum_{n=0}^{L} \frac{(-1)^{L-n}}{L!} \frac{L!}{n!(L-n)!} b^L g(L+k)$$

$$= \sum_{L=0}^{\infty} b^L \frac{g(L+k)}{L!} \sum_{n=0}^{L} \binom{L}{n} (-1)^{L-n}. \qquad (3.26)$$

By the binomial theorem, $\sum_{n=0}^{L} \binom{L}{n} (-1)^{L-n} = \sum_{n=0}^{L} \binom{L}{n} (-1)^{L-n}(1)^n = (1-1)^L$, so we get 0 here except for when $L=0$. Thus, (3.26) becomes

$$\frac{b^0}{0!} g(k+0)\cdot 1 = g(k) \qquad \square$$

Any function g which is nonnegative and bounded by an exponential will satisfy condition 3 by Lemma 3.15. For this example we must also choose g so that product densities satisfy the inequalities giving that X has CPCIC. That is, we need $p(x_1,\ldots,x_n)p(x_i,\ldots,x_j) \geq p(x_1,\ldots,x_j)p(x_i,\ldots,x_n)$ or, in terms of g, $g(n)g(j-i) \geq g(j)g(n-i)$.

<u>Lemma 3.16</u> $g(n)g(j-i) \geq g(j)g(n-i)$ if and only if $g(n+1)g(n-1) \geq [g(n)]^2$.

<u>Proof</u> That $g(n)g(j-i) \geq g(j)g(n-i)$ implies $g(n+1)g(n-1) \geq [g(n)]^2$ is obvious. Assume that $g(n+1)g(n-1) \geq [g(n)]^2$. If we assume also that $g(k) > 0$ for all k our assumption is equivalent to

$$\frac{g(n+1)}{g(n)} \geq \frac{g(n)}{g(n-1)} \tag{3.27}$$

Note that in each of the above fractions the argument in the numerator and the denominator differ by exactly one. Applying our original assumption and dividing again we also know that

$$\frac{g(n+2)}{g(n+1)} \geq \frac{g(n+1)}{g(n)}. \tag{3.28}$$

Thus from (3.27) and (3.28) it follows that

$$\frac{g(n+2)}{g(n+1)} \geq \frac{g(n)}{g(n-1)} \tag{3.29}$$

or, equivalently,

$$\frac{g(n+2)}{g(n)} \geq \frac{g(n+1)}{g(n-1)} \tag{3.30}$$

so that the desired inequality holds for $i = k \leq 2$.

Inductively assume that the inequality holds for any difference $k \leq i-1$. Then, since $i + j \geq n + 1$, $n - j \leq i - 1$ so that

$$\frac{g(n)}{g(j)} \geq \frac{g(n-i+1)}{g(j-i+1)} \tag{3.31}$$

and

$$\frac{g(n-i+1)}{g(j-i+1)} \geq \frac{g(n-i)}{g(j-i)} \tag{3.32}$$

so that

$$\frac{g(n)}{g(j)} \geq \frac{g(n-i)}{g(j-i)} \tag{3.33}$$

or, equivalently,

$$\frac{g(n)}{g(n-i)} \geq \frac{g(j)}{g(j-i)}, \tag{3.34}$$

which is the desired result.

Now, suppose that there is an n_0 so that $g(n_0) = 0$. Then we can not divide as in the above argument. In this case, by our initial assumption, we will have

$$g(n_0+1)^2 \leq g(n_0)g(n_0+2) \tag{3.35}$$

which implies that $g(n_0+1) = 0$. Inductively, this implies that $g(n) = 0$ for all $n \geq n_0$. We know that $g(0) = 1$. Now if $n_0 > 1$ then $g(n_0-1)^2 \leq g(n_0)g(n_0-2)$, by assumption, implies that $g(n_0-1) = 0$. Inductively, this implies that for every n with $0 < n \leq n_0$ we have $g(n) = 0$. Thus if there is an n_0 with $g(n_0) = 0$ then the densities correspond to a process with no points. Such a process clearly has CPCIC. In fact, it satisfies the FKG inequalities. □

We also want a condition which will guarantee that the process is not associated. Note that sets A and B are positively correlated (i.e. $Cov(1_A,1_B)\geq 0$) if and only if $Cov(1_{A^c},1_{B^c})\geq 0$. In fact these covariances are equal since $E[1_{A^c}\cdot 1_{B^c}]-E[1_{A^c}]E[1_{B^c}] \quad =P((A\cup B)^c)-P(A^c)P(B^c)= 1-P(A\cup B)-(1-P(A))(1-P(B))=1-P(A\cup B)-1+P(A)+P(B)-P(A)P(B)=P(A)+P(B)-P(A\cup B)-P(A)P(B)=P(A\cap B)-P(A)P(B)=Cov(1_A,1_B)$. Let A be the event that $X([0,\frac{b}{2}])=0$ and B the event that $X([\frac{b}{2},b])=0$. Then $A\cap B$ is the event that $X([0,b])=0$ and 1_{A^c}, 1_{B^c} are binary increasing functions of the random variables $X([0,\frac{b}{2}])$ and $X([\frac{b}{2},b])$ respectively. Thus, if $Cov(1_{A^c},1_{B^c})<0$ or equivalently $Cov(1_A,1_B)<0$ then X is not associated. That is, we will need to show

$$P\{X([0,b])=0\}<P\{X([0,\tfrac{b}{2}])=0\}P\{X([\tfrac{b}{2},b])=0\} \tag{3.36}$$

<u>Theorem</u> <u>3.17</u> Let $g(0)=1$, $g(1)=\alpha$, and $g(k)=\alpha^2\beta^{k-2}$ for $k\geq 3$, where $\alpha<\beta$. Then, there is a choice of α,β, and b (depending on α and β) so that the point process X with corresponding product densities $p(x_1,\ldots,x_k)=g(k)$ has CPCIC but is not associated.

<u>Proof</u> We check that X has CPCIC by using Lemma 3.16 and checking each case.

If $k=1$ $\quad g(2)g(0)=\alpha^2\cdot 1=\alpha^2$ and $g(1)^2=\alpha^2$.

If $k=2$ $\quad g(3)g(1)=\alpha^2\beta\cdot\alpha=\alpha^3\beta$ and $g(2)^2=\alpha^4$, but since $\beta>\alpha$, we have $\alpha^3\beta>\alpha^4$.

If $k>2$ $\quad g(k+1)g(k-1)=\alpha^2\beta^{k+1-2}\cdot\alpha^2\beta^{k-1-2}=\alpha^4\beta^{2k-4}$ whereas $g(k)^2=(\alpha^2\beta^{k-2})^2=\alpha^4\beta^{2k-4}$.

Thus, by Lemma 3.16 X has CPCIC.

To see that X is not associated we check that (3.36) holds. We also must check that α and β may be chosen so that X is well defined. In order to do this we first calculate the absolute product densities.

$$f_B(0) = \sum_{n=0}^{\infty}(-1)^n \frac{b^n}{n!}g(0+n) = 1 - b\alpha + \frac{\alpha^2}{\beta^2}[e^{-\beta b} - 1 + \beta b] \tag{3.37}$$

$$f_B(1) = \sum_{n=0}^{\infty}(-1)^n \frac{b^n}{n!}g(1+n) = \alpha + \frac{\alpha^2}{\beta}[e^{-\beta b} - 1] \tag{3.38}$$

$$f_B(k) = \sum_{n=0}^{\infty}(-1)^n \frac{b^n}{n!}g(k+n) = \alpha^2\beta^{k-2}[e^{-\beta b}], \text{ for } k \geq 2 \tag{3.39}$$

Let $\alpha=1$ and $\beta=2$ to get

$$f_B(0) = \frac{3}{4} - \frac{1}{2}b + \frac{1}{4}e^{-2b},\ f_B(1) = 1 + \frac{1}{2}[e^{-2b} - 1]$$

$$\text{and for } k \geq 2,\ f_B(k) = \frac{1}{4}2^k e^{-2b} \tag{3.40}$$

Note that $f_B(k) \geq 0$ for all values of b if $k \geq 1$. In order to be certain that these functions determine absolute product densities (with corresponding product densities g(k)) it remains to choose b so that $f_B(0) \geq 0$. In fact for the example we choose b so that $f_B(0)=0$. That such a b exists is guaranteed by the intermediate value theorem. We now have α, β, and b determining a well defined X. Finally, we show this X is not associated. Note that $P\big[X([0,b]) = 0\big] = f_B(0) = 0$, but $P\big[X([0,\frac{b}{2}]) = 0\big] = P\big[X([\frac{b}{2},b]) = 0\big] = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{b}{2}\right)^k f_B(k) > 0$ since $f_B(0)=0$ and $f_B(k) \geq 0$ for all $k \geq 1$. So the process satisfies (3.36) and is therefore not associated. □

This theorem, together with Theorem 3.7, also shows that X having CPCIC does not in general imply that X satisfies the strong FKG inequalities.

<u>Example</u> <u>3.18</u> This example shows that X being associated does not imply X has CPCIC. Consider a Neyman Scott cluster process on $\mathbb{R}$ where

the Poisson process of cluster centers is stationary with intensity $\lambda=1$. Assume also that there are exactly two points independently distributed at each cluster center according to a distance distribution F which has density f, continuous with the exception of a finite number of jump discontinuities. A sample realization might look like the diagram below.

● denotes a point occurrence (3.41)

* denotes locations of centers

This process was shown to be associated by Burton and Waymire [5] but can be adjusted (by choosing an appropriate distance distribution) so that if points x_1, x_2, and x_3 are chosen far enough apart the inequality $p(x_1,x_2,x_3)\,p(x_2) < p(x_1,x_2)\,p(x_2,x_3)$ holds, i.e. the process does not have CPCIC. This may be expected due to the fact that in this process one is more likely to observe points occurring in pairs than in triples or singles. Thus we would expect that both $p(x_1,x_2)$ and $p(x_2,x_3)$ are likely to be larger than either $p(x_1,x_2,x_3)$ or $p(x_2)$.

In order to show that a process does not have CPCIC we must first calculate some of the product densities. This is done by making use of Moyal's formula (2.57). In this case, from Example 2.36 of the previous section, the probability generating functional is given by

$$G(\xi) = \exp \int \Big[\Big(\int \xi(x+r) f(r)\,dr\Big)^2 - 1\Big]dx. \tag{3.42}$$

By (2.57) we can use this to calculate the product densities.

$$M_{x_1,\ldots,x_k} = \lim_{\eta\uparrow 1} \left\{ \frac{\partial^k}{\partial x_1 \ldots \partial x_k} \right.$$

$$\exp\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}[\eta+\sum_{i=1}^{k}\lambda_i 1_{[0,x_i]}(x+r)]f(r)\,dr\right)-1\,dx\Big\}_{\lambda_1=\cdots=\lambda_k=0} \tag{3.43}$$

Let $$h_k(\lambda_1,\ldots,\lambda_k)=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}[1+\sum_{i=1}^{k}\lambda_i 1_{[0,x_i]}(x+r)]f(r)\,dr\right)-1\,dx \tag{3.44}$$

and $$\hat{G}_k(\lambda_1,\ldots,\lambda_k)=\exp(h_k(\lambda_1,\ldots,\lambda_k)). \tag{3.45}$$

Then, first taking the limit as η increases to one,

$$M_{x_1,\ldots,x_k} = \frac{\partial^k}{\partial x_1\ldots\partial x_k}\left\{\hat{G}_k(\lambda_1,\ldots,\lambda_k)\right\}_{\lambda_1=\cdots=\lambda_k=0}. \tag{3.46}$$

We use this to calculate $p(x_1)$:

$$M_{x_1}=\frac{\partial}{\partial\lambda_1}\hat{G}_1(\lambda_1)|_{\lambda_1=0}=\exp(h_1(\lambda_1))\frac{\partial h}{\partial\lambda_1}|_{\lambda_1=0} \tag{3.47}$$

but $\frac{\partial h_1}{\partial\lambda_1}=$

$$\int_{-\infty}^{\infty}2\left(\int_{-\infty}^{\infty}[1+\lambda_1 1_{[0,x_1]}(x+r)]f(r)\,dr\right)\left(\int_{-\infty}^{\infty}1_{[0,x_1]}(x+r)\,f(r)\,dr\right)dx. \tag{3.48}$$

Let $$I_i=\int_{-\infty}^{\infty}1_{[0,x_i]}(x+r)\,f(r)\,dr \tag{3.49}$$

so, $$\frac{\partial}{\partial\lambda_1}\hat{G}_1=(\exp(h_1))\cdot\int_{-\infty}^{\infty}2\left(\int_{-\infty}^{\infty}[1+\lambda_1 1_{[0,x_1]}(x+r)]f(r)\,dr\right)\cdot I_1\,dx. \tag{3.50}$$

We evaluate this at $\lambda_1=0$ to get

$$M_{x_1}=\int_{-\infty}^{\infty}2\,I_1\,dx = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}1_{[0,x_1]}(x+r)f(r)\,dr\,dx. \tag{3.51}$$

To calculate $p(x_1)$ we differentiate M_{x_1} with respect to x_1. First we change the order of integration.

$$M_{x_1} = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{[0,x_1]}(x+r)f(r)\,dr\,dx = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{[0,x_1]}(x+r)f(r)\,dx\,dr$$

$$= 2\int_{-\infty}^{\infty}\int_{-r}^{x_1-r} f(r)\,dx\,dr = 2\int_{-\infty}^{\infty} f(r)x_1\,dr = 2x_1 \tag{3.52}$$

We differentiate this to get

$$p(x_1) = 2. \tag{3.53}$$

This is the intensity of the point process. Similarly, we calculate $p(x_1,x_2)$:

$$M_{x_1,x_2} = \frac{\partial^2}{\partial\lambda_1\partial\lambda_2}\hat{G}_2(\lambda_1,\lambda_2)|_{\lambda_1=\lambda_2=0}, \text{ and} \tag{3.54}$$

$$\frac{\partial^2}{\partial\lambda_1\partial\lambda_2}\hat{G}_2(\lambda_1,\lambda_2) = \exp(h_2(\lambda_1,\lambda_2))\frac{\partial^2 h_2}{\partial\lambda_1\partial\lambda_2} + \frac{\partial h_2}{\partial\lambda_1}\cdot\frac{\partial h_2}{\partial\lambda_2}\exp(h_2(\lambda_1,\lambda_2))$$

$$= \exp(h_2(\lambda_1,\lambda_2))\left(\frac{\partial^2 h_2}{\partial\lambda_1\partial\lambda_2} + \frac{\partial h_2}{\partial\lambda_1}\cdot\frac{\partial h_2}{\partial\lambda_2}\right) \tag{3.55}$$

But, $$\frac{\partial^2 h_2}{\partial\lambda_1\partial\lambda_2} = \frac{\partial}{\partial\lambda_2}\left(\int_{-\infty}^{\infty} 2I_1\left(\int_{-\infty}^{\infty}[1+\sum_{i=1}^{2}\lambda_i 1_{[0,x_i]}(x+r)]f(r)\,dr\right)dx\right)$$

$$= \int_{-\infty}^{\infty} 2I_1\left(\int_{-\infty}^{\infty} 1_{[0,x_2]}(x+r)\,f(r)\,dr\right)dx = \int_{-\infty}^{\infty} 2I_1 I_2\,dx. \tag{3.56}$$

Using (3.50) and (3.56) to substitute into (3.55) we find that

$$\frac{\partial^2}{\partial\lambda_1\partial\lambda_2}\hat{G}_2(\lambda_1,\lambda_2) = \exp(h_2(\lambda_1,\lambda_2))$$

$$\cdot\left(\int_{-\infty}^{\infty} 2I_1 I_2\,dx + \left(\int_{-\infty}^{\infty} 2I_1\left(\int_{-\infty}^{\infty}[1+\sum_{i=1}^{2}\lambda_i 1_{[0,x_i]}(x+r)]f(r)\,dr\right)dx\right)\right.$$

$$\left.\cdot\left(\int_{-\infty}^{\infty} 2I_1\left(\int_{-\infty}^{\infty}[1+\sum_{i=1}^{2}\lambda_i 1_{[0,x_i]}(x+r)]f(r)\,dr\right)dx\right)\right) \tag{3.57}$$

Evaluating (3.57) at $\lambda_1=\lambda_2=0$ we obtain

$$M_{x_1,x_2} = \int_{-\infty}^{\infty} 2I_1 I_2 \, dx + \Big(\int_{-\infty}^{\infty} 2I_1 dx \Big)\Big(\int_{-\infty}^{\infty} 2I_2 \, dx \Big)$$

$$= 4 \Big(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[0,x_1]}(x+r) f(r) \, dr \, dx \Big)\Big(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[0,x_2]}(x+r) f(r) \, dr \, dx \Big)$$

$$+ 2 \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{\infty} 1_{[0,x_1]}(x+r) f(r) \, dr \Big)\Big(\int_{-\infty}^{\infty} 1_{[0,x_2]}(x+r) f(r) \, dr \Big) dx. \tag{3.58}$$

Again we will differentiate, this time with respect to both x_1 and x_2, to obtain $p(x_1,x_2)$. First we change the order of integration in the integrals of the first term of (3.58).

$$M_{x_1,x_2} = 4 \Big(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[0,x_1]}(x+r) f(r) \, dx \, dr \Big)\Big(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[0,x_2]}(x+r) f(r) \, dx \, dr \Big)$$

$$+ 2 \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{\infty} 1_{[0,x_1]}(x+r) f(r) \, dr \Big)\Big(\int_{-\infty}^{\infty} 1_{[0,x_2]}(x+r) f(r) \, dr \Big) dx$$

$$= 4 \Big(\int_{-\infty}^{\infty} \int_{-r}^{x_1 - r} f(r) \, dx \, dr \Big)\Big(\int_{-\infty}^{\infty} \int_{-r}^{x_2 - r} f(r) \, dx \, dr \Big)$$

$$+ 2 \int_{-\infty}^{\infty} \Big(\int_{-x}^{x_1 - x} f(r) \, dr \Big)\Big(\int_{-x}^{x_2 - x} f(r) \, dr \Big) dx$$

$$= 4x_1 x_2 + 2 \int_{-\infty}^{\infty} \Big(\int_{-x}^{x_1 - x} f(r) \, dr \Big)\Big(\int_{-x}^{x_2 - x} f(r) \, dr \Big) dx. \tag{3.59}$$

Differentiating (3.59) with respect to x_1 and x_2 we get

$$p(x_1,x_2) = 4 + 2 \int_{-\infty}^{\infty} f(x_1 - x) f(x_2 - x) \, dx \tag{3.60}$$

Lastly we calculate $p(x_1,x_2,x_3)$:

$$M_{x_1,x_2,x_3} = \frac{\partial^3}{\partial\lambda_1 \partial\lambda_2 \partial\lambda_3} \hat{G}_3(\lambda_1,\lambda_2,\lambda_3)\big|_{\lambda_1=\lambda_2=\lambda_3=0} \tag{3.61}$$

$$\frac{\partial^3}{\partial\lambda_1\partial\lambda_2\partial\lambda_3}\hat{G}_3(\lambda_1,\lambda_2,\lambda_3)$$

$$=\exp(h_3)\left[\frac{\partial}{\partial\lambda_3}\left(\frac{\partial^2 h_3}{\partial\lambda_2\partial\lambda_1}\right)+\frac{\partial}{\partial\lambda_3}\left(\frac{\partial h_3}{\partial\lambda_2}\right)\cdot\left(\frac{\partial h_3}{\partial\lambda_1}\right)+\frac{\partial}{\partial\lambda_3}\left(\frac{\partial h_3}{\partial\lambda_1}\right)\cdot\left(\frac{\partial h_3}{\partial\lambda_2}\right)\right]$$

$$+\exp(h_3)\cdot\frac{\partial h_3}{\partial\lambda_3}\cdot\left[\frac{\partial^2 h_3}{\partial\lambda_2\partial\lambda_1}+\frac{\partial h_3}{\partial\lambda_1}\cdot\frac{\partial h_3}{\partial\lambda_2}\right]$$

$$=\exp(h_3)\left[\frac{\partial^3 h_3}{\partial\lambda_3\partial\lambda_2\partial\lambda_1}+\frac{\partial^2 h_3}{\partial\lambda_3\partial\lambda_2}\cdot\frac{\partial h_3}{\partial\lambda_1}+\frac{\partial^2 h_3}{\partial\lambda_3\partial\lambda_1}\cdot\frac{\partial h_3}{\partial\lambda_2}\right.$$

$$\left.+\frac{\partial^2 h_3}{\partial\lambda_2\partial\lambda_1}\cdot\frac{\partial h_3}{\partial\lambda_3}+\frac{\partial h_3}{\partial\lambda_3}\cdot\frac{\partial h_3}{\partial\lambda_2}\cdot\frac{\partial h_3}{\partial\lambda_1}\right]$$

$$=\exp(h_3)\left[\frac{\partial^2 h_3}{\partial\lambda_3\partial\lambda_2}\cdot\frac{\partial h_3}{\partial\lambda_1}+\frac{\partial^2 h_3}{\partial\lambda_3\partial\lambda_1}\cdot\frac{\partial h_3}{\partial\lambda_2}+\frac{\partial^2 h_3}{\partial\lambda_2\partial\lambda_1}\cdot\frac{\partial h_3}{\partial\lambda_3}\right.$$

$$\left.+\frac{\partial h_3}{\partial\lambda_3}\cdot\frac{\partial h_3}{\partial\lambda_2}\cdot\frac{\partial h_3}{\partial\lambda_1}\right],\quad \text{since } \frac{\partial^3 h_3}{\partial\lambda_3\partial\lambda_2\partial\lambda_1}=0. \tag{3.62}$$

Note that from the previous two calculations it follows that $\frac{\partial h_3}{\partial\lambda_i}$ evaluated at $\lambda_1=\lambda_2=\lambda_3=0$ is $2x_i$ for each i and that $\frac{\partial^2 h_3}{\partial\lambda_i\partial\lambda_j}$ evaluated at $\lambda_1=\lambda_2=\lambda_3=0$ is $4x_ix_j+\int_{-\infty}^{\infty} 2I_iI_j\,dx$ for each $i\neq j$. Thus,

$$\frac{\partial^3}{\partial\lambda_1\partial\lambda_2\partial\lambda_3}\hat{G}_3(\lambda_1,\lambda_2,\lambda_3)\Big|_{\lambda_1=\lambda_2=\lambda_3=0}$$

$$=\left(4x_2x_3+\int_{-\infty}^{\infty} 2I_2I_3\,dx\right)(2x_1)+\left(4x_1x_3+\int_{-\infty}^{\infty} 2I_1I_3\,dx\right)(2x_2)$$

$$+\left(4x_1x_2+\int_{-\infty}^{\infty} 2I_1I_2\,dx\right)(2x_3)+(2x_1)(2x_2)(2x_3)$$

$$=32x_1x_2x_3+4x_1\int_{-\infty}^{\infty} I_2I_3\,dx+4x_2\int_{-\infty}^{\infty} I_1I_3\,dx+4x_3\int_{-\infty}^{\infty} I_1I_2\,dx. \tag{3.63}$$

Differentiating with respect to x_1,x_2 and x_3 as before we get

$$p(x_1,x_2,x_3) = 32 + 4\int_{-\infty}^{\infty} f(x_2-x)f(x_3-x)\,dx + 4\int_{-\infty}^{\infty} f(x_1-x)f(x_2-x)\,dx + 4\int_{-\infty}^{\infty} f(x_1-x)f(x_3-x)\,dx. \tag{3.64}$$

Having calculated $p(x_1)$, $p(x_1,x_2)$, $p(x_1,x_2,x_3)$ we are now able to show that X does not have CPCIC by making an appropriate choice of f so that

$$p(x_1,x_2,x_3)\,p(x_2) < p(x_1,x_2)\,p(x_2,x_3). \tag{3.65}$$

The left hand side of (3.65) is (from (3.53) and (3.54))

$$p(x_1,x_2,x_3)\,p(x_2) = 64 + 8\Big(\int_{-\infty}^{\infty} f(x_2-x)f(x_3-x)\,dx + \int_{-\infty}^{\infty} f(x_1-x)f(x_2-x)\,dx + \int_{-\infty}^{\infty} f(x_1-x)f(x_3-x)\,dx\Big) \tag{3.66}$$

while on the right hand side (from (3.60)) we have

$$p(x_1,x_2)\,p(x_2,x_3) = 16 + 8\Big(\int_{-\infty}^{\infty} f(x_2-x)f(x_3-x)\,dx + \int_{-\infty}^{\infty} f(x_1-x)f(x_2-x)\,dx\Big) + 4\Big(\int_{-\infty}^{\infty} f(x_2-x)f(x_3-x)\,dx\Big)\Big(\int_{-\infty}^{\infty} f(x_1-x)f(x_2-x)\,dx\Big). \tag{3.67}$$

Cancelling like terms in (3.66) and (3.67) we are reduced to showing that

$$64 + 8\int_{-\infty}^{\infty} f(x_1-x)f(x_3-x)\,dx < 16 + 4\Big(\int_{-\infty}^{\infty} f(x_2-x)f(x_3-x)\,dx\Big)\Big(\int_{-\infty}^{\infty} f(x_1-x)f(x_2-x)\,dx\Big) \tag{3.68}$$

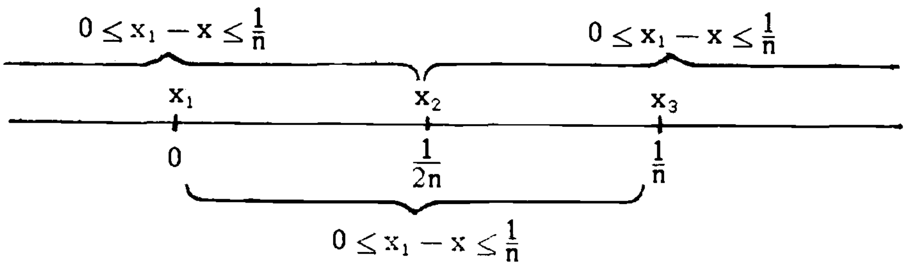
or,

$$12 + 2\int_{-\infty}^{\infty} f(x_1-x)f(x_3-x)\,dx < \Big(\int_{-\infty}^{\infty} f(x_2-x)f(x_3-x)\,dx\Big)\Big(\int_{-\infty}^{\infty} f(x_1-x)f(x_2-x)\,dx\Big). \tag{3.69}$$

We now determine an appropriate choice of f. For a fixed value of n

let
$$f_n(x) = \begin{cases} n & \text{if } 0 \le x \le n \\ 0 & \text{otherwise} \end{cases} \tag{3.70}$$

Let $x_1=0$, $x_2=\frac{1}{2n}$, and $x_3=\frac{1}{n}$. Then $f_n(x_1-x)f_n(x_3-x)=0$ because if $0 \le x_1-x \le \frac{1}{n}$ then $0 \le -x \le \frac{1}{n}$ which implies $-\frac{1}{n} \le x \le 0$ so that $\frac{1}{n}-x > \frac{1}{n}$, thus $f_n(x_3-x)=0$. Similarly, if $f_n(x_3-x) \neq 0$ then $f_n(x_1-x)=0$. (see diagram below)



(3.71)

For $f = f_n$ (3.69) becomes

$$\begin{aligned} 12 \quad &< \Big(\int_{-\infty}^{\infty} f(x_2-x)f(x_3-x)\,dx\Big)\Big(\int_{-\infty}^{\infty} f(x_1-x)f(x_2-x)\,dx\Big) \\ &= \Big(\int_0^{\frac{1}{2n}} n \cdot n\,dx\Big)\Big(\int_{\frac{1}{2n}}^{\frac{1}{n}} n \cdot n\,dx\Big) \\ &= n^2 \frac{1}{2n} n^2 \frac{1}{2n} = \frac{n^2}{4} \end{aligned} \tag{3.72}$$

Which holds as long as $n^2 \ge 48$. So, for example, as long as $n \ge 7$ we get the desired result. That is, we have shown

<u>Theorem</u> <u>3.19</u> If X is a Neyman Scott cluster process on $\mathbb{R}$ where the cluster center process is a stationary Poisson process with intensity $\lambda=1$ and there are exactly two points distributed about each center according to a common distance distribution having density $f_n(x)$ given by (3.70) for any given $n \ge 7$, then X is associated but not CPC.

<u>Proof of Theorem 3.13</u> First, to see the general idea, notice that $p(x_1,x_2,x_3)=q(x_1,x_2,x_3)+q(x_1,x_2)q(x_3)+q(x_1,x_3)q(x_2)+q(x_2,x_3)q(x_1)+q(x_1)q(x_2)q(x_3)$ Whereas, $p(x_1,x_2)p(x_3)=[q(x_1,x_2)+q(x_1)q(x_2)]q(x_3)=q(x_1,x_2)q(x_3)+q(x_1)q(x_2)q(x_3)$ so that $p(x_1,x_2,x_3)\geq p(x_1,x_2)p(x_3)$ since $p(x_1,x_2,x_3)$ contains all of the terms of $p(x_1,x_2)p(x_3)$ as well as additional, non-negative, terms.

In general $p(x_1,\ldots,x_k,x_{k+1},\ldots,x_{k+n})$ consists of the sum of products of $q(\cdot)$ terms where each such product is taken over a subdivision of $(x_1,\ldots,x_{k+n})$ and all such subdivisions are represented once in the sum. (e.g. one such product is $q(x_1,\ldots,x_k)q(x_{k+1},\ldots,x_{k+n})$ another is $q(x_1,x_2)q(x_3,x_5)q(x_4,x_6,\ldots,x_{n+k})$, etc.). On the other hand, $p(x_1,\ldots,x_k)p(x_{k+1},\ldots,x_{k+n})$ is a product of similar sums for $p(x_1,\ldots,x_k)$ and $p(x_{k+1},\ldots,x_{k+n})$. Clearly every term in the resulting sum for $p(x_1,\ldots,x_k)p(x_{k+1},\ldots,x_{k+n})$ is a term in the sum for $p(x_1,\ldots,x_k,x_{k+1},\ldots,x_{k+n})$. There are, however, in $p(x_1,\ldots,x_k,x_{k+1},\ldots,x_{k+n})$ additional terms in which the subdivisions allow for a combination of x_i's with $1\leq i\leq k$ and x_i's with $k+1\leq i\leq k+n$, e.g. $q(x_1,\ldots,x_{k+1},x_{k+2})q(x_{k+3},\ldots,x_{k+n})$ is a term of the sum corresponding to $p(x_1,\ldots,x_k,x_{k+1},\ldots,x_{k+n})$, but is not a term of the sum corresponding to $p(x_1,\ldots,x_k)p(x_{k+1},\ldots,x_{k+n})$. Since all of the cumulant densities are non-negative, adding in these additional terms makes $p(x_1,\ldots,x_k,x_{k+1},\ldots,x_{k+n})$ at least as large as $p(x_1,\ldots,x_k)p(x_{k+1},\ldots,x_{k+n})$. □

<u>Example 3.20</u> This example shows that X can have PCIC and still have cumulant densities that are negative. We consider a process on [0,1] where the product densities again depend only on the number of point occurrences and not on their locations. In this case we take $p(x_1,\ldots,x_n)=g(n)$ where

$$g(n)=\begin{cases} c^{\frac{n}{2}} & \text{if } n \text{ is even} \\ c^{\frac{n-1}{2}} & \text{if } n \text{ is odd} \end{cases} \tag{3.73}$$

There are three things we must check:

(1) The product densities p given above determine a completely regular point process, that is g(n) satisfies the conditions of Lemma 3.15.

(2) The process has positively correlated increasing cylinder sets. That is, $p(x_1,\ldots,x_{k+n}) \geq p(x_1,\ldots,x_k)p(x_{k+1},\ldots,x_{k+n})$ or, $g(n+k) \geq g(n)g(k)$.

(3) The process has at least one negative cumulant density function.

We begin by checking (2) and (3). For (2) we list the possible cases and note that in each case the desired inequality holds.

	g(n+k)	g(n)	g(k)
(a) n even, k even	$c^{\frac{n+k}{2}}$	$c^{\frac{n}{2}}$	$c^{\frac{k}{2}}$
(b) n even, k odd	$c^{\frac{n+k-1}{2}}$	$c^{\frac{n}{2}}$	$c^{\frac{k-1}{2}}$
(c) n odd, k odd	$c^{\frac{n+k}{2}}$	$c^{\frac{n-1}{2}}$	$c^{\frac{k-1}{2}}$

In cases (a) and (b) it is clear that $g(n+k) \geq g(n)g(k)$, no matter what value c takes. In fact in these two cases equality holds. In case (c), $g(n+k) \geq g(n)g(k)$ if and only if $c^{\frac{n+k}{2}} \geq c^{\frac{n-1}{2}} \cdot c^{\frac{n-1}{2}} = c^{\frac{n+k-2}{2}}$. That is, if and only if $1 \geq c^{-\frac{1}{2}}$. Thus, as long as we take $c \geq 1$, the process will have PCIC.

To guarantee that at least one cumulant density is negative we will calculate $q(x_1,x_2,x_3)$.

$$
\begin{aligned}
q(x_1,x_2,x_3) &= p(x_1,x_2,x_3) - q(x_1,x_2)q(x_3) - q(x_1,x_3)q(x_2) \\
&\quad - q(x_2,x_3)q(x_1) - q(x_1)q(x_2)q(x_3) \\
&= p(x_1,x_2,x_3) - [p(x_1,x_2) - p(x_1)p(x_2)]\,p(x_3) \\
&\quad - [p(x_1,x_3) - p(x_1)p(x_3)]\,p(x_2) \\
&\quad - [p(x_2,x_3) - p(x_2)p(x_3)]\,p(x_1) - p(x_1)p(x_2)p(x_3)
\end{aligned}
$$

$$=g(3)-3g(2)g(1)+2(g(1))^3$$

$$=c^1-3c^1c^0+2(c^0)^3=c-3c+2=2-2c \tag{3.74}$$

Thus, $q(x_1,x_2,x_3)<0$ as long as $c>1$.

It remains to determine at least one value of c for which the conditions of Lemma 3.15 hold. That $g(0)=1$ and $g(k)\geq 0$ for $k=0,1,2,\ldots$ are given by the definition. That $\sum_{L=0}^{\infty}\frac{2^L}{L!}g(k+L)<\infty$ for $k=0,1,2,\ldots$ is due to the fact that g is bounded by an exponential function. It remains to show that $\sum_{n=0}^{\infty}(-1)^n\frac{1}{n!}g(k+n)\geq 0$ for $k=0,1,2,\ldots$, that is, $f(k)\geq 0$ for each value of k.

We consider separately the case where k is even and the case where k is odd.

$$f(2n)=\sum_{j=0}^{\infty}(-1)^j\frac{1}{j!}g(j+2n)=\sum_{j=0}^{\infty}(-1)^j\frac{1}{j!}\,c^{\log_c g(j+2n)} \tag{3.75}$$

But
$$\log_c g(j+2n)=\begin{cases}\dfrac{j+2n}{2} & \text{if } j \text{ is even}\\[2mm] \dfrac{j+2n-1}{2} & \text{if } j \text{ is odd}\end{cases}$$

$$=\begin{cases} n+\dfrac{j}{2} & \text{if } j \text{ is even}\\[2mm] n+\dfrac{j-1}{2} & \text{if } j \text{ is odd}\end{cases}$$

$$=n+\log_c g(j) \tag{3.76}$$

Thus,
$$f(2n)=\sum_{j=0}^{\infty}(-1)^j\frac{1}{j!}\,c^{n+\log_c g(j)}$$

$$=c^n\sum_{j=0}^{\infty}(-1)^j\frac{1}{j!}\,c^{\log_c g(j)}=c^n f(0) \tag{3.77}$$

Likewise, $$f(2n+1) = c^n f(1) \tag{3.78}$$

Thus we need only check that $f(0) \geq 0$ and $f(1) \geq 0$.

$$\begin{aligned} f(0) &= \sum_{j=0}^{\infty} (-1)^j \frac{1}{j!} g(j) \\ &= \sum_{j \text{ even}} \frac{1}{j!} c^{\frac{j}{2}} - \sum_{j \text{ odd}} \frac{1}{j!} c^{\frac{j-1}{2}} \end{aligned} \tag{3.79}$$

We will compare the sum for j even with the sum for j odd by comparing them term by term. The n^{th} term of the first sum in (3.79) (starting with $n=0$) is $\frac{1}{2n!} c^n$, whereas the n^{th} term of the second sum is (again starting with $n=0$) is $\frac{1}{(2n+1)!} c^n$. Since $c^n(\frac{1}{2n!} - \frac{1}{(2n+1)!}) \geq 0$ for all values of n and c, $f(0) \geq 0$.

Unfortunately, showing that $f(1) \geq 0$ is a little more complicated and requires some restrictions on c.

$$\begin{aligned} f(1) &= \sum_{j=0}^{\infty} (-1)^j \frac{1}{j!} g(j+1) \\ &= \sum_{j \text{ even}} \frac{1}{j!} c^{\frac{j}{2}} - \sum_{j \text{ odd}} \frac{1}{j!} c^{\frac{j+1}{2}} \end{aligned} \tag{3.80}$$

We again compare the last two sums term by term. Starting with $n=0$, the n^{th} term of the first sum is again $\frac{1}{2n!} c^n$, however now the n^{th} term of the second sum is $\frac{1}{(2n+1)!} c^{n+1}$. f(0) is clearly positive for $c \leq 1$ but this will not give us the desired result for comparing the property of PCIC with that of having non-negative cumulant densities. Thus we must look for values of c that are greater than one and still make the above sum non-negative.

In this case we compare the n^{th} term of the sum over evens with the $n+1^{st}$ term of the sum over odds. Note that this leaves the 0^{th} term of the sum over odds unpaired with a term from the evens. That is, there will be a quantity of -c "left over". Thus we must not only check that the pairings give non-negative quantities, but also that when they are summed they sum at least to c, making the entire sum non-negative.

$$\frac{1}{(2n)!}\, c^n - \frac{1}{(2(n+1)+1)!}\, c^{n+2}$$

$$=\frac{1}{(2n)!}\, c^n - \frac{1}{(2n+3)!}\, c^{n+2}$$

$$=\frac{1}{(2n)!}\, c^n \left(1 - \frac{c^2}{(2n+3)(2n+2)(2n+1)}\right) \tag{3.81}$$

(3.81) is non-negative as long as $1 - \frac{c^2}{(2n+3)(2n+2)(2n+1)} \geq 0$. That is, as long as $c^2 \leq (2n+3)(2n+2)(2n+1)$. The smallest value n can take is 0, so we need $c^2 \leq 6$, i.e. $c \leq \sqrt{6}$.

We have now found that as long as $c \leq \sqrt{6}$ the pairings we have chosen provide non-negative sums. It remains to check that we can also make these non-negative terms add up to a quantity at least as large as c in order to insure that $f(1) \geq 0$. To do this we will choose a particular value of $c > 1$ (to guarantee PCIC and at least one negative cumulant density) and $\leq \sqrt{6}$ as required above.

Let $c = 1.1$. For $n = 0$, from (3.81) we get the term

$$1\left(1 - \frac{1.21}{6}\right) \approx .7983 \tag{3.83}$$

For $n = 1$ we get
$$\frac{1.1}{2}\left(1 - \frac{1.21}{69.88}\right) \approx .5405 \tag{3.84}$$

Summing (3.83) and (3.84) we get $1.3388 > c = 1.1$ so that summing the positive terms does outweigh the $-c$, guaranteeing that $f(1) \geq 0$. We have now shown

Theorem 3.21 If X is a point process with product densities given by $p(x_1,\ldots,x_n) = g(n)$ where

$$g(n) = \begin{cases} c^{\frac{n}{2}} & \text{if } n \text{ is even} \\ c^{\frac{n-1}{2}} & \text{if } n \text{ is odd} \end{cases}$$

with $c > 1$ chosen so that the conditions of Lemma 3.15 hold (such values of c do exist) then X has PCIC but also has at least one negative cumulant density.

This example also gives further evidence that a process which has PCIC need not have CPCIC. We can easily check that the given product densities do not satisfy Lemma 3.16. We have $g(0) = g(1) = 1$ and $g(2) = g(3) = c$. Thus $g(1)g(3) = c$, whereas $(g(2))^2 = c^2$. Since $c > 1$, this shows that X does not have CPCIC.

BIBLIOGRAPHY

[1] Barlow, R.E. and Frank Proschan, (1981). Statistical Theory of Reliability and Life Testing. Holt, Reinhart and Winston, Inc.

[2] Van den Berg, J. and R. M. Burton, (1987). FKG and equivalent conditions for binary random variables. preprint.

[3] Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

[4] Billingsley, P. (1979). Probability and Measure. Wiley, New York.

[5] Burton, R. M. and E. Waymire, (1985). Classical scaling limits for associated random measures. Ann. of Prob. 13 (4), 1267-1278.

[6] Burton, R. M. and E. Waymire, (1986). The central limit problem for infinitely divisible random measures, in Dependence in Probability and Statistics, pg 383-395, ed. M. Taqqu and E. Eberlein. Birkhauser Press, Boston.

[7] Cohn,D.L., (1980). Measure Theory. Birkhauser Press, Boston.

[8] Daley, D. J. and D. Vere-Jones (1972). A Summary of the theory of point processes, in Stochastic Point Processes, pg. 299-383, ed. Lewis. Wiley, New York.

[9] Esary, J. D., F. Proschan and D.W. Walkup, (1967). Association of random variables, with applications. Ann. Mat. Satatist. 38, 1466-1474.

[10] Feller, W. (1971). An Introduction to Probability Theory and Its Applications vol. I. Wiley, New York.

[11] Fisher, L. (1972). A survey of the mathematical theory of multidimensional point processes, in Stochastic Point Processes, pg. 468-513, ed. Lewis. Wiley, New York.

[12] Fortuin, C., P. Kastelyn and J. Ginibre (1971). Correlation inequalities on some partially ordered sets. Comm. Math. Phys. (22), 89-103.

[13] Gupta, and E. Waymire. The mathematical structure of rainfall, part II, a review of the theory of point processes. Water Resources Research 17 (5), 1275 -1285.

[14] Hoeffding, W. (1940). Masstabinvariante korrelations theorie. Shriften Math. Inst. Univ. Berlin (5) 181-233.

[15] Kallenberg, O. (1983). Random Measures, 3rd edition. Akademic Press, New York.

[16] Kallenberg, O. (1975). On symmetrically distributed random measures. Trans Am. Math. Soc. (202), 105-121.

[17] Karr, A. F. , (1986). Point Processes and their Statistical Inference. Marcel Dekker, Inc.

[18] Kemperman, J. H. B. (1977). On the FKG inequality for measures on a partially ordered space. Indag. Math. 39, 313 - 331.

[19] Kinderman, R. and J.L. Snell (1980). Markov Random Fields and their Applications. AMS, Providence.

[20] Lehman, E.L., (1966). Some concepts of dependence. Ann. Math. Statist. (37), 1137-1153.

[21] Macchi, O., (1975). The coincidence approach to point processes. Adv. Appl. Prob. (7), 83-122.

[22] Matthes, K., J. Kerstan, and J. Mecke, (1978). Infinitely Divisible Point Processes. Wiley, New York.

[23] Moyal, J. E., (1962). The general theory of stochastic population processes. Acta. Math. (108), 1-31.

[24] Newman, C.M. (1980). Normal fluctuations and the FKG inequalities. Comm. Math. Phys. (74), 119-128.

[25] Newman, C.M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables, in Inequalities in Statistics and Probability, IMS Lecture Notes - Monograph Series (5), 127-140.

[26] Newman, C.M. and L.S. Schulman (1981). Infinite clusters in percolation models. J. Statist. Phys. (26), 613-628.

[27] Neyman, J. and E. L. Scott, (1958). A statistical approach to problems of cosmology. J. R. Statist. Soc. B 20, 1-43.

[28] Neyman, J. and E. L. Scott, (1972). Processes of clustering, in Stochastic Point Processes, pg. 646-681, ed. Lewis. Wiley, New York.

[29] Perlman, M. D. and I. Olkin (1980). Uniqueness of invariant tests for MANOVA and other multivariate problems. Annals of Stat. 8, 1326 - 1341.

[30] Royden, H.L., (1968). Real Analysis 2nd edition. McMillan, Toronto.

[31] Rudin, W., (1966). Real and Complex Analysis. McGraw-Hill, New York.

[32] Srinivasan, S.K. (1974). Stochastic Point Processes and Their Applications. Hafner Press, New York.

[33] Westcott, M. (1972). The probability generating functional. J. Austral. Math. Soc. (14), 448-466.

APPENDICES

These appendices summarize some basic material on weak convergence of random elements. For more complete details and proofs see Billingsley [4] (in the case of random elements of a metric space) and Kallenberg [15] (in the case of random measures) or Karr [17].

A.1 Weak Convergence (Convergence in Distribution) of Random Elements in a Metric Space

Let S be a metric space and $\mathcal{S}$ be the Borel σ-field on S (i.e. $\mathcal{S}$ is the σ-field generated by the open sets in S). Let P_n, $n=1,2,\ldots$ and P be probability measures on $(S,\mathcal{S})$.

Definition A.1.1 If $\int f\, dP_n \rightarrow \int f\, dP$ for every bounded, continuous, real valued function f on S then we say P_n converges weakly to P and write $P_n \Rightarrow P$.

Let $(\Omega,\mathcal{F},\tilde{P})$ be any probability space.

Definition A.1.2 A random element on S is a measurable mapping $X:\Omega \rightarrow S$. The distribution of X is the probability measure $P = \tilde{P}X^{-1}$ on $(S,\mathcal{S})$. That is, $P(A) = \tilde{P}X^{-1}(A) = \tilde{P}\{w: X(w) \in A\}$.

Weak convergence can also be understood in terms of the convergence of the distributions of random elements.

Definition A.1.3 A sequence $\{X_n\}$ of random elements converges in distribution to the random element X, written $X_n \xrightarrow{d} X$, if the corresponding distributions P_n converge weakly to P.

In many situations it may be difficult to check the requirement given for convergence in distribution. The next theorem states that it is enough to check that the integrals given in definition A.1.1 converge for bounded uniformly continuous functions f. It also gives several other equivalent definitions for convergence in distribution.

Theorem A.1.4 The following are equivalent

a. $X_n \xrightarrow{d} X$.

b. $\lim_{n\to\infty} E\{f(X_n)\} = E\{f(X)\}$ for all bounded, uniformly continuous, real f.

c. $\limsup_{n\to\infty} P\{X_n \in F\} \leq P\{X \in F\}$ for all closed sets F in S.

d. $\liminf_{n\to\infty} \{X_n \in G\} \leq P\{X \in G\}$ for all open sets G in S.

e. $\lim_{n\to\infty} P\{X_n \in A\} = P\{X \in A\}$ for all X continuity sets A (that is all sets A for which $P\{X \in \partial A\} = 0$).

In the special case where $S = \mathbb{R}^d$ for some d, the random element $X:(\Omega,\mathcal{F},\tilde{P}) \to (\mathbb{R}^d, \mathcal{B}^d)$ (where $\mathcal{B}^d$ denotes the Borel σ-field on $\mathbb{R}^d$) is called a random variable (in the case where $d=1$) or a random vector (in the cases where $d>1$). The distributions of random variables and random vectors are usually understood in terms of their distribution functions.

<u>Definition A.1.5</u> The <u>distribution function</u> of a random vector X is the function F(x) given by $F(x) = P\{y: y \leq x\}$ for $x \in \mathbb{R}^d$ where $y \leq x$ means that $y_i \leq x_i$ for $i = 1,2,\ldots d$.

F has the properties that it is (1) nondecreasing in each variable and (2) continuous from above. F is continuous at x if and only if $\{y: y \leq x\}$ is an X- continuity set.

Another function which is often used to understand information about the distribution of X is the characteristic function of X

<u>Definition A.1.6</u> The <u>characteristic function</u> $\Phi_X(t)$ for $t \in \mathbb{R}^d$ corresponding to a random vector X on $\mathbb{R}^d$ is defined by

$$\Phi_X(t) = \int e^{i<t,X>} \tilde{P}(dX) = \int e^{i<t,x>} dP(x)$$

where $<t,x> = \sum_{i=1}^{d} t_i\, x_i$.

In the special case where $S = \mathbb{R}^d$ Theorem A.1.4 becomes

Theorem A.1.7 The following are equivalent

a. $X_n \xrightarrow{d} X$.

b. $\lim_{n\to\infty} E\{f(X_n)\} = E\{f(X)\}$ for all bounded, uniformly continuous, real f.

c. $\limsup_{n\to\infty} P\{X_n \in F\} \le P\{X \in F\}$ for all closed sets F in $\mathbb{R}^d$.

d. $\liminf_{n\to\infty} \{X_n \in G\} \le P\{X \in G\}$ for all open sets G in $\mathbb{R}^d$.

e. $\lim_{n\to\infty} F_n(x) = F(x)$ for all points $x \in \mathbb{R}^d$ at which F is continuous.

f. If Φ_{X_n}, Φ_X are the characteristic function corresponding to X_n, X then $\Phi_{X_n}(t) \to \Phi_X(t)$ for all $t \in \mathbb{R}^d$.

A.2 Weak Convergence and Convergence in Distribution For Random Measures

Let $\mathbb{R}^d$ be d- dimensional Euclidean space and $D \subseteq \mathbb{R}^d$ a fixed, possibly infinite, subrectangle. Let $\mathfrak{B}^d$ be the collection of Borel subsets of D. Denote the subset of $\mathfrak{B}^d$ consisting of bounded sets (i.e. sets with compact closures) by $\hat{\mathfrak{B}}^d$. A measure μ on $(D,\mathfrak{B}^d)$ is called Radon if $\mu(B) < \infty$ for all sets $B \in \hat{\mathfrak{B}}^d$. Let M denote the set of all Radon measures on $(D,\mathfrak{B}^d)$ and N the subset of M consisting of counting measures. Thus, $\mu \in N$ if and only if $\mu(B) \in \mathbb{Z}^+ = \{0,1,2,\cdots\}$ and $\mu(B) < \infty$ for all $B \in \hat{\mathfrak{B}}^d$. N is naturally identified with the set of all finite or infinite configurations of points (including multiplicities) in D without limit points.

Let $\mathcal{M}$ be the σ-algebra on M generated by sets of the form $\{\mu \in M | \mu(A) < k\}$ for all $A \in \hat{\mathfrak{B}}^d$ and $0 \le k < \infty$. Likewise $\mathcal{N}$ is the σ-algebra generated by such sets of measures in N. Note that $\mathcal{N} \subset \mathcal{M}$ and so $\mathcal{N}$ is the restriction of $\mathcal{M}$ to N. $\mathcal{N}$ is the σ-algebra on N which allows us to count the points in bounded regions of D. We will use the notation μf to denote the integral $\int f \, d\mu$.

<u>Definition</u> <u>A.2.1</u> The <u>vague</u> <u>topology</u> on M (or N) is the topology for which $\mu_n f \to \mu f$ for all functions $f \in \mathcal{F}_C$, where $\mathcal{F}_C = \{ f\colon \mathbb{R}^d \to \mathbb{R}^+ = [0,\infty] \mid f$ is continuous and has compact support$\}$

Consequently the vague topology on M (or N) is the topology generated by the class of all finite intersections of subsets of M (resp. N) of the form $\{\mu \in M \mid s < \int f \, d\mu < t\}$ for all $f \in \mathcal{F}_C$ and $s, t \in \mathbb{R}$.

M is metrizable as a complete separable metric space (i.e. M is a Polish space) in the vague topology. Since N is a closed subset of M in the vague topology, N is also a Polish space. $\mathcal{M}$ and $\mathcal{N}$ are the Borel sets generated by these topologies.

<u>Definition</u> <u>A.2.2</u> A <u>random</u> <u>measure</u> is a measurable mapping X from a probability space $(\Omega,\mathcal{F},P)$ into $(M,\mathcal{M})$.

<u>Definition</u> <u>A.2.3</u> The <u>characteristic</u> <u>functional</u> Φ_X corresponding to a random measure X is defined by

$$\Phi_X(\phi) = E\left\{\exp\left(i \int_{\mathbb{R}^d} \phi(x) \, dX(x)\right)\right\}$$

for real valued functions ϕ which are bounded, measurable and have compact support.

Definition A.2.4 the Laplace functional $\mathcal{L}_X$ corresponding to X is defined by $\mathcal{L}_X(f) = E\{e^{-Xf}\}$ for all measurable functions $f:\mathbb{R}^d \to [0,\infty)$.

Definition A.2.5 X_n converges in distribution to X (written $X_n \xrightarrow{d} X$) if $\lim_{n\to\infty} E[\xi(X_n)] = E[\xi(X)]$ for all bounded continuous functions $\xi:\mathcal{M} \to \mathbb{R}$.

Note that $\xi(X_n)$ and $\xi(X)$ are functions from Ω to $\mathbb{R}$, that is they are random variables. Convergence in distribution in this setting is thus exactly convergence in distribution as described for random elements in definition A.1.3. In this case the random elements are the elements X_n and X of the space $\mathcal{M}$ equipped with a metric making it a complete separable metric space. This definition depends on considering the class of all bounded continuous functions from $\mathcal{M}$ to $\mathbb{R}$, which is certainly awkward in practice. The following theorem gives equivalent definitions of convergence in distribution, including one in which the class of functions to be considered is a more natural class of functions to deal with, continuous functions with compact support from $\mathbb{R}^d$ to $\mathbb{R}$.

Theorem A.2.6 The following are equivalent

a. $X_n \xrightarrow{d} X$.

b. For each function $f:\mathbb{R}^d \to \mathbb{R}$, continuous with compact support, $X_nf \to Xf$ in distribution as random variables.

c. The characteristic functionals corresponding to X_n converge to that corresponding to X.

d. The Laplace functionals corresponding to X_n converge to that corresponding to X.

e. For each $B_1,\ldots,B_k \in \hat{\mathfrak{B}}^d$ with $\mu(\partial B_i)=0$ a.s. for each i, $(\mu_n(B_1),\ldots,\mu_n(B_k)) \to (\mu(B_1),\ldots,\mu(B_k))$.

f. For each $B_1,\ldots,B_k \in \hat{\mathfrak{B}}^d$ with $\mu(\partial B_i)=0$ a.s. for each i, and for every choice of a_i, $\sum a_i\,\mu_n(B_i) \to \sum a_i\,\mu(B_i)$.

If X and Y have the same distribution we will write $X \stackrel{d}{=} Y$. The following theorem is a special case of Theorem A.2.6 since if $X_n \stackrel{d}{=} Y$ for all n then $X_n \stackrel{d}{\to} Y$ if and only if $X \stackrel{d}{=} Y$.

<u>Theorem</u> <u>A.2.7</u> The following are equivalent

a. $X \stackrel{d}{=} Y$.

b. $Xf \stackrel{d}{=} Yf$ for each function $f:\mathbb{R}^d \to \mathbb{R}$, continuous with compact support.

c. $\mathcal{L}_X(f) = \mathcal{L}_Y(f)$ for each function $f:\mathbb{R}^d \to [0,\infty]$, continuous with compact support.

d. For each $B_1,\ldots,B_k \in \hat{\mathfrak{B}}^d$, $(X(B_1),\ldots,X(B_k)) \stackrel{d}{=} (Y(B_1),\ldots,Y(B_k))$.

We will now consider the special class of random measures called point processes.

<u>Definition</u> <u>A.2.8</u> A <u>point</u> <u>process</u> is a measurable mapping X from a probability space $(\Omega,\mathcal{F},P)$ into $(N,\mathcal{N})$. The <u>distribution</u> of X is the induced measure on $(N,\mathcal{N})$ given by $P_X = PX^{-1}$.

Since $(N,\mathcal{N})$ is a restriction of $(M,\mathcal{M})$ a point process may be considered a random measure. Thus if $A \in \mathfrak{B}^d$ we set $X(A) =$ the (random) number of occurrences in A. If f is a function with compact support we set $Xf = \int f\, dX$.

Definition A.22.9 Define the translation operator $T_x: N \to N$ for $x \in \mathbb{R}^d$, $w = \{\delta_{x_i}\} \in N$ by $T_x w = \{\delta_{x_i + x}\}$. For $A \in \mathcal{N}$ let $T_x(A)$ denote the set $\{T_x w \mid w \in A\}$. X (or its distribution P_X) is called stationary if for very $x \in \mathbb{R}^d$, T_x is P_X invariant. That is, $P_X(T_x(A)) = P_X(A)$ for every $x \in \mathbb{R}^d$ and $A \in \mathcal{N}$.

Alternative representations for stationarity, representations in which we need only consider the random measure X itself instead of dealing with its distribution, are given by the next theorem, which follows immediately from Theorem A.2.7.

Theorem A.2.10 The following are equivalent

a. X is stationary.

b. For any Borel subsets $B_1, \ldots, B_n$ of $\mathbb{R}^d$ $\text{dist}(X(B_1), \ldots, X(B_n)) = \text{dist}(X(B_1 + x), \ldots, X(B_n + x))$.

c. $\text{dist } X(T_x f) = \text{dist } (Xf)$ where $T_x f(y) = f(T_x y)$.

Definition A.2.11 the probability generating functional corresponding to X is defined by

$$G(\xi) = E\left\{\exp\left[\int_{\mathbb{R}^d} \log \xi(x)\, dX(x)\right]\right\},$$

for all real valued, measurable functions ξ on $\mathbb{R}^d$ satisfying (i) $0 \le \xi(x) \le 1$ for all $x \in \mathbb{R}^d$ and (ii) $\xi(x) = 1$ on the complement of a bounded subset of $\mathbb{R}^d$.

In the case of point processes, theorem A.2.7 becomes the following.

Theorem A.2.12 The following are equivalent

a. $X_n \xrightarrow{d} X$.

b. For each function $f:\mathbb{R}^d \to \mathbb{R}$, continuous with compact support, $X_n f \to Xf$.

c. The characteristic functionals corresponding to X_n converge to that corresponding to X.

d. The probability generating functionals corresponding to X_n converge to that corresponding to X.

e. For each $B_1,\ldots,B_k \in \hat{\mathfrak{B}}^d$ with $X(\partial B_i)=0$ a.s. for each i, $(X_n(B_1),\ldots,X_n(B_k)) \to (X(B_1),\ldots,X(B_k))$.

f. For each $B_1,\ldots,B_k \in \hat{\mathfrak{B}}^d$ with $X(\partial B_i)=0$ a.s. for each i, and for every choice of a_i, $\sum a_i X_n(B_i) \to \sum a_i X(B_i)$.

Note that if X is a stationary point process $E[X(B)]=\lambda|B|$, where $|B|$ denotes the Lebesgue measure of B. Thus $X(B)=0$ with probability 1 when $|B|=0$. Since bounded rectangular boxes generate the Borel sets on $\mathbb{R}^d$ and $|\partial B|=0$ for rectangles B it is enough in e and f of the above theorem to consider $B_1,\ldots,B_k$ bounded rectangles. This gives the following corollary.

<u>Corollary</u> <u>A.2.13</u> If X is a stationary point process then $X_n \xrightarrow{d} X$ if and only if for each collection of bounded rectangular boxes $B_1,\ldots,B_k$ in $\mathbb{R}^d$ $(X_n(B_1),\ldots,X_n(B_k)) \to (X(B_1),\ldots,X(B_k))$.

The above corollary remains true even if λ is infinite since $\mathbb{R}^d$ can be written as an uncountable union of translations of any affine $d-1$ dimensional hyperplane. If any translate of such a hyperplane were allowed to have positive measure, then all of its translates would have that same measure by stationarity, thus contradicting the fact that we must have a

Radon measure. Consequently, boundaries of d-dimensional rectangles must still have measure zero, even if λ is infinite.