4

Point processes II – General theory

4.1 Basic properties

4.1.1 Introduction

Point processes provide models for irregular patterns of points. The theory was developed in response to various problems of physics, biology, and queueing theory. Important pioneers include Neyman (1939) (as reported by Neyman and Scott, 1972), Palm (1943), Bartlett (1954), Cox (1955), Khintchin (1955) and Matérn (1960); see Guttorp and Thorarinsdottir (2012) for more details of the history of various point process models. The work by Kallenberg, Kerstan, Krickeberg, Matthes, Mecke and Ryll-Nardzewski during 1960 to 1975 laid the foundation of the modern theory of point processes. Books on the subject include: Srinivasan (1974), Kallenberg (1976b, 1983a, 1986, 2002), Brillinger (1978), Matthes *et al.* (1978), Cox and Isham (1980), Brémaud (1981), Karr (1986, 1991), Daley and Vere-Jones (1988, 2003, 2008), Snyder and Miller (1991), Reiss (1993), Last and Brandt (1995), Sigman (1995), van Lieshout (2000), Baccelli and Brémaud (2003), Srinivasanm and Vijayakumar (2003a,b), Møller and Waagepetersen (2004), Jacobsen (2006), Last (2010) and Gelfand *et al.* (2010). This chapter presents the general theory, which is a development of the special case of the Poisson process that is discussed in Chapter 2. Chapter 5 presents some special point process models.

The mathematical definition of a point process on $\mathbb{R}^d$ is as a random variable taking values in a measurable space $[\mathbb{N}, \mathcal{N}]$, where $\mathbb{N}$ is the family of all sequences φ of points of $\mathbb{R}^d$ satisfying two regularity conditions:

(i) the sequence φ is *locally finite*, that is, each bounded subset of $\mathbb{R}^d$ must contain only a finite number of points of φ,

(ii) the sequence is *simple*, that is, $x_i \neq x_j$ if $i \neq j$.

Stochastic Geometry and its Applications, Third Edition.
Sung Nok Chiu, Dietrich Stoyan, Wilfrid S. Kendall and Joseph Mecke.

The condition (ii) of simplicity is relaxed in some extensions of the theory, but not in this book.

For convenience sometimes the notation $\varphi = \{x_n\}$ is used in order to emphasise the sequential nature of φ. The reader should note that the x_n are dummy variables, and have no particular interpretation. Thus for example x_1 need not be the point closest to the origin.

The σ-algebra $\mathcal{N}$ is defined as the smallest σ-algebra on $\mathbb{N}$ to make all mappings $\varphi \mapsto \varphi(B)$ measurable, for B running through the bounded Borel sets. Here $\varphi(B)$ is simply the number of points in the set B. This $\mathcal{N}$ contains the so-called *configuration sets*, often denoted as Y and Z. An example is $Y = \{\varphi : \varphi(B) = 0\}$, the set of all point sequences that have no point in a given set B.

Thus formally a point process Φ is a measurable mapping of a probability space $[\Omega, \mathcal{A}, \mathbf{P}]$ into $[\mathbb{N}, \mathcal{N}]$. More intuitively it is a random choice of one of the φ in $\mathbb{N}$. It generates a distribution on $[\mathbb{N}, \mathcal{N}]$, the *distribution* P of Φ.

Perhaps the following imaginary scenario may help readers who lack training in measure theory. A simple example of a random variable describes the outcome of a roll of a die. Here the random variable takes on the values 1 to 6, and when the die is fair the probability for each of the six values is $\frac{1}{6}$. Now assume that it is possible to construct a die with infinitely many sides, with a (two-dimensional) point pattern on each of the sides. Every time the die is thrown, a point pattern is generated. This die represents the point process Φ, and a point pattern $\varphi = \Phi(\omega)$ is assigned to each sample point ω represented by a roll of die. Now assume somebody is observing the independent repetitive rolling of the die. This person sees different point patterns and, in particular, fluctuating values of the number of points in some fixed, deterministic subset B of the plane. In other words, every throw of the 'point pattern die' produces a different realisation of the random variable 'number of points in B'.

This way of thinking of point processes (and later of random sets and other structures) is fundamental to the way in which mathematicians approach point process theory. The way in which application-oriented people consider point patterns is sometimes different. They often focus on only one irregular point pattern and study its variability. The mathematical theory can be used in such applied studies using the idea of ergodicity, see p. 114.

The word 'process' in the term 'point process' does not imply a dynamic evolution over time. Historically, the first researchers into the theory had in mind random sequences of temporal events, such as instants of arrivals of calls at telephone exchanges, arrivals of customers at a queue, occurrences of earthquakes or mine disasters, etc. However, a notion of time is often absent in applications of point processes in $\mathbb{R}^2$ or $\mathbb{R}^3$. The phrase 'random point field' would be a more exact term:

$$\boxed{\text{point process } = \text{ random point field}}.$$

It is used in Stoyan and Stoyan (1994) and Ohser and Schladitz (2009). Nevertheless, the abstract-thinking mathematicians retained the term 'point process'. *Spatio-temporal point processes* explicitly involving temporal as well as spatial dispersion of points constitute a separate theory; see Diggle (2007), Daley and Vere-Jones (2008), Illian *et al.* (2008) and Cressie and Wikle (2011).

A similar point can be made about the adjective 'stationary', as in 'stationary point processes', occurring later in the chapter. In this context 'stationary' and 'homogeneous' are equivalent terms and sometimes an adverb such as 'statistically' (as in Torquato, 2002),

'spatially' or 'macroscopically' (as in Ohser and Schladitz, 2009) is added to the latter to distinguish it from other kinds of homogeneity:

$$\boxed{\text{stationary} \;=\; \text{homogeneous} \;=\; \text{statistically / spatially / macroscopically homogeneous}}.$$

4.1.2 The distribution of a point process

The distribution P of a point process Φ is determined by the probabilities

$$\begin{aligned} P(Y) &= \mathbf{P}(\Phi \in Y) \\ &= \mathbf{P}(\{\omega \in \Omega : \Phi(\omega) \in Y\}) \qquad \text{for } Y \in \mathcal{N}. \end{aligned}$$

The term $\Phi \in Y$ means that Φ has some property, for example that it has no point in the set B. Then $\mathbf{P}(\Phi \in Y)$ denotes the probability that Φ has this property.

The *finite-dimensional distributions* are of particular importance. These are probabilities of the form

$$\mathbf{P}\big(\Phi(B_1) = n_1, \ldots, \Phi(B_k) = n_k\big)$$

where $B_1, \ldots, B_k$ are bounded Borel sets and $n_1, \ldots, n_k \geq 0$. The term denotes the probability that Φ has n_1 points in the set B_1, ..., and n_k points in B_k. The distribution of Φ on $[\mathbb{N}, \mathcal{N}]$ is uniquely determined by the system of all these values for all $k = 1, 2, \ldots$. In fact, it is determined by the subsystem for which the constituent B_i are pairwise disjoint.

An even smaller but still sufficient subsystem is that of the *void-probabilities*

$$\begin{aligned} v_B &= P(\{\varphi \in \mathbb{N} : \varphi(B) = 0\}) \\ &= \mathbf{P}\big(\Phi(B) = 0\big) \\ &= \mathbf{P}(\Phi \cap B = \emptyset) \qquad \text{for Borel sets } B. \end{aligned}$$

The quantity v_B is the probability that B is empty (i.e. does not contain a point of Φ). If the point process is simple, as described above, then P is determined by the system of values of v_K as K ranges through the compact sets. This last and strongest characterisation follows by interpreting Φ as a random closed set with capacity functional $T_\Phi(K) = 1 - v_K$, and then applying the Choquet theorem in Chapter 6. (A simple direct proof is possible; see for example Møller and Waagepetersen, 2004, Theorem B.1.) The system of void-probabilities (or the 'avoidance function') specifies the probabilities that points do *not* occur in a given region.

An example of the use of void-probabilities is the proof of the Slivnyak–Mecke theorem (4.69) on p. 132.

4.1.3 Notation

Point processes can be considered either as random sets of discrete points or as random measures counting the numbers of points lying in spatial regions. Corresponding to these two interpretations is the following notation:

$x \in \Phi$	asserts that the point x belongs to the random sequence Φ;
$\Phi(B) = n$	asserts that the set B contains n points of Φ.

Since Φ is a random closed set, results from the theory of random closed sets can also be applied to point processes.

Correspondingly there is a variety of notation for point process formulae. The process Φ can be written as

$$\Phi = \{x_1, x_2, \ldots\} = \{x_n\}$$

which emphasises the interpretation of Φ as a sequence of points or a random closed set. Let f be a measurable function on $\mathbb{R}^d$. The sum of $f(x)$ over x in Φ can be written variously as

$$f(x_1) + f(x_2) + \cdots, \qquad \sum_{x\in\Phi} f(x),$$

or (as physicists would write)

$$\int_{\mathbb{R}^d} f(x)p(x)\,\mathrm{d}x \quad \text{where } p(x) = \sum_{y\in\Phi} \delta(x-y) \text{ and } \delta \text{ is the Dirac delta function,}$$

or

$$\int f(x)\Phi(\mathrm{d}x).$$

The mean value of the sum is written variously as

$$\mathbf{E}\left(\sum_{x\in\Phi} f(x)\right), \qquad \int_{\mathbb{N}} \sum_{x\in\varphi} f(x)P(\mathrm{d}\varphi) \qquad \text{or} \qquad \int_{\mathbb{N}}\int_{\mathbb{R}^d} f(x)\varphi(\mathrm{d}x)P(\mathrm{d}\varphi).$$

The number of points of φ in a set B can be written as

$$\varphi(B) = \sum_{x\in\varphi} \mathbf{1}_B(x),$$

where $\mathbf{1}_B(x)$ is the indicator function of B. Its mean value is written as

$$\begin{aligned}\mathbf{E}\big(\Phi(B)\big) &= \int \varphi(B)P(\mathrm{d}\varphi)\\ &= \mathbf{E}\left(\sum_{x\in\Phi} \mathbf{1}_B(x)\right)\\ &= \int \sum_{x\in\varphi} \mathbf{1}_B(x)P(\mathrm{d}\varphi)\\ &= \int\int \mathbf{1}_B(x)\varphi(\mathrm{d}x)P(\mathrm{d}\varphi).\end{aligned}$$

Such a plethora of notation is perhaps confusing for the beginner but reflects the variety of approaches to the theory.

4.1.4 Stationarity and isotropy

This book mainly considers point processes with infinitely many points; physicists would perhaps speak about the 'thermodynamic limit'. But no limit procedures are used in introducing the fundamental notions of stationarity and isotropy; limits will only appear later, in the context of ergodicity.

A point process Φ, or its distribution P, is said to be *stationary* if its characteristics are invariant under translation: the processes $\Phi = \{x_n\}$ and $\Phi_x = \{x_n + x\}$ have the same distribution for all x in $\mathbb{R}^d$. So

$$\mathbf{P}(\Phi \in Y) = \mathbf{P}(\Phi_x \in Y) \tag{4.1}$$

for all configuration sets Y and all x in $\mathbb{R}^d$. Equivalent to (4.1) is

$$P(Y) = P(Y_x), \tag{4.2}$$

where $Y_x = \{\varphi_x : \varphi \in Y\}$.

The notion of *isotropy* is entirely analogous: Φ is isotropic if its characteristics are invariant under rotation; that is to say, Φ and $\boldsymbol{r}\Phi$ have the same distribution for every rotation $\boldsymbol{r}$ around the origin. So

$$\mathbf{P}(\Phi \in Y) = \mathbf{P}(\boldsymbol{r}\Phi \in Y) \tag{4.3}$$

and

$$P(Y) = P(\boldsymbol{r}Y) \tag{4.4}$$

for all $\boldsymbol{r}$ and all Y, with $\boldsymbol{r}Y = \{\boldsymbol{r}\varphi : \varphi \in Y\}$.

Stationarity and isotropy together yield *motion-invariance*. A motion-invariant point process Φ has the same distribution as $\boldsymbol{m}\Phi$ for all rigid motions $\boldsymbol{m}$ of $\mathbb{R}^d$.

Figures 2.3, 2.4 and 4.1 show point patterns which behave like typical realisations of nonstationary point processes, while Figures 2.1, 5.4, 5.6, 5.7 and 5.9 resemble point patterns of stationary processes. Figure 4.1 illustrates that nonstationarity can arise even if the point density is in some sense constant in space. Here the kind of distribution of the points depends on spatial location: lower in the figure there is a stronger degree of clustering.

4.1.5 Intensity measure and intensity

The *intensity measure* Λ of Φ is a characteristic analogous to the mean of a real-valued random variable. Its definition is

$$\Lambda(B) = \mathbf{E}\big(\Phi(B)\big) = \int \varphi(B) P(\mathrm{d}\varphi) \qquad \text{for Borel sets } B. \tag{4.5}$$

So $\Lambda(B)$ is the mean number of points in B.

In particular, if Φ is stationary, then the intensity measure simplifies; it must be translation-invariant since

$$\Lambda(B) = \mathbf{E}\big(\Phi(B)\big) = \mathbf{E}\big(\Phi_x(B)\big) = \mathbf{E}\big(\Phi(B_{-x})\big) = \Lambda(B_{-x}) \tag{4.6}$$

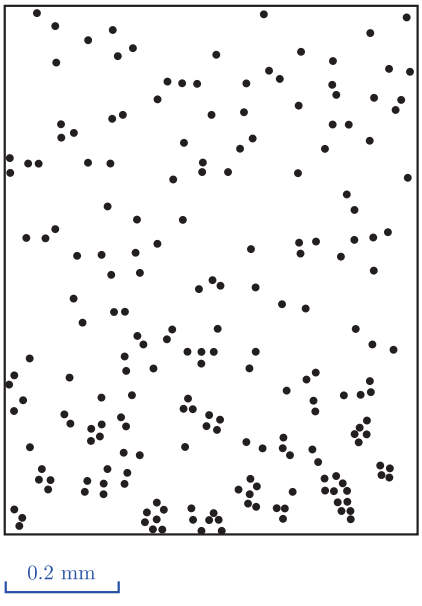

Figure 4.1 A point pattern which should by no means be modelled by a stationary point process. There is obviously a gradient in the vertical direction. The dots are the midpoints of cell nuclei in a specimen of joint cartilage. The lower part of the figure is the side near the bone tissue, while the upper part near the joint.

for all x. By (1.83) this translation-invariance implies that

$$\Lambda(B) = \lambda \nu_d(B) \tag{4.7}$$

for some (possibly infinite) nonnegative constant λ, which is called the *intensity* of Φ. Choosing B to have volume 1 shows that λ may be interpreted as the mean number of points of Φ per unit volume or the *point density*. It is assumed that $0 < \lambda < \infty$.

In general, under some continuity conditions, which are usually satisfied in practical applications of point process theory, $\Lambda(x)$ has a density function $\lambda(x)$, which is called the *intensity function* of Φ:

$$\Lambda(B) = \int_B \lambda(x)\,\mathrm{d}x. \tag{4.8}$$

The continuity condition is violated if, for example, the points are arranged on a lattice. It is clear that $\lambda(x)$ is proportional to the local point density around a location x. If $\mathrm{d}x$ is the volume of an infinitesimal ball centred at x, then $\lambda(x)\mathrm{d}x$ is the probability that there is a point in this ball. This shows that the intensity function can be also interpreted in a local sense; see Daley and Vere-Jones (2008, Section 9.3).

Some calculations may be simplified by using the *Campbell theorem*:

Theorem 4.1. *For any nonnegative measurable function* $f(x)$,

$$\mathbf{E}\left(\sum_{x\in\Phi} f(x)\right) = \int \sum_{x\in\varphi} f(x)P(\mathrm{d}\varphi) = \int\int f(x)\varphi(\mathrm{d}x)P(\mathrm{d}\varphi)$$
$$= \int f(x)\Lambda(\mathrm{d}x). \tag{4.9}$$

This is essentially an application of Fubini's theorem. (The name Campbell theorem refers to early work by Campbell, 1909.) In the stationary case the Λ-integral in (4.9) becomes a volume integral so

$$\mathbf{E}\left(\sum_{x\in\Phi} f(x)\right) = \lambda \int f(x)\,\mathrm{d}x. \tag{4.10}$$

Similarly, when the point process possesses an intensity function then (4.9) can be re-written as:

$$\mathbf{E}\left(\sum_{x\in\Phi} f(x)\right) = \int f(x)\lambda(x)\,\mathrm{d}x. \tag{4.11}$$

4.1.6 Ergodicity and central limit theorem

Often one is interested in *spatial averages*

$$\lim_{n\to\infty} \frac{1}{\nu_d(W_n)} \int_{W_n} f(\Phi_x)\,\mathrm{d}x$$

or in *spatial point averages*

$$\lim_{n\to\infty} \frac{1}{\nu_d(W_n)} \sum_{x_i\in W_n} f(\Phi_{-x_i})$$

for a nonnegative integrable (with respect to **P**) function f and $W_n \uparrow \mathbb{R}^d$. Under ergodicity, these limits exist as deterministic numbers. The following gives some facts of ergodic theory, for details the reader is referred to Kallenberg (2002, Chapter 10), Daley and Vere-Jones (2008, Section 12.2) and Baccelli and Błaszczyszyn (2009a, Section 1.6).

The sequence $\{W_n\}$ is a *convex averaging sequence*, that is, a sequence of subsets of $\mathbb{R}^d$ satisfying:

(1) each W_n is convex and compact;

(2) $W_n \subset W_{n+1}$;

(3) $\sup\{r \geq 0 : B(x, r) \subset W_n \text{ for some } x\} \to \infty$ as $n \to \infty$.

An example is $W_n = B(o, n)$.

A stationary point process Φ is said to be *ergodic* if

$$\lim_{t\to\infty} \frac{1}{(2t)^d} \int_{[-t,t]^d} \mathbf{1}(\Phi_x \in Y, \Phi \in Z)\,\mathrm{d}x = \mathbf{P}(\Phi \in Y)\,\mathbf{P}(\Phi \in Z) \tag{4.12}$$

for all configuration sets Y and Z.

A property that implies ergodicity is *mixing*; Φ is said to be mixing if

$$P(Y \cap Z_x) \to P(Y)P(Z) \qquad \text{for } \|x\| \to \infty \tag{4.13}$$

for all configuration sets Y and Z. Various mixing conditions are given in Daley and Vere-Jones (2008, Section 12.2).

The homogeneous Poisson process is mixing, whereas the mixed Poisson process (see p. 166) is not ergodic.

For a stationary point process which is ergodic, it holds

$$\lim_{n\to\infty} \frac{1}{\nu_d(W_n)} \int_{W_n} f(\Phi_x)\,\mathrm{d}x = \mathbf{E}\big(f(\Phi)\big), \tag{4.14}$$

where the convergence is almost surely. A consequence is

$$\lim_{n\to\infty} \frac{\Phi(W_n)}{\nu_d(W_n)} = \lambda. \tag{4.15}$$

Also *central limit theorems*, that is, theorems ensuring the convergence of the distribution of point numbers in large sets towards normal distributions, can be proved; see Schreiber (2010). Formally, for $\Phi(W_n)$ with a convex averaging sequence $\{W_n\}$

$$\frac{\Phi(W_n) - \lambda\nu_d(W_n)}{\sqrt{\nu_d(W_n)}} \to N(0,\sigma^2) \qquad \text{for } n \to \infty, \tag{4.16}$$

where

$$\sigma^2 = \lim_{n\to\infty} \frac{\mathbf{var}\big(\Phi(W_n)\big)}{\nu_d(W_n)},$$

and the stationary point process Φ has to satisfy strong mixing conditions; see Ivanoff (1982) and Heinrich and Schmidt (1985).

4.1.7 Contact distributions

Let B be a convex compact (or, more generally, star-shaped compact) set in $\mathbb{R}^d$ with $o \in B$ and $\nu_d(B) > 0$. The *contact distribution function* $H_B(r)$ of the stationary point process Φ with respect to the structuring element B is defined by

$$H_B(r) = 1 - \mathbf{P}\big(\Phi(rB) = 0\big) \qquad \text{for } r \geq 0. \tag{4.17}$$

It is a distribution function. In the special case of $rB = B(o,r)$ the contact distribution function is written as $H_s(r)$ and called the *spherical contact distribution function*, *empty space distribution function* or *first contact distribution function*. It can be interpreted as the distribution function of the distance from o to the nearest point of Φ. In the language of statisticians it is the distribution of the distance from an arbitrary test point to its nearest neighbour in Φ.

Another interpretation sees $H_s(r)$ as the volume fraction of the random set $\Phi_{\oplus r}$. This aspect is considered in more detail in Sections 3.2.3 and 6.3.3.

For $H_B(r)$ there always exists a probability density function; see Hansen *et al.* (1999). An excellent survey on contact distribution functions is Hug *et al.* (2002b), which contains formulae for many point process models and considers also the nonstationary case.

Closely related are Klaus Mecke's morphological functions, which describe the behaviour of other geometrical characteristics of the random set $\Phi_{\oplus r}$ as a function of r; see pp. 84 and 229.

Note that the notion of a linear contact distribution function as in Section 6.3.3 is ineffective for point processes since for 'sufficiently random' point processes a line starting in a process point will never hit another process point, and $\Phi \oplus s(o, r)$ has null Lebesgue measure, where $s(o, r)$ denotes a segment of length r of a given direction. (It is $v_d\big(s(o, r)\big) = 0$.)

4.2 Marked point processes

4.2.1 Fundamentals

A point process is made into a *marked point process* by attaching a characteristic (a *mark*) to each point of the process. Thus a marked point process on $\mathbb{R}^d$ is a random sequence $\Psi = \{[x_n; m_n]\}$ from which the points x_n together constitute a point process (unmarked, called the *ground process*) in $\mathbb{R}^d$ and the m_n are the marks corresponding to the respective points x_n. The marks m_n belong to a given *space of marks* $\mathbb{M}$ which is assumed to be a Polish space. The Borel σ-algebra of $\mathbb{M}$ is denoted by $\mathcal{M}$. Important special cases are discrete marks, where $\mathbb{M}$ is a finite set, or real-valued marks, where $\mathbb{M}$ is $\mathbb{R}$.

Specific examples are:

- for x the centre of a particle, m the volume of the particle;
- for x the position of a tree, m the stem diameter of the tree or its growth function;
- for x the centre of an atom, m the type of the atom;
- for x the location (suitably defined) of a convex compact set, m the centred (shifted to origin) set itself.

The marks can be continuous variables, as in the first two examples, indicators of types as in the third example (in which case the terms 'multivariate point process' or 'multitype point process' are often used; 'bivariate' point processes have only two marks), or rather complicated indeed, as in the last example, which occurs in the marked point process interpretation of a germ–grain model (see Section 6.5). Figures 4.2 and 4.3(a) on p. 117 show two point patterns that can be interpreted as samples of marked point processes.

A formally important case, which often serves as null-model, is that of independent marks, where the m_n form a sequence of i.i.d. random variables.

Matthes (1963) and his co-workers established the theory of marked point processes. Its initial application was in queueing theory; such applications up to early 1980s are discussed in the books of Franken *et al.* (1981) and Baccelli and Brémaud (2003). Today marked point processes are a valuable tool in many branches of applied probability.

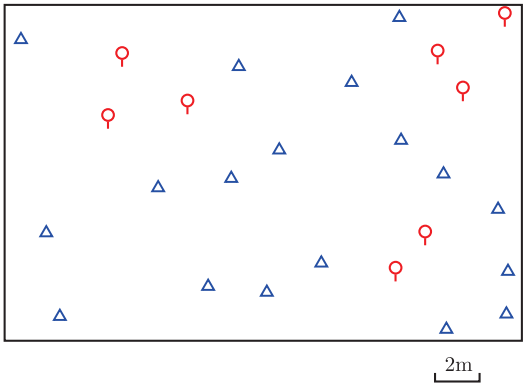

Figure 4.2 Positions of spruces and birches in a forest. The point pattern can be interpreted as a sample of a bivariate point process; that is to say, a marked point process with marks 0 and 1.

It is of course possible to interpret the marked point process Ψ as an ordinary point process in $\mathbb{R}^d \times \mathbb{M}$. The notation of a random counting measure can be used: for Borel $B \in \mathcal{B}^d$ and $L \in \mathcal{M}$ the number of points of Ψ in B with marks in L is denoted by $\Psi(B \times L)$. However, there is a particular feature of marked point processes which makes it worthwhile to consider them separately: usually one defines rigid motions of marked point processes as transforms which move the points but leave the marks unchanged. So Ψ_x, the translate of $\Psi = \{[x_n, m_n]\}$ by x, is given by

$$\Psi_x = \{[x_1 + x; m_1],\ [x_2 + x; m_2],\ \ldots\}. \tag{4.18}$$

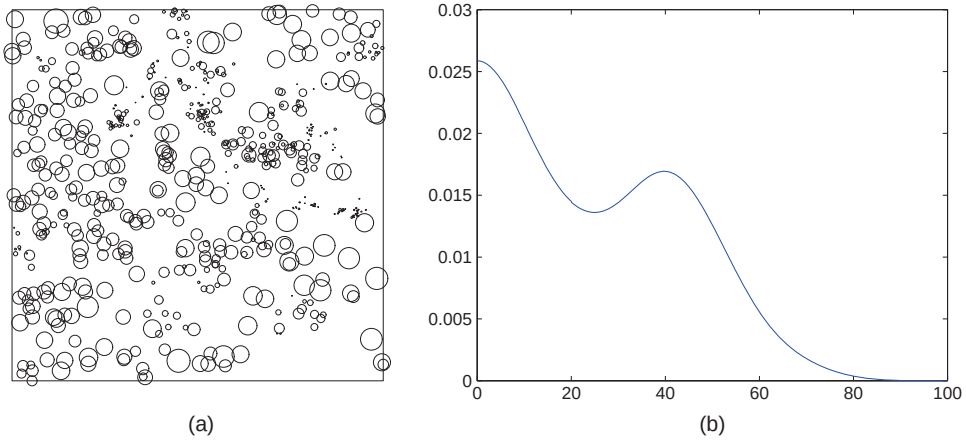

Figure 4.3 (a) Locations of longleaf pines in a North American forest. The circles indicate the stem diameters. They can be interpreted as marks for the point process of locations. (b) Statistically determined mark (stem diameter in cm) probability density function for the point pattern of longleaf pine trees.

Rotations act on marked point processes by rotating the points but *not* altering the marks. (However, in the case of set-marks as above this may be not so natural: it may be appropriate to subject the set-marks to rotations as well; see Baddeley and Jensen, 2005, p. 271; in contrast, Rau and Chiu, 2011, considered independent random rotations of the set-marks without moving the points.)

4.2.2 Intensity and mark distribution

The distribution of a marked point process is defined analogously as that of a unmarked point process. A marked point process is said to be *stationary* if for all x the translated process, Ψ_x has the same distribution as Ψ; it is *motion-invariant* if for all rigid motions $\boldsymbol{m}$, the process $\boldsymbol{m}\Psi$ has the distribution of Ψ.

The definition of the *intensity measure* Λ of a marked point process Ψ is analogous to that of the intensity measure of Ψ when Ψ is interpreted as an unmarked point process:

$$\Lambda(B \times L) = \mathbf{E}\big(\Psi(B \times L)\big). \tag{4.19}$$

$\Lambda(B \times L)$ is the mean number of points in B having mark in L.

A *Campbell theorem* (cf. Theorem 4.1) also holds:

$$\mathbf{E}\left(\sum_{[x;m]\in\Phi} f(x,m)\right) = \int f(x,m)\Lambda\big(\mathrm{d}(x,m)\big) \tag{4.20}$$

for any nonnegative measurable function f on $\mathbb{R}^d \times \mathbb{M}$.

The measure $\Lambda(\cdot \times L)$ for fixed L in $\mathcal{M}$ is absolutely continuous with respect to the intensity measure Λ_{g} of the ground point process $\{x_n\}$. Consequently, it can be shown that

$$\Lambda\big(\mathrm{d}(x,m)\big) = M_x(\mathrm{d}m)\Lambda_{\mathrm{g}}(\mathrm{d}x), \tag{4.21}$$

where M_x is a probability measure on $[\mathbb{M}, \mathcal{M}]$, which can be interpreted as the *distribution of the mark* of a point at x. Section 4.4 explains the close relationship of M_x to Palm distribution theory. The Polish space property of $\mathbb{M}$ is needed here to establish that M_x is indeed a measure.

When Ψ is stationary

$$\Lambda(B \times L) = \mathbf{E}\big(\Psi(B \times L)\big) = \mathbf{E}\big(\Psi(B_x \times L)\big) = \Lambda(B_x \times L) \qquad \text{for all } x \in \mathbb{R}^d. \tag{4.22}$$

If L is a fixed member of $\mathcal{M}$, then $\Lambda(\cdot \times L)$ is a translation-invariant measure. Hence, by (1.83)

$$\Lambda(B \times L) = \lambda_L \nu_d(B). \tag{4.23}$$

The quantity λ_L is the *intensity of* Ψ *with respect to* L; it is the mean number of points of Ψ per unit volume with marks in L. Of course,

$$\lambda_{\mathbb{M}} = \lambda, \tag{4.24}$$

where λ is the intensity of the ground point process.

As a function of L the quantity λ_L is a measure on $[\mathbb{M}, \mathcal{M}]$; the quotients λ_L/λ provide the so-called *mark distribution* M of the stationary marked point process Ψ:

$$M(L) = \frac{\lambda_L}{\lambda}. \tag{4.25}$$

In the case of stationarity, the intensity measure satisfies

$$\Lambda(B \times L) = \lambda \cdot \nu_d(B) \cdot M(L) \tag{4.26}$$

and $M = M_x$ for all x, where M_x is given in Formula (4.21).

In the case of independent marks, M is simply the distribution of the m_n.

In the ergodic case the following holds for every nonnegative integrable function $h(m)$ and any convex averaging sequence $\{W_n\}$

$$\frac{1}{\nu_d(W_n)} \sum_{i:\, x_i \in W_n} h(m_i) \to \int h(m) M(\mathrm{d}m) \qquad \text{as } n \to \infty. \tag{4.27}$$

Thus M can be interpreted as the distribution of the mark of a 'randomly chosen point' of Ψ:

$$M(L) = \lim_{n\to\infty} \frac{\Psi(W_n \times L)}{\Psi(W_n \times \mathbb{M})} \tag{4.28}$$

for convex averaging sequences $\{W_n\}$. In this book the term *typical point* is often used instead of 'randomly chosen point'.

In the stationary case the Campbell theorem (4.20) takes the form

$$\mathbf{E}\left(\sum_{[x;m]\in\Psi} f(x, m) \right) = \lambda \int \int f(x, m) M(\mathrm{d}m)\, \mathrm{d}x. \tag{4.29}$$

A typical application of this theorem is to establish Formula (6.124). An important particular case is

$$\mathbf{E}\left(\sum_{[x;m]\in\Psi} \mathbf{1}_B(x) h(m) \right) = \lambda \int h(m) M(\mathrm{d}m) \tag{4.30}$$

for any Borel set B with $\nu_d(B) = 1$, and for $h(m)$ a nonnegative measurable function on $\mathbb{M}$.

In the case of real-valued marks the following characteristics are useful. The *mark distribution function* $F_M(m)$ is given by

$$F_M(m) = \int_{\mathbb{R}} \mathbf{1}_{(-\infty,x]}(k) M(\mathrm{d}k) \tag{4.31}$$

and the *mean mark* is

$$\overline{m} = \int_{-\infty}^{\infty} m\, \mathrm{d}F_M(m). \tag{4.32}$$

If the marks are positive, then the *mark sum measure* S_m can be considered:

$$S_m(B) = \sum_{[x;m]\in\Psi} m \mathbf{1}_B(x) \qquad \text{for Borel sets } B. \tag{4.33}$$

This is a random measure, namely the sum of the random marks of all points within B. In the stationary case its mean satisfies

$$\mathbf{E}\big(S_m(B)\big) = \lambda \overline{m} v_d(B). \tag{4.34}$$

The second-order properties of the mark sum measure are studied in Stoyan (1984c).

Example 4.1. *Longleaf pines in a North American forest (Cressie, 1993, p. 579ff)*
Figure 4.3(a) is the point pattern of the positions of 583 longleaf pines in a 200 m × 200 m region of a forest. The marks shown by circles are the stem diameters. Figure 4.3(b) shows a kernel-smoothed estimate of the mark probability density function (giving the stem diameter distribution). The estimated mean mark is

$$\hat{\overline{m}} = 26.8\,\mathrm{cm},$$

and the intensity of the ground point process is estimated as

$$\hat{\lambda} = 0.0146\,\mathrm{m}^{-2}.$$

In the remainder of this chapter most formulae are given for unmarked point processes. Corresponding formulae for a marked point process Ψ can be deduced either by interpreting Ψ as a point process in $\mathbb{R}^d \times \mathbb{M}$ or by noting that for each fixed L in $\mathcal{M}$ the ground process $\{x_n : [x_n; m_n] \in \Psi, m_n \in L\}$ is an unmarked point process.

4.3 Moment measures and related quantities

4.3.1 Moment measures

In the classical theory of random variables the moments (particularly mean and variance), generating functions, and Laplace–Stieltjes transforms are important tools and provide useful means of describing distributions. Point process theory has analogues to these. For example, numerical means and variances of random variables are replaced by moment *measures*, as described below and in more detail in Daley and Vere-Jones (2008, Section 9.5).

Throughout the section (except when noted otherwise), B, $B_1, \ldots, B_n$ stand for Borel sets in $\mathbb{R}^d$.

General definition

The n^{th} moment measure of the point process Φ is the measure $\mu^{(n)}$ defined on $\mathcal{B}^{nd}$ by

$$\begin{aligned}\int_{\mathbb{R}^{nd}} f(x_1, \ldots, x_n)\mu^{(n)}\big(\mathrm{d}(x_1, \ldots, x_n)\big) &= \int_{\mathbb{N}} \sum_{x_1, \ldots, x_n \in \varphi} f(x_1, \ldots, x_n) P(\mathrm{d}\varphi) \\ &= \mathbf{E}\left(\sum_{x_1, \ldots, x_n \in \Phi} f(x_1, \ldots, x_n) \right), \end{aligned} \tag{4.35}$$

where f is any nonnegative measurable function on $\mathbb{R}^{nd}$. In particular, if $f(x_1, \ldots, x_n) = \mathbf{1}_{B_1}(x_1) \cdot \ldots \cdot \mathbf{1}_{B_n}(x_n)$, then

$$\mu^{(n)}(B_1 \times \cdots \times B_n) = \mathbf{E}\big(\Phi(B_1) \cdot \ldots \cdot \Phi(B_n)\big), \tag{4.36}$$

and if $B_1 = \cdots = B_n = B$, then

$$\mu^{(n)}(B^n) = \mathbf{E}\big(\Phi(B)^n\big). \tag{4.37}$$

Thus $\mu^{(n)}$ yields the n^{th} moment of the real-valued random variable $\Phi(B)$, which is the number of points in B.

Particular cases

$$n = 1: \qquad \mu^{(1)}(B) = \mathbf{E}\big(\Phi(B)\big) = \Lambda(B); \tag{4.38}$$

$$n = 2: \qquad \mu^{(2)}(B_1 \times B_2) = \mathbf{E}\big(\Phi(B_1)\Phi(B_2)\big), \tag{4.39}$$

$$\mathbf{var}\big(\Phi(B)\big) = \mu^{(2)}(B \times B) - \big(\Lambda(B)\big)^2. \tag{4.40}$$

The covariance of the random variables $\Phi(B_1)$ and $\Phi(B_2)$, that is, of the point numbers in two different sets, satisfies

$$\begin{aligned} \mathbf{cov}\big(\Phi(B_1), \Phi(B_2)\big) &= \mathbf{E}\big(\Phi(B_1)\Phi(B_2)\big) - \mathbf{E}\big(\Phi(B_1)\big)\mathbf{E}\big(\Phi(B_2)\big) \\ &= \mu^{(2)}(B_1 \times B_2) - \Lambda(B_1)\Lambda(B_2). \end{aligned} \tag{4.41}$$

If Φ is stationary, then the $\mu^{(n)}$ are translation-invariant in an extended sense:

$$\mu^{(n)}(B_1 \times \cdots \times B_n) = \mu^{(n)}\big((B_1 + x) \times \cdots \times (B_n + x)\big) \qquad \text{for all } x \text{ in } \mathbb{R}^d. \tag{4.42}$$

Section 4.5 gives a simplified description of the second moment measure in the stationary case.

4.3.2 Factorial moment measures

The n^{th}-order factorial moment measure $\alpha^{(n)}$ of the point process Φ is defined on $\mathcal{B}^{nd}$ by

$$\int f(x_1, \ldots, x_n)\, \alpha^{(n)}\big(\mathrm{d}(x_1, \ldots, x_n)\big) = \int \sum_{x_1, \ldots, x_n \in \varphi}^{\neq} f(x_1, \ldots, x_n) P(\mathrm{d}\varphi). \tag{4.43}$$

where, again, f is any nonnegative measurable function on $\mathbb{R}^{nd}$. Now the summation is over all n-tuples of *distinct* points in φ including all permutations of given points, as indicated by notation $\Sigma^{\neq}$. This is the difference between $\alpha^{(n)}$ and $\mu^{(n)}$. In (4.43) the sum omits all n-tuples with two or more equal members. Thus if $B_1, \ldots, B_n$ are pairwise disjoint, then

$$\mu^{(n)}(B_1 \times \cdots \times B_n) = \alpha^{(n)}(B_1 \times \cdots \times B_n). \tag{4.44}$$

For $n = 2$,

$$\mu^{(2)}(B_1 \times B_2) = \Lambda(B_1 \cap B_2) + \alpha^{(2)}(B_1 \times B_2) \tag{4.45}$$

as can be seen from the definitions of $\mu^{(2)}$, Λ and $\alpha^{(2)}$, and the relation

$$\sum_{x_1,\, x_2 \in \varphi} \mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_2) = \sum_{x_1 \in \varphi} \mathbf{1}_{B_1 \cap B_2}(x_1) + \sum_{x_1,\, x_2 \in \varphi}^{\neq} \mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_2).$$

Since

$$\alpha^{(n)}(B^n) = \mathbf{E}\big(\Phi(B)(\Phi(B)-1)\cdot\ldots\cdot(\Phi(B)-n+1)\big), \tag{4.46}$$

the quantity $\alpha^{(n)}(B^n)$ is the n^{th}-order factorial moment of the random variable $\Phi(B)$: this motivates the term 'factorial moment measure'.

The factorial moment measures for a Poisson process are given in Section 2.3.3.

4.3.3 Product densities

Suppose that $\alpha^{(n)}$ is locally finite and absolutely continuous with respect to Lebesgue measure ν_{nd}. Then $\alpha^{(n)}$ has a density $\varrho^{(n)}$, the n^{th}*-order product density*:

$$\alpha^{(n)}(B_1 \times \cdots \times B_n) = \int_{B_n} \ldots \int_{B_1} \varrho^{(n)}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n. \tag{4.47}$$

Moreover, for any nonnegative bounded measurable function f

$$\mathbf{E}\left(\sum_{x_1,\ldots,\, x_n \in \Phi}^{\neq} f(x_1, \ldots, x_n) \right) = \int \ldots \int f(x_1, \ldots, x_n)\varrho^{(n)}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

The product densities have an intuitive infinitesimal interpretation, which probably accounts for their historical precedence over product measures. Suppose that $C_1, \ldots, C_n$ are pairwise disjoint balls with centres $x_1, \ldots, x_n$ and infinitesimal volumes $\mathrm{d}V_1, \ldots, \mathrm{d}V_n$, respectively. Then $\varrho^{(n)}(x_1, \ldots, x_n)\, \mathrm{d}V_1 \ldots \mathrm{d}V_n$ is the probability that there is a point of Φ in each of $C_1, \ldots, C_n$.

Clearly, $\varrho^{(1)}(x)$ is the same as the intensity function $\lambda(x)$.

For pairwise disjoint $B_1, \ldots, B_n$, the Formulae (4.44) and (4.47) yield

$$\mu^{(n)}(B_1 \times \cdots \times B_n) = \int_{B_1} \ldots \int_{B_n} \varrho^{(n)}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

If Φ is stationary, then $\varrho^{(1)} \equiv \lambda$, and $\varrho^{(2)}$ depends only on the difference of its arguments:

$$\varrho^{(2)}(x_1, x_2) = f_{\text{st}}(x_1 - x_2) \qquad \text{for all } x_1, x_2 \in \mathbb{R}^d.$$

If furthermore Φ is motion-invariant, then this f_s depends only on the distance r between x_1 and x_2:

$$\varrho^{(2)}(x_1, x_2) = f_{\text{inv}}(\|x_1 - x_2\|) = f_{\text{inv}}(r).$$

The function $f_{\text{inv}}(r)$ is often still denoted by $\varrho^{(2)}(r)$ and is also called the second-order product density.

Product densities for the Poisson process can be found in Sections 2.3.3 and 2.4, and for other point processes models in Chapter 5.

4.3.4 The Campbell measure

The Campbell measure $\mathscr{C}$ is defined as a measure on $[\mathbb{R}^d \times \mathbb{N}, \mathcal{B}^d \times \mathcal{N}]$ by

$$\int \sum_{x\in\varphi} f(x,\varphi)P(\mathrm{d}\varphi) = \int f(x,\varphi)\mathscr{C}\big(\mathrm{d}(x,\varphi)\big), \tag{4.48}$$

where f is any nonnegative measurable function on $\mathbb{R}^d \times \mathbb{N}$. Since

$$\mathscr{C}(B\times Y) = \int \varphi(B)\mathbf{1}_Y(\varphi)P(\mathrm{d}\varphi) \tag{4.49}$$

for Borel $B \in \mathcal{B}^d$ and $Y \in \mathcal{N}$, one can also write

$$\mathscr{C}(B\times Y) = \mathbf{E}\big(\Phi(B)\mathbf{1}_Y(\Phi)\big) = \mathbf{E}\big(\Phi(B)\mathbf{1}(\Phi\in Y)\big). \tag{4.50}$$

It is sometimes useful to consider the *reduced Campbell measure* $\mathscr{C}^!$ defined by

$$\begin{aligned}\int \sum_{x\in\varphi} f(x,\varphi\setminus\{x\})P(\mathrm{d}\varphi) &= \int \sum_{x\in\varphi} f(x,\varphi-\delta_x)P(\mathrm{d}\varphi)\\ &= \int f(x,\varphi)\mathscr{C}^!\big(\mathrm{d}(x,\varphi)\big).\end{aligned} \tag{4.51}$$

(Note that Daley and Vere-Jones, 2008, used the term *modified Campbell measure* for $\mathscr{C}^!$.) Here $\varphi\setminus\{x\}$ and $\varphi-\delta_x$ are alternative notations, in set-theoretic and measure-theoretic language respectively, for the point pattern φ with the point $x\in\varphi$ deleted.

Campbell measures and reduced Campbell measures of higher order can also be defined; see Daley and Vere-Jones (2008, Section 13.1). For example, the second-order Campbell measure is

$$\mathscr{C}^{(2)}(B_1\times B_2\times Y) = \mathbf{E}\big(\Phi(B_1)\Phi(B_2)\mathbf{1}(\Phi\in Y)\big).$$

4.3.5 The mark correlation function

Moment measures for marked point processes can be defined in a fashion similar to the unmarked case. Details are given in Illian *et al.* (2008) both for discrete and continuous (i.e. real-valued) marks.

Here only one example of a useful second-order characteristic for marked point processes is considered, to give the reader some idea: the *mark correlation function* $k_{mm}(r)$ for a motion-invariant marked point process $\Psi = \{[x_n; m_n]\}$ with continuous marks. Heuristically $k_{mm}(r)$ is a normalised mean of the product of marks of points at two positions separated by a distance r, under the condition that there are indeed points of Ψ in these two positions. By motion-invariance it suffices to consider the positions o and $\mathbf{r}$, with $r = \|\mathbf{r}\|$. So

$$k_{mm}(r) = \frac{\mathbf{E}_{o,\mathbf{r}}\big(m(o)m(\mathbf{r})\big)}{\overline{m}^2} \qquad \text{for } r>0, \tag{4.52}$$

where $\mathbf{E}_{o,\mathbf{r}}$ is expectation subject to the conditioning that Ψ has points at positions o and $\mathbf{r}$ with marks $m(o)$ and $m(\mathbf{r})$; $\overline{m}$ is the mean mark.

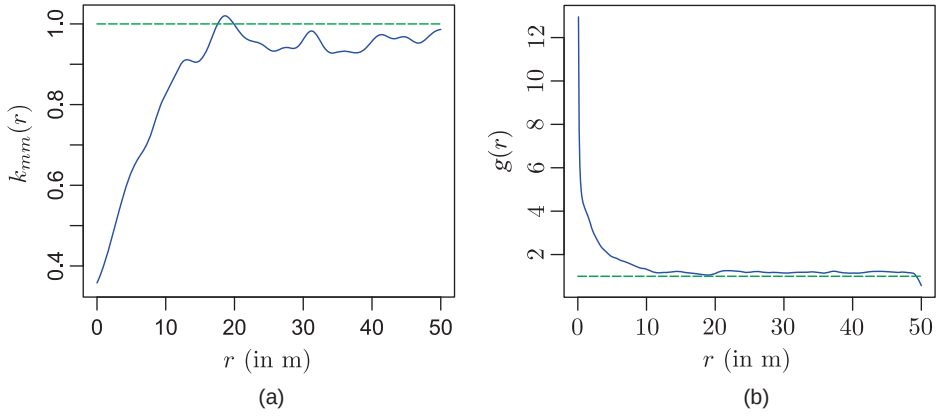

Figure 4.4 (a) Statistically estimated mark correlation function $k_{mm}(r)$ for the longleaf pines stem diameters. This function tends to increase with r up to $r = 15$ m, and the values smaller than 1 indicate that trees close together tend to have diameters smaller than the mean diameter. The distance $r = 15$ m may be seen as an estimate of the range of correlation. (b) Statistically estimated pair correlation function for the longleaf pines. Comparing it with Figure 4.7 suggests that it is similar to the pair correlation function of a Poisson cluster process.

Example 4.1 (Continued). *Longleaf pines in a North American forest*

The mark correlation function $k_{mm}(r)$ was calculated using as marks the stem diameters of the 583 pine trees. The result shown in Figure 4.4(a) is as perhaps expected: for small r (≤ 10 m) $k_{mm}(r)$ is smaller than one. This indicates that trees close together tend to have smaller diameters than the typical tree; this is the price trees have to pay for being close together. That this correlation goes only up to 15 m shows that the forest considered has only a weak short-range order; it is similar in most forests. For completeness Figure 4.4(b) shows also the estimated pair correlation function. It looks like the pair correlation function of a Poisson cluster process; see also the discussion in Cressie (1993) and Stoyan and Stoyan (1996).

Similar studies for other forests and for other tree characteristics such as damage characteristics can be found in Penttinen *et al.* (1992), Stoyan and Penttinen (2000) and Illian *et al.* (2008); growth functions as marks appear in Comas *et al.* (2013). Stoyan (1993) analysed a collage of Hans Arp by this method.

Stoyan (1984c) and Penttinen and Stoyan (1989) gave a mathematical description of $k_{mm}(r)$. The main idea is as follows: consider the second-order factorial moment measure of the marked point process in question

$$\alpha_m^{(2)}(B_1 \times B_2 \times L_1 \times L_2) = \mathbf{E}\left(\sum_{\substack{[x_1;m_1]\in\Psi \\ [x_2;m_2]\in\Psi}}^{\neq} \mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_2)\mathbf{1}_{L_1}(m_1)\mathbf{1}_{L_2}(m_2)\right),$$

where the sum is taken over all pairs $[x_1; m_1]$ and $[x_2; m_2]$ in Ψ but excluding the cases when $x_1 = x_2$, as indicated by the symbol $\Sigma^{\neq}$. Let $\alpha^{(2)}$ be the second-order factorial moment

measure of the unmarked ground point process $\Phi = \{x_n\}$. Then for fixed Borel subsets L_1 and L_2 of $[0, \infty)$, the measure $\alpha_m^{(2)}(\cdot \times \cdot \times L_1 \times L_2)$ is absolutely continuous with respect to $\alpha^{(2)}$. If $\alpha_m^{(2)}$ is σ-finite, then by the Radon–Nikodym theorem there is a 'density' M_{x_1,x_2} such that

$$\alpha_m^{(2)}(B_1 \times B_2 \times L_1 \times L_2) = \int_{B_1 \times B_2} M_{x_1,x_2}(L_1 \times L_2)\, \alpha^{(2)}\big(\mathrm{d}(x_1, x_2)\big).$$

For fixed x_1 and x_2 the term M_{x_1,x_2} can be interpreted as a measure on $\mathbb{M} \times \mathbb{M}$, the 'two-point mark distribution'. It gives the joint distribution of the marks of two points at the locations x_1 and x_2, under the condition that there are indeed points of Ψ at x_1 and x_2.

Under the assumption of motion-invariance, M_{x_1,x_2} depends only on $r = \|x_1 - x_2\|$ and the simpler notation M_r is used. In that case

$$k_{mm}(r) = \int m_1 m_2 M_r\big(\mathrm{d}(m_1, m_2)\big) \Big/ \overline{m}^2 \qquad \text{for } r \geq 0. \tag{4.53}$$

For more details see Illian *et al.* (2008), as well as Schlather (2001a), who also discussed the case of $r = 0$.

4.3.6 The probability generating functional

Recall from probability theory that the probability generating function G_X of a nonnegative integer-valued random variable X is given by

$$G_X(t) = \mathbf{E}(t^X) = \sum_{n=0}^{\infty} t^n \mathbf{P}(X = n) \qquad \text{for } t \in [0, 1].$$

The mean and variance of X can be derived from G_X:

$$\mathbf{E}(X) = G'(1) \qquad \text{and} \qquad \mathbf{var}(X) = G''(1) + G'(1) \cdot \big(1 - G'(1)\big)$$

under suitable regularity conditions.

The probability generating functional G of a point process $\Phi = \{x_n\}$ is defined by analogy with G_X. Let $\mathbf{U}$ be the family of all nonnegative bounded measurable functions u on $\mathbb{R}^d$ whose support $\{x \in \mathbb{R}^d : u(x) > 0\}$ is bounded. Furthermore let $\mathbf{V}$ be the family of all functions $v = 1 - u$ for u in $\mathbf{U}$, $0 \leq u \leq 1$. Then G of Φ is defined by

$$\begin{aligned} G(v) = \mathbf{E}(v(x_1)v(x_2) \cdot \ldots) &= \mathbf{E}\left(\prod_{x \in \Phi} v(x)\right) \\ &= \int \prod_{x \in \varphi} v(x) P(\mathrm{d}\varphi) \qquad \text{for } v \in \mathbf{V}. \end{aligned} \tag{4.54}$$

Since Φ is locally finite, the functional G is well-defined for any measurable function $v = 1 - u$ on $\mathbb{R}^d$ such that $0 \leq u \leq 1$ and $\int \big(1 - u(x)\big)\Lambda(\mathrm{d}x) < \infty$; see Daley and Vere-Jones (2008, p. 59).

The distribution P of Φ is determined uniquely by G. The joint probability generating function $G_{\Phi(B_1),\,\ldots,\,\Phi(B_n)}$ of the random vector $(\Phi(B_1), \ldots, \Phi(B_n))$ for bounded Borel sets $B_1, \ldots, B_n$ is given by

$$G_{\Phi(B_1),\,\ldots,\,\Phi(B_n)}(t_1, \ldots, t_n) = \mathbf{E}(t_1^{\Phi(B_1)} \cdot \ldots \cdot t_n^{\Phi(B_n)}) = G(v),$$

where $v(x) = v_1(x) \cdot \ldots \cdot v_n(x)$ in which $v_i(x) = 1 + (t_i - 1)\mathbf{1}_{B_i}(x)$ and $t_1, \ldots, t_n$ all belong to $[0, 1]$. Thus in particular for $n = 1$

$$G_{\Phi(B)}(t) = G(v) \qquad \text{with } v(x) = 1 + (t - 1)\mathbf{1}_B(x).$$

The factorial moment measures of Φ can be derived from G

$$\alpha^{(n)}(B_1 \times \cdots \times B_n) = (-1)^n \left(\left(\frac{\partial}{\partial s_1} \ldots \frac{\partial}{\partial s_n} \right) G(1 - s_1\mathbf{1}_{B_1} - \cdots - s_n\mathbf{1}_{B_n}) \right) \Bigg|_{s_1 = \cdots = s_n = 0}.$$

Conversely, under suitable regularity conditions the probability generating functional is determined by the factorial moment measures; for u in $\mathbf{U}$ with $0 \le u \le 1$

$$G(1 - u) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int u(x_1) \cdot \ldots \cdot u(x_n) \alpha^{(n)}\big(\mathrm{d}(x_1, \ldots, x_n)\big)$$

if the $\alpha^{(n)}$ are locally finite measures for all n and if the right-hand side converges.

Example 4.2. *The probability generating functional of the Poisson process*
Let Φ be a Poisson process of intensity measure Λ. Then

$$G(v) = \exp\left(- \int \big(1 - v(x)\big) \Lambda(\mathrm{d}x) \right) \qquad \text{for } v \in \mathbf{V}. \tag{4.55}$$

Proof. Consider $v(x) = 1 - \sum_{i=1}^{n} (1 - t_i)\mathbf{1}_{C_i}(x)$ for t_i all in $[0, 1]$ and $C_1, \ldots, C_n$ pairwise disjoint compact sets. Then

$$\begin{aligned} G(v) &= \int t_1^{\varphi(C_1)} \cdot \ldots \cdot t_n^{\varphi(C_n)} P(\mathrm{d}\varphi) \\ &= \mathbf{E}(t_1^{\Phi(C_1)} \cdot \ldots \cdot t_n^{\Phi(C_n)}) \\ &= \mathbf{E}(t_1^{\Phi(C_1)}) \cdot \ldots \cdot \mathbf{E}(t_n^{\Phi(C_n)}), \end{aligned}$$

where the last step uses the independence of $\Phi(C_1), \ldots, \Phi(C_n)$ for disjoint $C_1, \ldots, C_n$. Then each $\mathbf{E}(t_i^{\Phi(C_i)})$ can be determined, since $\Phi(C_i)$ has a Poisson distribution of mean $\Lambda(C_i)$. Therefore

$$\begin{aligned} G(v) &= \exp\bigl(-(1-t_1)\Lambda(C_1)\bigr) \cdot \ldots \cdot \exp\bigl(-(1-t_n)\Lambda(C_n)\bigr) \\ &= \exp\left(-\sum_{i=1}^{n} \int_{C_i} (1-t_i)\Lambda(\mathrm{d}x)\right) \\ &= \exp\left(-\int \bigl(1-v(x)\bigr)\Lambda(\mathrm{d}x)\right). \end{aligned}$$

Thus (4.55) holds for piecewise constant v. The general result follows by standard arguments from measure theory. □

A point process can be considered also as a random counting measure. Thus ideas of the theory of random measures can be transferred to point processes. For example, the Laplace functional of a point process can be defined analogously to Section 7.2.1.

4.4 Palm distributions

4.4.1 Heuristic introduction

In the study of point processes it is often natural to ask questions concerning the conditional expectations or distributions of some random quantities, given that there is a point of the process at the fixed location x. Section 2.3.4 gives an elementary example, which is related to the nearest-neighbour distance distribution and which is reconsidered in this section. Suppose $\mathbf{P}\bigl(\Phi(B(x,r)) = 1 \parallel x\bigr)$ represents the conditional probability, given that Φ has a point at x, that there is no further point within the ball of radius r centred at x. This conditional probability is ambiguous without further explanation since the conditioning event has probability zero. It must be defined by requiring it to satisfy some extra properties to be described in the following.

What properties might one reasonably require of such conditional distributions $\mathbf{P}(\cdot \parallel x)$? To begin with one would require in the stationary case that the relation

$$\mathbf{P}(\Phi \in Z_x \parallel x) = \mathbf{P}(\Phi \in Z \parallel o)$$

holds for $Z \in \mathcal{N}$ and $Z_x = \{\varphi_x : \varphi \in Z\}$. Substituting $Z = Y_{-x}$ into it yields

$$\mathbf{P}(\Phi \in Y \parallel x) = \mathbf{P}(\Phi_x \in Y \parallel o) = \mathbf{P}(\Phi \in Y_{-x} \parallel o) \tag{4.56}$$

for $x \in \mathbb{R}^d$ and $Y \in \mathcal{N}$, where Y_{-x} is defined in the same manner as Y_x after Formula (4.1). Given (4.56) one could define a function

$$D(r) = 1 - \mathbf{P}\bigl(\Phi(B(o,r)) = 1 \parallel o\bigr),$$

which could be interpreted as the distribution function of the distance of the typical point to its nearest neighbour. (If it is sure that the ball $B(o,r)$ contains o as a point of Φ, then the event $\Phi \in Z$ with $Z = \{\varphi \in \mathbb{N} : \varphi\bigl(B(o,r)\bigr) = 1\}$ is the same as that the distance from o to its nearest neighbour is larger than r.)

In order to introduce further properties required, consider again the example related to $D(r)$. Suppose that Φ is a point process, not necessarily stationary, of distribution P and locally finite intensity measure Λ. For the set Y in $\mathcal{N}$ given by

$$Y = \{\varphi \in \mathbb{N} : \varphi\big(B(o, r)\big) = 1\},$$

consider the function h on $\mathbb{R}^d \times \mathbb{N}$ defined by

$$h(x, \varphi) = \mathbf{1}_B(x)\mathbf{1}_Y(\varphi - x)$$

for some bounded Borel set B; as usual if $\varphi = \{x_n\}$, then

$$\varphi_{-x} = \varphi - x = \{x_n - x\}.$$

With the distribution function $D(r)$ given above in mind, one might wish to evaluate the mean number of points of Φ in B whose neighbours are all at distance at least r:

$$\mathbf{E}\left(\sum_{x\in\Phi} h(x, \Phi)\right) = \int \sum_{x\in\varphi} h(x, \varphi)P(\mathrm{d}\varphi).$$

More generally one might wish to evaluate this for arbitrary nonnegative measurable functions h on $\mathbb{R}^d \times \mathbb{N}$.

If $\mathbb{R}^d$ is partitioned into domains $D_1, D_2, \ldots$ each of positive volume, then the expectation above can be written as

$$\mathbf{E}\left(\sum_{x\in\Phi} h(x, \Phi)\right) = \sum_k \mathbf{E}\left(\sum_{x\in\Phi\cap D_k} h(x, \Phi) \;\Big|\; \Phi(D_k) > 0\right) \cdot \mathbf{P}\big(\Phi(D_k) > 0\big).$$

Suppose that each D_k tends to an infinitesimal volume element $\mathrm{d}x$. Then $\mathbf{P}\big(\Phi(D_k) > 0\big)$ should converge to $\Lambda(\mathrm{d}x)$, while the conditional mean should converge to the mean $\mathbf{E}\big(h(x, \Phi) \parallel x\big)$ of $h(x, \varphi)$ with respect to $\mathbf{P}(\cdot \parallel x)$. Hence one is led to the relation

$$\mathbf{E}\left(\sum_{x\in\Phi} h(x, \Phi)\right) = \int \mathbf{E}\big(h(x, \Phi) \parallel x\big)\Lambda(\mathrm{d}x). \tag{4.57}$$

Under the assumption that Φ is stationary, it is $\Lambda(\cdot) = \lambda\nu_d(\cdot)$ with $0 < \lambda < \infty$. The relations (4.56) and (4.57) then yield

$$\mathbf{E}\left(\sum_{x\in\Phi} h(x, \Phi)\right) = \lambda \int \mathbf{E}\big(h(x, \Phi_x) \parallel o\big)\,\mathrm{d}x. \tag{4.58}$$

If the notation $P_o(Y) = \mathbf{P}(\Phi \in Y \parallel o)$ is used for $Y \in \mathcal{N}$, then (4.58) takes the form

$$\mathbf{E}\left(\sum_{x\in\Phi} h(x, \Phi)\right) = \lambda \int \int h(x, \varphi_x) P_o(\mathrm{d}\varphi)\, \mathrm{d}x \tag{4.59}$$

and so for $h(x, \varphi) = \mathbf{1}_B(x) \cdot \mathbf{1}_Y(\varphi_{-x})$ as above

$$\int \sum_{x\in\varphi\cap B} \mathbf{1}_Y(\varphi - x) P(\mathrm{d}\varphi) = \lambda \nu_d(B) P_o(Y) \tag{4.60}$$

for Borel $B \in \mathcal{B}^d$ and $Y \in \mathcal{N}$.

In the following, a distribution on $[\mathbb{N}, \mathcal{N}]$ will be constructed which will have the behaviour of the hypothetical P_o above: it will satisfy (4.59) and its definition will be inspired by (4.60).

4.4.2 The Palm distribution: First definition

Suppose that Φ is a stationary point process, with finite nonzero intensity λ. The *Palm distribution* (at o) of P is a distribution defined on $[\mathbb{N}, \mathcal{N}]$ by

$$P_o(Y) = \frac{1}{\lambda \nu_d(B)} \int \sum_{x\in\varphi\cap B} \mathbf{1}_Y(\varphi_{-x}) P(\mathrm{d}\varphi) \qquad \text{for } Y \in \mathcal{N}, \tag{4.61}$$

where B is an arbitrary Borel set of positive volume; the definition does not depend on the choice of the set B.

The definition can be understood intuitively using an approach of Matthes, based on ideas from the theory of marked point processes as follows. To each point x of Φ a mark of 1 or 0 is given according as to whether the shifted process Φ_{-x} belongs to the configuration set Y or not. For example, consider $Y = \{\varphi \in \mathbb{N} : \varphi \cap B(o, r) = \{o\}\}$. Then x has mark 1 precisely when its nearest neighbour is further away than r. The result of this is a stationary marked point process of marks 0 and 1. Its mark distribution M, defined in (4.25), is a 'ratio of rates', given by

$$M(\{1\}) = \frac{\lambda_{\{1\}}}{\lambda},$$

where $\lambda_{\{1\}}$ is the intensity of the subpoint process of points with mark 1. (It is analogous for the points with mark 0.) Thus for any fixed Borel set B the number of points of Φ in B of mark 1 has the mean $\lambda_{\{1\}} \cdot \nu_d(B)$. The same value is yielded by the integral in (4.61) and so

$$P_o(Y) = M(\{1\}), \tag{4.62}$$

and the choice of B does not play a rôle.

Note in passing that this approach is closely related to methods of statistical estimation of Palm probabilities, see around Formula (4.128).

If Φ is ergodic, then the Palm probability $P_o(Y)$ may be interpreted as the probability that the typical point x of Φ has the property that Φ_{-x}, the point process shifted with x into the origin o, belongs to Y. This is comparable with the ergodic interpretation of a mark distribution as on p. 119. The following ergodic theorem holds:

$$\lim_{n\to\infty} \frac{1}{\Phi(W_n)} \sum_{x_i \in W_n} f(\Phi_{-x_i}) = \int f(\varphi) P_o(\mathrm{d}\varphi). \tag{4.63}$$

It is plausible when one considers the definition of the Palm distribution in (4.61) and Formula (4.15).

Because of (4.61) the Palm distribution can be shown to satisfy the so-called *Campbell–Mecke theorem*:

Theorem 4.2. *(Mecke, 1967) For any nonnegative measurable function h on $\mathbb{R}^d \times \mathbb{N}$,*

$$\begin{aligned} \mathbf{E}\left(\sum_{x\in\Phi} h(x, \Phi)\right) &= \int \sum_{x\in\varphi} h(x, \varphi)\, P(\mathrm{d}\varphi) \\ &= \lambda \int \int h(x, \varphi_x) P_o(\mathrm{d}\varphi)\, \mathrm{d}x. \end{aligned} \tag{4.64}$$

Substituting $h(x, \varphi) = f(x)$ into (4.64) gives (4.10), the simple Campbell theorem.

Many problems in point process theory can be resolved by calculations of quantities in the form of the left-hand side of (4.64). The Campbell–Mecke theorem helps reduce these calculations to relatively simple integrations.

4.4.3 The Palm distribution: Second definition

The approach in this section uses more measure-theoretic tools, namely the Radon–Nikodym theorem and the Campbell measure $\mathscr{C}$. Suppose that the intensity measure Λ is σ-finite: this implies that $\mathscr{C}$ is likewise σ-finite, and furthermore, for any configuration set Y the measure $\mathscr{C}(\cdot \times Y)$ is absolutely continuous with respect to Λ. So there is a density function $d(x)$ with

$$\mathscr{C}(B \times Y) = \int_B d(x) \Lambda(\mathrm{d}x) \qquad \text{for Borel sets } B.$$

Since this density function depends on Y, it is denoted by $P_x(Y)$. Soon it will turn out that it can be interpreted as a distribution. Thus

$$\mathscr{C}(B \times Y) = \int_B P_x(Y) \Lambda(\mathrm{d}x). \tag{4.65}$$

Indeed, for fixed x, $P_x(\cdot)$ can be taken to be a distribution on $[\mathbb{N}, \mathcal{N}]$: the *Palm distribution of P with respect to x*. In the stationary case, with $0 < \lambda < \infty$, it *can be taken to satisfy* relation (4.56) and so

$$P_o(Y) = P_x(Y_x) \qquad \text{for } Y \in \mathcal{N}, \tag{4.66}$$

as

$$\lambda \int_B P_z(Y)\mathrm{d}z = \mathscr{C}(B \times Y) = \mathscr{C}(B_x \times Y_x)$$
$$= \lambda \int_{B_x} P_z(Y_x)\,\mathrm{d}z = \lambda \int_B P_{x+z}(Y_x)\,\mathrm{d}z$$

for all $B \in \mathcal{B}^d$, $x \in \mathbb{R}^d$ and $Y \in \mathcal{N}$. This implies $P_z(Y) = P_{x+z}(Y_x)$. Putting $z = 0$ yields (4.66).

The elliptical language here ('can be taken to satisfy ...') is used because relations such as (4.65) define $P_x(Y)$ for all x off a null-set that may *a priori* depend on Y. The measure-theoretic machinery of *regular conditional probabilities* (see Kallenberg, 2002) can be used to resolve this problem. Nevertheless, it can be ignored for the practical purposes of Palm distribution theory.

The Campbell–Mecke theorem given in Theorem 4.2 for stationary point processes is a consequence of this second definition:

$$\mathbf{E}\left(\sum_{x\in\Phi} h(x, \Phi)\right) = \int h(x, \varphi)\mathscr{C}\big(d(x, \varphi)\big)$$
$$= \lambda \int \int h(x, \varphi)P_x(\mathrm{d}\varphi)\,\mathrm{d}x = \lambda \int \int h(x, \varphi_x)P_o(\mathrm{d}\varphi)\,\mathrm{d}x. \qquad (4.67)$$

The two definitions agree in the stationary case (the first definition is not otherwise applicable).

A further possible way to define the Palm distribution locally is sketched in Section 2.3.4: it is to consider the conditional distribution of Φ, given that $\Phi\big(B(o, \varepsilon)\big) \geq 1$, and to take the limit as $\varepsilon \to 0$. Mathematical details can be found in Daley and Vere-Jones (2008, pp. 306–8).

4.4.4 Reduced Palm distributions

As the ordinary Palm distributions are related to $\mathscr{C}$, so the reduced Palm distributions are related to the reduced Campbell measure $\mathscr{C}^!$. If the appropriate replacement is made in (4.65), then the *reduced Palm distribution* $P_x^!$ is defined:

$$P_x^!(Y) = P(\Phi \setminus \{x\} \in Y \parallel x) \qquad \text{for } Y \in \mathcal{N}, \qquad (4.68)$$

where $P_x^!(Y)$ denotes the probability that Φ without point x belongs to the configuration set Y, under the condition that Φ has a point at x. In the stationary case, Formula (4.61) can be employed to define $P_o^!(Y)$ if $\mathbf{1}_Y(\varphi - x)$ is replaced by $\mathbf{1}_Y\big((\varphi - x) \setminus \{o\}\big)$.

Analogously as the usual Palm distribution, the reduced Palm distribution characterises the distribution of the point process under the condition that at o there is a point. But now the distribution of the point process is given with the point o removed after conditioning. This simplifies notation, as the following two examples show.

The *nearest-neighbour distribution function* $D(r)$ can be expressed via P_o and $P_o^!$, by

$$\begin{aligned} D(r) &= 1 - P_o\big(\varphi \in \mathbb{N} : \varphi(B(o,r)) = 1\big) \\ &= 1 - P_o^!\big(\varphi \in \mathbb{N} : \varphi(B(o,r)) = 0\big). \end{aligned}$$

The latter formula shows more clearly that the ball $B(o,r)$ is empty as it is expected for the complement of the nearest-neighbour distance distribution function.

Example 4.3. *The Palm distribution of a Poisson process*

The *theorem of Slivnyak–Mecke* gives the Palm distribution P_x of a Poisson process of intensity measure Λ and distribution P:

$$P_x = P * \delta_{\delta_x} \qquad \text{for all } x \in \mathbb{R}^d. \tag{4.69}$$

Here '$*$' denotes convolution of distributions, which corresponds to the superposition of point processes; see Section 5.1. The symbol δ_{δ_x} denotes the distribution of the degenerate point process that consists solely of the (nonrandom) point x. Formula (4.69) can be interpreted as

$$P_x(Y) = \mathbf{P}(\Phi \in Y \parallel x) = \mathbf{P}(\Phi \cup \{x\} \in Y) \qquad \text{for } Y \in \mathcal{N},$$

or

$$\int f(\varphi) P_x(\mathrm{d}\varphi) = \int f(\varphi \cup \{x\}) P(\mathrm{d}\varphi)$$

for all measurable nonnegative functions f. In words: the Palm distribution with respect to x is simply the distribution of the Poisson process plus an added point at x.

If the reduced Palm distribution is used, then (4.69) takes on a more elegant form, called *Mecke's formula*:

$$P_x^! = P \qquad \text{for all } x \in \mathbb{R}^d. \tag{4.70}$$

That is to say, the Palm conditioning does not change the distribution of the Poisson process, apart from a given point at x. An equivalent equation is

$$\mathbf{E}\left(\sum_{x \in \Phi} f(x, \Phi \setminus \{x\})\right) = \lambda \mathbf{E}\left(\int f(x, \Phi)\mathrm{d}x\right). \tag{4.71}$$

The Poisson process is characterised by (4.69) or (4.70); see Mecke (1968) and Daley and Vere-Jones (2008, p. 281).

Proof that a Poisson process satisfies Formula (4.69). Both $P * \delta_{\delta_x}$ and P_x are distributions of simple point processes. So by the discussion in Section 4.1 their equality is established if the corresponding systems of void-probabilities are equal, that is,

$$P * \delta_{\delta_x}(V_K) = P_x(V_K) \qquad \text{for all compact } K \text{ in } \mathbb{R}^d,$$

where $V_K = \{\varphi \in \mathbb{N} : \varphi(K) = 0\}$ is the set of all point sequences without a point in K.

Suppose A is any bounded Borel set. Then

$$\begin{aligned}\int_A (P * \delta_{\delta_x})(V_K)\Lambda(\mathrm{d}x) &= \int_{A\setminus K} P(V_K)\Lambda(\mathrm{d}x) = P(V_K)\cdot\Lambda(A\setminus K)\\ &= \mathbf{E}\big(\Phi(A\setminus K)\cdot\mathbf{1}(\Phi(K)=0)\big)\\ &= \mathscr{C}\big((A\setminus K)\times V_K\big),\end{aligned}$$

where the second line results from the independence property of the Poisson process. Clearly $\mathscr{C}\big((A\cap K)\times V_K\big) = \mathbf{E}\big(\Phi(A\cap K)\mathbf{1}(\Phi(K)=0)\big) = 0$. Hence

$$\int_A P * \delta_{\delta_x}(V_K)\Lambda(\mathrm{d}x) = \mathscr{C}(A\times V_K) = \int_A P_x(V_K)\Lambda(\mathrm{d}x)$$

by using (4.65). Since A was arbitrary, a standard argument of measure theory establishes (4.69). □

4.4.5 Isotropy of Palm distribution

Clearly the Palm distribution P_o is never a stationary distribution since under P_o a point process must always contain o; but Example 4.3 shows that $P_o^!$ can be stationary. However, a Palm distribution can be isotropic, that is, invariant with respect to rotations about the origin o: if the point process Φ is motion-invariant, then its Palm distribution is isotropic.

Proof. Rotation invariance means

$$P_o(Y) = P_o(\boldsymbol{r}Y) \qquad \text{for } Y\in\mathcal{N} \tag{4.72}$$

for every rotation $\boldsymbol{r}$ about o. To establish this, consider

$$\begin{aligned}\lambda P_o(Y)v_d\big(B(o,r)\big) &= \int\int_{B(o,r)} \mathbf{1}_Y(\varphi-x)\varphi(\mathrm{d}x)P(\mathrm{d}\varphi)\\ &= \int\int_{B(o,r)} \mathbf{1}_Y(\boldsymbol{r}^{-1}\varphi-x)\,\boldsymbol{r}^{-1}\varphi(\mathrm{d}x)P(\mathrm{d}\varphi) \qquad \text{(by isotropy of } \Phi)\\ &= \int\int_{B(o,r)} \mathbf{1}_Y\big(\boldsymbol{r}^{-1}(\varphi-y)\big)\varphi(\mathrm{d}y)P(\mathrm{d}\varphi) \qquad \text{(by letting } x=\boldsymbol{r}^{-1}y)\\ &= \int\int_{B(o,r)} \mathbf{1}_{\boldsymbol{r}Y}(\varphi-y)\varphi(\mathrm{d}y)P(\mathrm{d}\varphi)\\ &= \lambda P_o(\boldsymbol{r}Y)\cdot v_d\big(B(o,r)\big).\end{aligned}$$

□

Suppose a system of polar coordinates is based on the typical point of a motion-invariant point process. Then as a direct consequence of (4.72) the random angle α of the coordinates of the nearest neighbour is uniformly distributed over $(0, 2\pi]$; see Figure 4.5.

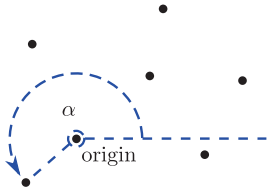

Figure 4.5 The angle α between a point x and its nearest neighbour $n(x)$.

4.4.6 Inversion formulae

It is theoretically possible to determine the distribution of Φ from λ and P_o if the point process Φ is stationary. Here, for example, is a simple relationship between Φ and P_o:

$$k\,\mathbf{P}\big(\Phi(B)=k\big)=\lambda\int_B P_o\big(\{\varphi\in\mathbb{N}:\varphi(B-x)=k\}\big)\,\mathrm{d}x \tag{4.73}$$

for $k = 1, 2, \ldots$ and all bounded Borel sets B. The stationary distribution can be retrieved by integration over the typical Voronoi cell (to be discussed in Chapter 9); see Daley and Vere-Jones (2008, p. 306):

Theorem 4.3. *For every integrable nonnegative function f, it holds*

$$\mathbf{E}\big(f(\Phi)\big)=\lambda\int\int_{\mathbb{R}^d} f(\varphi_{-x})\mathbf{1}\big(x\in C_0(\varphi)\big)\mathrm{d}x P_o(\mathrm{d}\varphi), \tag{4.74}$$

where $C_0(\varphi)$ is the zero cell of the Voronoi tessellation with respect to the point pattern φ.

4.4.7 *n*-fold Palm distributions

The n-fold Palm distributions of Φ are derived from its higher-order Campbell measures; see Kallenberg (1983a, 1986) and Hanisch (1982). An n-fold Palm distribution is a conditional distribution for a point process given that points of the process occur at n fixed positions. (The mark correlation function $k_{mm}(r)$ described on p. 123 is closely related to a two-fold Palm distribution.)

The n-fold Palm distribution of a Poisson process with distribution P is, analogously to (4.69), equal to

$$P*\delta_{\delta_{x_1}}*\cdots*\delta_{\delta_{x_n}}\qquad\text{for } x_1,\ldots,x_n\in\mathbb{R}^d. \tag{4.75}$$

This means, it is the distribution of a Poisson process plus n points at the (deterministic) locations $x_1, \ldots, x_n$.

4.4.8 Palm distributions for marked point processes

The definitions of these distributions follow the case of an unmarked point process, using either the Campbell measure or, in the stationary case, a generalised version of Formula (4.61). For

a marked point process Ψ with distribution P the Campbell measure is defined by

$$\mathscr{C}(B \times L \times Y) = \mathbf{E}\big(\Psi(B \times L)\mathbf{1}(\Psi \in Y)\big) \tag{4.76}$$

for Borel $B \in \mathcal{B}^d$, $L \in \mathcal{M}$ and $Y \in \mathcal{N}_M$, where $\mathcal{N}_M$ is the smallest σ-algebra on $\mathbb{N}_M$ (the family of all marked point sequences) that makes every mapping

$$\psi \mapsto \psi(B \times L)$$

measurable for all Borel $B \in \mathcal{B}^d$ and $L \in \mathcal{M}$.

Again, assume that Λ is σ-finite. Since for all Y the measure $\mathscr{C}(\cdot \times \cdot \times Y)$ is absolutely continuous with respect to Λ, the Radon–Nikodym derivative $P_x^m(Y)$ exists such that

$$\mathscr{C}(B \times L \times Y) = \int_{B \times L} P_x^m(Y)\, \Lambda\big(\mathrm{d}(x, m)\big). \tag{4.77}$$

Here, for fixed x and m, P_x^m is a distribution on $\mathcal{N}_M$. It can be interpreted as the probability of Y given that Ψ has a point at x of mark m.

From (4.26) and (4.66) in the stationary case, Formula (4.77) takes the form

$$\mathscr{C}(B \times L \times Y) = \lambda \int_B \int_L P_o^m(Y_{-x}) M(\mathrm{d}m)\, \mathrm{d}x, \tag{4.78}$$

where M is the mark distribution of Ψ.

The *Campbell–Mecke theorem* for a stationary marked point process is

$$\int \sum_{[x;m] \in \psi} h(x, m, \psi) P(\mathrm{d}\psi) = \lambda \int \int \int h(x, m, \psi_x) P_o^m(\mathrm{d}\psi) M(\mathrm{d}m)\, \mathrm{d}x \tag{4.79}$$

for any nonnegative measurable function h on $\mathbb{R}^d \times \mathbb{M} \times \mathbb{N}_M$.

Again in the stationary case, suppose that $L \in \mathcal{M}$ with $M(L) > 0$. Then one can define the *Palm distribution of* Ψ *with respect to the mark set* L, denoted by P_L, as

$$P_L(Y) = \int_L P_o^m(Y) M(\mathrm{d}m)/M(L) \qquad \text{for } Y \in \mathcal{N}_M.$$

It has an interpretation as the conditional distribution of Ψ given that Ψ has a point at o with mark in L.

The mark distribution M itself can be thought of as a Palm distribution. Set $Y_L = \{\psi \in \mathbb{N}_M : [o; m] \in \psi, m \in L\}$. Then

$$P_{\mathbb{M}}(Y_L) = \int_{\mathbb{M}} P_o^m(Y_L) M(\mathrm{d}m) = M(L),$$

since $P_o^m(Y_L)$ is 1 if $m \in L$, and 0 otherwise.

For brevity and convenience, throughout this book and also in the literature the phrase *typical point* is used. This is to be interpreted using Palm distributions and understood intuitively as a 'randomly chosen point'. For example, if under the Palm distribution the distance from o to the nearest point has an exponential distribution, then it is said that the distance of the typical point to its nearest neighbour is exponentially distributed. Similarly the statement that the mark of the typical point has probability p of lying in L means that $M(L) = p$.

4.4.9 Point-stationarity

Let Φ be a stationary point process with Palm distribution P_o. It has been noticed in Section 4.4.5 that P_o is not stationary. Still P_o has important invariance properties. To discuss these it is convenient to introduce the notion of a *Palm version* of Φ, that is a point process Φ^o with distribution P_o.

In one dimension it is possible to write Φ^o as a point sequence, $\Phi^o = \{T_n\}$, where n is running through all integers, with $T_0 := 0$ and $T_n < T_{n+1}$ for all n. It can be shown that

$$\mathbf{P}(\Phi^o_{-T_n} \in Y) = \mathbf{P}(\Phi^o \in Y) \tag{4.80}$$

for all integers n and all configuration sets Y. This means, moving the origin $o \in \Phi^o$ to the n^{th} point of Φ^o does not change the distribution of Φ^o, and the sequence $\{T_n - T_{n-1}\}$ of inter-point distances is stationary. Moreover, there is an essentially unique correspondence between such sequences and stationary point processes on the line; see for example Kallenberg (2002, Theorem 11.4) or Daley and Vere-Jones (2008, Theorem 13.3.I). These facts are of fundamental importance in one-dimensional point process theory and its applications.

In higher dimensions there is no obvious analogue of (4.80). Assume for instance that Φ is a homogeneous Poisson process. According to Example 4.3 one can then take $\Phi^o = \Phi \cup \{o\}$. Let $T \in \Phi$ be the nearest neighbour of o in Φ. Then Φ^o_{-T} has not the same distribution as Φ^o. This is because $-T$ is a point of Φ^o_{-T} that is closer to the origin than to any other point of Φ^o_{-T}. It is not hard to see that a Poisson process cannot have this property almost surely.

A generalisation of (4.80) can be based on the concept of the *bijective point map*, introduced in Thorisson (2000). A point map is a mapping $\tau : \mathbb{N} \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying

$$\tau(\varphi, x) \in \varphi \qquad \text{whenever } x \in \varphi,$$

and the covariance property

$$\tau(\varphi_x, y - x) = \tau(\varphi, y) - x \qquad \text{for all } \varphi \in \mathbb{N}, \text{ and } x, y \in \mathbb{R}^d. \tag{4.81}$$

A point map τ is called bijective if $\tau(\varphi, \cdot)$ is a bijection on φ for all $\varphi \in \mathbb{N}$. In this case

$$\mathbf{P}(\Phi^o_{-T} \in Y) = \mathbf{P}(\Phi^o \in Y) \qquad \text{for all } Y \in \mathcal{N}, \tag{4.82}$$

where $T = \tau(\Phi, o)$. This was proved in Thorisson (2000) and Heveling and Last (2005), but can also be derived from results in Mecke (1975). Note that (4.82) contains (4.80) as a special case.

An example of bijective point map τ is as follows: For $\varphi \in \mathbb{N}$ and $x \in \varphi$, let $\tau(\varphi, x) = y$ if $y \in \varphi$ is the unique nearest neighbour of $x \in \varphi$ and if x is the unique nearest neighbour of y. In all other cases let $\tau(\varphi, x) := x$. Hence (4.82) holds if $T = \tau(\Phi, o)$ and o are (unique) mutually nearest neighbours (even though $\mathbf{P}(T = o) > 0$ for most Φ). Other examples of bijective point maps can be found in Heveling and Last (2005), which also contains a construction of a bijective point map with $\mathbf{P}(T = o) = 0$.

Heveling and Last (2005) also proved that (4.82) is even characterising Palm versions. The attractive feature of this result is that (4.82) is an *intrinsic* invariance property, which does not refer to a stationary distribution. The first intrinsic characterisation of Palm measures was discovered in Mecke (1967).

Proof of (4.82). Let B be a Borel set of $\mathbb{R}^d$ with $v_d(B) = 1$. The definition (4.61) implies for all nonnegative measurable functions g on $\mathbb{N}$ that

$$\lambda \mathbf{E}\big(g(\Phi^o_{-T})\big) = \mathbf{E}\left(\sum_{x\in\Phi\cap B} g\big((\Phi_{-x})_{-\tau(\Phi-x,o)}\big)\right) = \mathbf{E}\left(\sum_{x\in\Phi\cap B} g(\Phi_{-\tau(\Phi,x)})\right),$$

where (4.81) has been used with $x = y$. For any $\varphi \in \mathbb{N}$ let $\tau^{-1}(\varphi, \cdot)$ denote the inverse of the bijection $\tau(\varphi, \cdot)$. It is easy to check that τ^{-1} has the covariance property (4.81). Now one can replace $\tau(\Phi, x)$ in the above summation by y and x by $\tau^{-1}(\Phi, y)$ to obtain

$$\begin{aligned}\lambda \mathbf{E}\big(g(\Phi^o_{-T})\big) &= \mathbf{E}\left(\sum_{y\in\Phi} \mathbf{1}\big(\tau^{-1}(\Phi, y) \in B\big) g(\Phi_{-y})\right) = \mathbf{E}\left(\sum_{y\in\Phi} \mathbf{1}\big(\tau^{-1}(\Phi_y, o) + y \in B\big) g(\Phi_{-y})\right)\\ &= \lambda \iint \mathbf{1}\big(\tau^{-1}(\varphi, o) + y \in B\big) g(\varphi)\,\mathrm{d}y P_o(\mathrm{d}\varphi),\end{aligned}$$

in which the Campbell–Mecke theorem (4.64) and the Fubini theorem have been used. Since the inner integration yields $g(\varphi)v_d(B) = g(\varphi)$, this implies $\mathbf{E}\big(g(\Phi^o_{-T})\big) = \mathbf{E}\big(g(\Phi^o)\big)$ and therefore (4.82). □

4.4.10 Stationary and balanced partitions

Given a homogeneous Poisson process Φ one may ask for the existence of a random variable T such that $T \in \Phi$ and Φ_{-T} have the same distribution as $\Phi \cup \{o\}$. This section will show that not only does such a T exist but also it may even be chosen as a function $T \equiv T(\Phi)$ of Φ. In fact the result can be formulated for any ergodic point process.

Let Φ be a stationary point process on $\mathbb{R}^d$ with finite intensity λ. A *stationary partition* (based on Φ) is a measurable mapping $\pi : \mathbb{N} \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying the covariance property (4.81) and such that almost surely $\pi(\Phi, x) \in \Phi$ for all $x \in \mathbb{R}^d$; see Last (2006, 2010). For $x \in \varphi \in \mathbb{N}$, the set

$$C_x(\varphi) = \{y \in \mathbb{R}^d : \pi(\varphi, y) = x\}$$

is referred to as *cell* with *centre* x. It is not required that $x \in C_x(\varphi)$ and some of the cells might be empty. The cells $C_x(\Phi)$, $x \in \Phi$, form a random partition of $\mathbb{R}^d$ into Borel sets. By stationarity and (4.81) the statistical properties of $\{C_x(\Phi) + z : x \in \Phi\}$ are the same for all $z \in \mathbb{R}^d$. For simplicity (and in contrast to Last, 2006, 2010), the attention here is restricted to partitions that do only depend on Φ but not on any additional source of randomness. An example is the Voronoi tessellation generated by Φ; see Section 9.2.

Let π be a stationary partition based on Φ and let Φ^o be a Palm version of Φ. It is easy to see that again almost surely $\pi(\Phi^o, x) \in \Phi^o$ for all $x \in \mathbb{R}^d$. By Last (2006, Theorem 4.1),

$$\mathbf{E}\big(f(\Phi)g(\Phi_{-\pi(\Phi,o)})\big) = \lambda \mathbf{E}\left(g(\Phi^o)\int_{C_0(\Phi^o)} f(\Phi^o_{-x})\,\mathrm{d}x\right) \tag{4.83}$$

for all measurable nonnegative functions f, g on $\mathbb{N}$. This result is a simple consequence of (4.64). A special case will be proved in Proposition 9.1.

Letting $g \equiv 1$ in (4.83) generalises the inversion formula (4.74), and letting $f \equiv 1$ yields

$$\mathbf{E}\big(g(\Phi_{-\pi(\Phi,o)})\big) = \lambda \mathbf{E}\big(g(\Phi^o)v_d(C_0(\Phi^o))\big). \tag{4.84}$$

Applying this with $g(\varphi) = v_d\big(C_0(\varphi)\big)^\alpha$ for some $\alpha \geq 0$ gives

$$\mathbf{E}\big(v_d(V)^\alpha\big) = \lambda \mathbf{E}\big(v_d(C_0(\Phi^o))^{\alpha+1}\big), \tag{4.85}$$

where $V = C_{\pi(\Phi,o)}(\Phi) = \{x \in \mathbb{R}^d : \pi(\Phi, x) = \pi(\Phi, o)\}$ is the *zero cell* of π. In particular,

$$\mathbf{E}\big(v_d(C_0(\Phi^o))\big) = \lambda^{-1}, \tag{4.86}$$

cf. (9.15). In the notation of Chapter 9 it is $C_0(\Phi^o) = C_0$.

If $\mathbf{P}\big(0 < v_d(C_0(\Phi^o)) < \infty\big) = 1$, then (4.85) extends to all real α, cf. Last (2006).

A stationary partition π is called *balanced* if

$$\mathbf{P}\big(v_d(C_x(\Phi)) = \lambda^{-1} \text{ for all } x \in \Phi\big) = 1. \tag{4.87}$$

In the ergodic case such a partition can be used to construct the Palm version Φ^o from Φ by a random shift. By stationarity, the origin $o \in \mathbb{R}^d$ might be interpreted as a uniformly distributed random test point in $\mathbb{R}^d$. This point picks one of the cells $C_x(\Phi)$, $x \in \Phi$, namely $C_T(\Phi)$, where $T := \pi(\Phi, o)$. If π is balanced, then all cells have the same chance of being picked by the origin. Therefore T is the typical point of Φ. Ergodicity suggests that Φ_{-T} and Φ^o should have the same distribution. The following result of Holroyd and Peres (2005) shows in particular that this is indeed true.

Theorem 4.4. *A stationary partition π is balanced if and only if*

$$\mathbf{P}(\Phi_{-\pi(\Phi,o)} \in Y) = P_o(Y) \qquad \textit{for all } Y \in \mathcal{N}. \tag{4.88}$$

Proof. If π is balanced, then (4.88) is a direct consequence of (4.84). Assume, conversely, that (4.88) holds. Taking $g(\varphi) := v_d\big(C_0(\varphi)\big)$ in (4.84) yields

$$\mathbf{E}\big(v_d(C_0(\Phi^o))\big) = \lambda \mathbf{E}\big(v_d(C_0(\Phi^o))^2\big).$$

In view of (4.86) this means that $\mathbf{E}\big(v_d(C_0(\Phi^o))^2\big) = \lambda^{-2}$. Therefore the variance of $v_d\big(C_0(\Phi^o)\big)$ vanishes, implying that $\mathbf{P}\big(v_d(C_0(\Phi^o)) = \lambda^{-1}\big) = 1$. This is equivalent to (4.87). □

By the ergodic theorem for stationary point processes (see e.g. in Kallenberg, 2002, Corollary 10.19) a balanced partition can only exist if the *sample intensity* $\lim_{n\to\infty} \Phi(W_n)/v_d(W_n)$ equals λ almost surely, where $\{W_n\}$ is a convex averaging sequence as in Section 4.1.6. This is the case, for instance, if Φ is ergodic. (In the general case the definition of a balanced partition has to be modified.) It is a quite remarkable fact that this assumption is already enough to guarantee the existence of a balanced partition.

The following construction is due to Holroyd and Peres (2005). Place a small ball around each point of the point process Φ and expand the balls simultaneously at the same rate in all directions. Each expanding ball occupies space that has not yet been previously occupied by other balls; the expansion of a ball stops when the size of its occupied space is λ^{-1}. In this way $\mathbb{R}^d$ is partitioned into regions (cells) of size λ^{-1}, each containing one point of Φ. Now

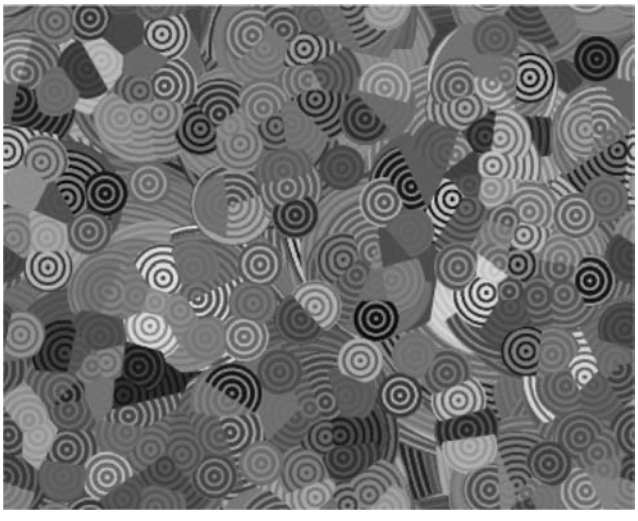

Figure 4.6 A balanced partition for a sample of a homogeneous Poisson process (with periodic boundary conditions). All cells have the same area, but not all cells are connected. The figure shows the generating points with the occupied areas. Concentric circles around the points are used to aid the identification of the cells. Courtesy of A. E. Holroyd.

map each location to the point of its cell. It should be noted that the cells of this partition need not be spherical or convex, they can be even disconnected. Figure 4.6 shows a balanced partition for a sample of a homogeneous Poisson process.

4.5 The second moment measure

Just as the variance is a fundamental parameter of the distribution of a random variable, the second moment measure is an important characteristic of a point process Φ. A simple description is provided in the stationary case by the Palm distribution. The second-order factorial moment measure $\alpha^{(2)}$ can be written as

$$\begin{aligned}\alpha^{(2)}(B_1 \times B_2) &= \lambda^2 \int_{B_1} \mathscr{K}(B_2 - x)\,\mathrm{d}x \\ &= \lambda^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{B_1}(x)\mathbf{1}_{B_2}(x+h)\,\mathscr{K}(\mathrm{d}h)\,\mathrm{d}x \qquad \text{for Borel sets } B_1 \text{ and } B_2,\end{aligned} \tag{4.89}$$

where $\mathscr{K}$ is a measure defined by

$$\begin{aligned}\lambda\mathscr{K}(B) &= \int_{\mathbb{N}} \varphi(B \setminus \{o\})\, P_o(\mathrm{d}\varphi) \\ &= \int \varphi(B)\, P_o^!(\mathrm{d}\varphi) \qquad \text{for Borel sets } B.\end{aligned} \tag{4.90}$$

So $\lambda\mathcal{K}(B)$ can be interpreted as the mean number of points in $B \setminus \{o\}$ under the condition that at o there is a point of Φ; $\mathcal{K}$ is called the *reduced second moment measure*. It possesses a symmetry property under reflection:

$$\mathcal{K}(B) = \mathcal{K}(\check{B}) \qquad \text{for Borel sets } B. \tag{4.91}$$

Formula (4.89) follows from the Campbell–Mecke theorem (4.64):

$$\begin{aligned}\alpha^{(2)}(B_1 \times B_2) &= \mathbf{E}\left(\sum_{x_1,x_2\in\Phi}^{\neq} \mathbf{1}_{B_1}(x_1)\mathbf{1}_{B_2}(x_2)\right)\\ &= \int_{\mathbb{N}} \sum_{x\in\varphi} \mathbf{1}_{B_1}(x)\varphi(B_2 \setminus \{x\})P(\mathrm{d}\varphi)\\ &= \lambda \int_{\mathbb{R}^d}\int_{\mathbb{N}} \mathbf{1}_{B_1}(x)\varphi\big((B_2 - x) \setminus \{o\}\big)P_o(\mathrm{d}\varphi)\,\mathrm{d}x.\end{aligned}$$

From (4.45) one can derive the following formula for the second moment measure:

$$\mu^{(2)}(B_1 \times B_2) = \lambda\nu_d(B_1 \cap B_2) + \lambda^2 \int\int \mathbf{1}_{B_1}(x)\mathbf{1}_{B_2}(x+h)\,\mathcal{K}(\mathrm{d}h)\,\mathrm{d}x. \tag{4.92}$$

It shows that the second moments given on $\mathcal{B}^d \times \mathcal{B}^d$ can be described in terms of intensity λ and measure $\mathcal{K}$.

Moreover, it holds the general formula for nonnegative measurable function f:

$$\mathbf{E}\left(\sum_{x_1,x_2\in\Phi}^{\neq} f(x_1,x_2)\right) = \lambda^2 \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x_1,x_2)\,\mathcal{K}(\mathrm{d}h)\,\mathrm{d}x. \tag{4.93}$$

According to Section 4.3 the second-order product density $\varrho^{(2)}$, if it exists, also describes $\mu^{(2)}$. So there is also a relationship between $\varrho^{(2)}$ and $\mathcal{K}$, namely

$$\lambda^2\mathcal{K}(B) = \int_B \varrho^{(2)}(x)\,\mathrm{d}x \qquad \text{for Borel sets } B. \tag{4.94}$$

In the planar case one can introduce polar coordinates and describe the reduced second moment measure by the function $K(r,\alpha)$,

$$K(r,\alpha) = \mathcal{K}\big(S(r,\alpha)\big) \qquad \text{for } r \geq 0 \text{ and } 0 \leq \alpha \leq \pi,$$

where $S(r,\alpha)$ is the sector of radius r, centred at the origin and given by the angle α.

For fixed r, the ratio

$$K(r,\alpha)/K(r,\pi)$$

is a distribution function for directions. This idea is useful for directional analysis of point processes; see Illian *et al.* (2008, Section 4.5). The present book limits itself to the case of isotropic point processes. Note that it makes sense to consider also 'local anisotropies', for example local tendencies to parallelism.

The description of the second moment measure simplifies in the motion-invariant case (when isotropy is added to stationarity). Then the reduced second moment measure can be

replaced by the *reduced second moment function* $K(r)$ which is a function that can be plotted. This function, often called *Ripley's K-function*, is defined by

$$K(r) = \mathscr{K}\big(B(o,r)\big) \qquad \text{for } r \geq 0. \tag{4.95}$$

The quantity $\lambda K(r)$ is the mean number of points of Φ within a ball of radius r centred at the typical point, which is not itself counted.

In the case of a homogeneous Poisson process

$$K(r) = b_d r^d \qquad \text{for } r \geq 0, \tag{4.96}$$

which follows from the Slivnyak–Mecke theorem as

$$\begin{aligned} \lambda K(r) &= \int \varphi\big(B(o,r)\big)\, P_o^!(\mathrm{d}\varphi) \\ &= \int \varphi\big(B(o,r)\big)\, P(\mathrm{d}\varphi) \\ &= \mathbf{E}\big(\Phi(B(o,r))\big) = \lambda \nu_d\big(B(o,r)\big) = \lambda b_d r^d. \end{aligned}$$

Asymptotically, for all stationary and isotropic point processes $K(r)$ behaves as $b_d r^d$.

Other functions than $K(r)$ are often used to describe the second-order behaviour of a point process. Which function is to be preferred depends mainly on convenience, but also on statistical considerations and traditions. Some functions originate from the physical literature in which they have been used for a long time, much longer than by statisticians. An example is the pair correlation function, for which an early reference is Ornstein and Zernike (1914).

Four examples of such functions are

product density $\varrho^{(2)}$:

$$\varrho^{(2)}(r) = \lambda^2\, \frac{\mathrm{d}K(r)}{\mathrm{d}r} \Big/ (db_d r^{d-1}) \qquad \text{for } r \geq 0, \tag{4.97}$$

pair correlation function g:

$$g(r) = \frac{\varrho^{(2)}(r)}{\lambda^2} \qquad \text{for } r \geq 0, \tag{4.98}$$

radial distribution function RDF:

$$RDF(r) = \lambda \frac{\mathrm{d}K(r)}{\mathrm{d}r} = \frac{db_d r^{d-1} \varrho^{(2)}(r)}{\lambda} \qquad \text{for } r \geq 0, \tag{4.99}$$

L-function:

$$L(r) = \left(\frac{K(r)}{b_d}\right)^{1/d} \qquad \text{for } r \geq 0. \tag{4.100}$$

In the Poisson process case all these functions satisfy simple formulae:

$$\varrho^{(2)}(r) = \lambda^2, \tag{4.101}$$

$$g(r) \equiv 1, \tag{4.102}$$

$$RDF(r) = db_d r^{d-1}\lambda, \tag{4.103}$$

$$L(r) = r. \tag{4.104}$$

For other models the forms of these functions correspond to various properties of the underlying point process. Maxima of $g(r)$, or values of $K(r)$ larger than $b_d r^d$ for r in specific intervals, indicate frequent occurrences of inter-point distances at such values; likewise, minima of $g(r)$ or low values of $K(r)$ indicate inhibition at these distances. Either way, some form of inner order or clustering in the point pattern may be responsible. (For this purpose, $g(r)$ is the more suitable tool, while the cumulative nature of $K(r)$ or $L(r)$ makes the interpretation sometimes difficult.) Model identification may be suggested by comparison of empirical pair correlation functions, or reduced second moment functions, with their theoretical versions. Figure 4.7 exhibits pair correlation functions for four different spatial point process models.

The possible forms of pair correlation functions and their interpretation are discussed in of Illian *et al.* (2008, Section 4.3.4).

The K-function (as well as $g(r)$) does *not* represent all distributional information about a stationary isotropic point process. Baddeley and Silverman (1984) gave an example of a planar point process which is quite different from a Poisson process and yet has the same K-function:

$$K(r) = \pi r^2.$$

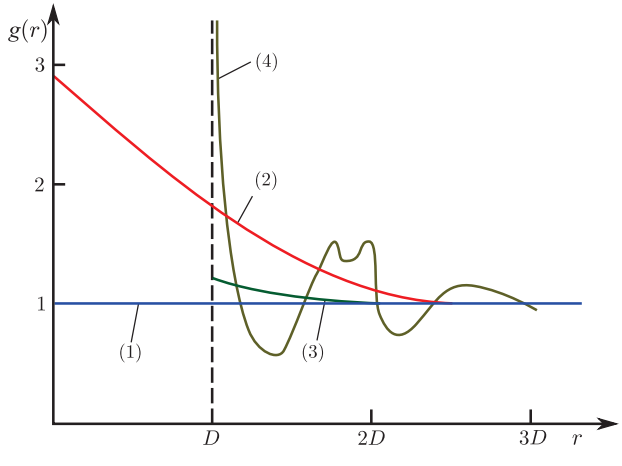

Figure 4.7 Pair correlation functions g for various point process models: (1) Poisson process; (2) Matérn cluster process with parameters $R = 2.5$ and $\lambda = 0.027$, $\mu = 5$ (the values of $g(r)$ greater than 1 for small r result from clustering); (3) Matérn hard-core process with parameters $D = 2$ and $p_h \simeq 0.03$ (the hard-core distance is 2); (4) random close packing of balls of diameter D. The peaks occur in a typical pattern at D, at around $\sqrt{3}D$ and below $2D$.

It is simple to give also examples of pairs of other processes with equal $K(r)$. For example, Stoyan (1991) gave an example, where one of the processes is even anisotropic; there $K(r)$ is defined as $\mathscr{K}\big(B(o,r)\big)$.

If $\mathscr{K}$, K or g is given, then (at least in principle) all variances and covariances of point-counts can be calculated. In particular, if B is a bounded Borel set, then

$$\mathbf{var}\big(\Phi(B)\big) = \lambda^2 \int \gamma_B(h)\,\mathscr{K}(\mathrm{d}h) + \lambda\nu_d(B) - \big(\lambda\nu_d(B)\big)^2, \tag{4.105}$$

and in the isotropic case

$$\mathbf{var}\big(\Phi(B)\big) = \lambda^2 \int_0^\infty \overline{\gamma}_B(r)\,\mathrm{d}K(r) + \lambda\nu_d(B) - \big(\lambda\nu_d(B)\big)^2, \tag{4.106}$$

where $\gamma_B(h)$ is the set covariance of B.

If B is large ('spherically infinite' in the terminology of Girling, 1982), then

$$\frac{\mathbf{var}\big(\Phi(B)\big)}{\nu_2(B)} \simeq \lambda + 2\pi\lambda^2 \int_0^\infty \big(g(r)-1\big) r\,\mathrm{d}r \qquad (d=2),$$

and

$$\frac{\mathbf{var}\big(\Phi(B)\big)}{\nu_3(B)} \simeq \lambda + 4\pi\lambda^2 \int_0^\infty \big(g(r)-1\big) r^2\,\mathrm{d}r \qquad (d=3).$$

For point processes with a tendency to regularity these formulae tend to produce approximations somewhat too large, while approximations for cluster processes tend to be too small.

In the case of marked point processes there exist similar characteristics; see Illian *et al.* (2008, Section 5.3). They are of a different nature in the cases of discrete and continuous marks.

4.6 Summary characteristics

It would not be surprising if one feels that the distribution P of a point process is a rather complex mathematical thing and much too complicated for intuitive understanding or for visualisation. Therefore various summary characteristics have been proposed which describe certain aspects of P. These are typically real numbers or functions, usually of distance r as they are based on inter-point distances.

The following tries to give the reader some orientation for the case of stationary point processes. For the case of nonstationary and of marked point processes the reader is referred to Illian *et al.* (2008).

It is recommended that one uses some summary characteristics in parallel since each of them gives only specific information. A single summary characteristic can hardly completely describe all interesting aspects of a point process distribution.

The correct interpretation of summary characteristics is a question of experience and somewhat of an art.

Intensity λ

The intensity gives very important global information about the point process. It is another term for 'point density'.

Of course, λ alone is of little value, similarly as the mean for a random variable in classical statistics. And λ influences the other summary characteristics, though some of them are constructed with the aim of eliminating the influence of point density.

Second-order characteristics

Second-order characteristics are for many scientists the favourite summary characteristics, in particular for physicists, who have used the pair correlation function since the beginning of the 20^{th} century and very rarely consider other second-order characteristics and non-second-order summary characteristics.

Second-order characteristics give information on many scales of distances, describe so-to-say the average behaviour of the point process of interest.

The K-function is only rarely used in practical statistics and model visualisation. The L-function is more popular because of its simpler graphical form and the better properties in the statistical context. It is useful in statistical tests. However, on the other hand its interpretation is a bit difficult because of its cumulative character. Frequent inter-point distances in a distance interval (r_1, r_2) lead to large values of $L(r)$ not only in the interval (r_1, r_2), but also for r-values larger than r_2.

The pair correlation function $g(r)$ does not have the 'cumulative' disadvantage just mentioned. Hence it is ideal for the interpretation and understanding of point process distribution; but it would be wrong to assume that it contains more information than $K(r)$ or $L(r)$. However, its statistical estimation is a bit difficult since a bandwidth must be carefully chosen. (This is similar to the case of density function estimation in classical statistics.) Moreover, the interpretation needs experience or some support; see Illian *et al.* (2008).

All three characteristics, $K(r)$, $L(r)$ and $g(r)$, are normalised by division of λ^2. Nevertheless, it should not be said that they are independent of intensity λ. Still intensity influences them, for example via the hard-core distance, which may be closely related to λ.

Nearest-neighbour distance distribution function $D(r)$

The nearest-neighbour distance distribution function $D(r)$ is a rather natural characteristic. Frequently it is used spontaneously as the first functional summary characteristic. It is easy to understand, and therefore is used in Section 4.4 to explain Palm characteristics.

For lattice-like patterns $D(r)$ yields the lattice-distance distribution. For processes with clusters it gives information on the distances of the points within clusters, but hardly on distances between clusters. The distribution function is said to be 'short-sighted', since it gives information only about the nearest neighbour; what happens further away does not play a rôle. This is different for second-order characteristics.

Spherical contact distribution function $H_s(r)$

Also the spherical contact distribution function is quite natural. It goes also under the name 'empty space function', with the colourless symbol $F(r)$. Now the distance to the nearest neighbour from a random test point is considered. So it is clear that $H_s(r)$ is also short-sighted. In contrast to $D(r)$ it is valuable for describing the extent of empty space between clusters, but of little value for describing the situation within clusters.

4.7 Introduction to statistics for stationary spatial point processes

4.7.1 General remarks

The theory of statistics for point processes comprises a part of spatial statistics as described by Ripley (1981), Cressie (1993) and Gelfand *et al.* (2010). This section presents some ideas of exploratory analysis for stationary spatial point processes; model-based statistics will play some rôle in Chapter 5. All is limited to the case of stationary point processes. The basic ideas of point process statistics apply also to other structures of stochastic geometry, in particular for line and fibre processes. The general aim of this section is to describe methods to obtain unbiased estimators or, at least, *ratio-unbiased* estimators (quotients where numerator and denominator are unbiased) of important summary characteristics. It is always assumed the Cartesian coordinates of all points are given. For other data situations see Illian *et al.* (2008).

Point process statistics in this book are considered for the planar $d = 2$ and spatial $d = 3$ case. The statistics of point processes on the real line ($d = 1$, often described as series of events) have a relatively large literature; see Cox and Lewis (1966), Brillinger (1975, 1978), Snyder (1975), Karr (1991) and Andersen *et al.* (1993). The special nature of the real line makes this case more amenable.

Books providing more details for $d \geq 2$ are Diggle (1983, 2003), Ripley (1981, 1988), Stoyan and Stoyan (1994), Martínez and Saar (2002), Møller and Waagepetersen (2004), Illian *et al.* (2008) and Gelfand *et al.* (2010).

Often statistical analysis of a stationary point process depends on observation of *one sample* only, and that via a bounded sampling window W. Patterns arising in astronomy, ecology, geography and geology are often truly unique samples of a stochastic phenomenon. In other cases data collection is so complicated that only one sample is collected. Typically, in such cases it is assumed that the observed patterns are samples of *stationary ergodic* point processes, an assumption not susceptible to statistical analysis if there is only one sample, but one that is necessary if any statistical analysis is sensible. Therefore ergodicity is also assumed throughout this section. In practice, either it is plausible from the very nature of the data, or else one must proceed on an *ad hoc* basis as if the assumption were true, and subject one's conclusions to the proviso that while they may be of possible value in the nonstationary case, they will not then have the same interpretation. For example, an empirical point density (the mean number of points per unit area) can still be calculated even if the observed point pattern is nonstationary. It has value as a description of the spatial average behaviour of the pattern but does not possess all the properties of an estimator of a stationary point process intensity. See the interesting discussion in Matheron (1989), who justified this approach.

4.7.2 Edge-corrections

An ever present problem of spatial statistics is that of *edge-effects*; Ripley (1982) and Illian *et al.* (2008) discuss these in detail. The problem intervenes in point process statistics in the estimation of g, $\mathscr{K}$, K, H_s and D. As an illustration, consider the problem of estimating the nearest-neighbour distribution function $D(r)$. Two naïve methods suggest themselves: plus-sampling and minus-sampling named according as to whether they require more, or less, information than is given through the window W. For rectangular W the method of periodic boundary conditions is also feasible.

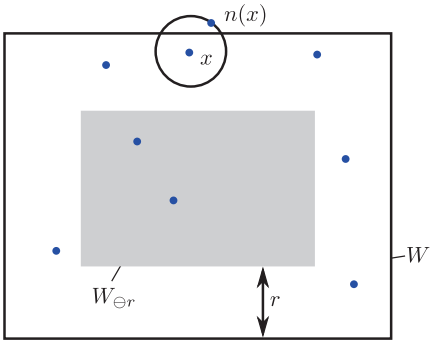

Figure 4.8 The window W and the eroded window $W_{\ominus r}$ for minus-sampling of $D(r)$. The shaded area is the eroded window. The nearest neighbour $n(x)$ of the point x is not in the window and so cannot be used for estimating of $D(r)$ by minus-sampling. Ignoring this edge problem will introduce bias.

Plus-sampling

In this method $D(r)$ is estimated by the empirical distribution function of the sample of all distances $\|x - n(x)\|$, where x runs through the points of the pattern in the whole window W, and $n(x)$ is the nearest neighbour of x, *no matter whether this neighbour is in the sampling window W or not*. If x is near the boundary of W, then its true nearest neighbour may lie outside W. Therefore, plus-sampling may require more information than is contained in the window W. When the point pattern can only be observed within W, this means that for some x it may not be possible to determine $n(x)$. Figure 4.8 gives an illustration. If one simply ignores this edge problem and uses the nearest neighbour to x *lying within W*, then bias is introduced; plus-sampling itself is unbiased.

Tscheschel and Chiu (2008) introduced the so-called *quasi-plus-sampling*. The idea is to reconstruct the point pattern (see Section 6.7) in a larger region containing W and then apply plus-sampling as if the reconstructed pattern were observed.

Minus-sampling

In estimating $D(r)$, one might use only the points x in $W_{\ominus r}$ in creating the sample of distances $\|x - n(x)\|$; see Figure 4.8. Clearly, for these points the nearest neighbours, if within distance r, lie in W. Thus, for each x in the eroded window, one can correctly determine whether its nearest-neighbour distance is larger than r or not. This 'border method' will avoid bias but loses much information for large r.

Periodic edge-correction

If the window W is a rectangle or a parallelepiped, the given point patterns can be continued outside of W, as shown in Figure 4.9 for the planar case. The resulting point pattern can then be analyzed by means of plus-sampling. To do this, the distances have to be redefined. In the planar case this is termed as 'torus metric': if W is the rectangle with side lengths a and b with

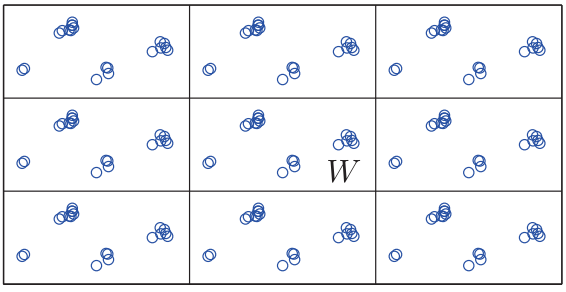

Figure 4.9 A point pattern in a rectangular window W and its periodic continuation.

left lower vertex at the origin, the distance between two points x and $y \in W$ with $x = (\xi_1, \xi_2)$ and $y = (\eta_1, \eta_2)$ is

$$\|x - y\| = \sqrt{(\min\{|\xi_1 - \eta_1|, a - |\xi_1 - \eta_1|\})^2 + (\min\{|\xi_2 - \eta_2|, b - |\xi_2 - \eta_2|\})^2}. \quad (4.107)$$

By this method every point in W gets a natural nearest neighbour, but in some cases they are incorrect. It is difficult to determine the errors of this method; strange and spurious point configurations may appear along the borders, and the method is merely a cheap trick to provide more points.

Clearly, this method is absolutely suitable when the pattern in W is itself a result of a simulation under 'periodic boundary conditions'.

These three edge-correction methods are applied in many cases and also outside of point process statistics. However, for particular summary characteristics there exist adapted, better edge-correction methods, which will be demonstrated in the following.

4.7.3 Estimation of the intensity λ

The natural and best estimator of the intensity λ is $\hat{\lambda}$ defined by

$$\hat{\lambda} = \frac{\Phi(W)}{\nu_d(W)}. \quad (4.108)$$

It uses the idea that the intensity is the mean number of points per volume unit. Straightforward calculation shows that it is unbiased.

If Φ is ergodic, then $\hat{\lambda}$ is strongly consistent, in the sense that $\hat{\lambda} \to \lambda$ almost surely as the window size increases as a convex averaging sequence.

The variance of $\hat{\lambda}$ depends both on the window's shape and size and on the second moment measure of the point process. In the isotropic case (4.106) implies

$$\mathbf{var}(\hat{\lambda}) = \left(db_d \lambda^2 \int_0^\infty r^{d-1} \overline{\gamma}_W(r) g(r)\, \mathrm{d}r + \lambda \nu_d(W) - \left(\lambda \nu_d(W)\right)^2 \right) \Big/ \left(\nu_d(W)\right)^2, \quad (4.109)$$

where $\overline{\gamma}_W(r)$ is the isotropised set covariance of the convex compact window W.

4.7.4 Estimation of the reduced second moment measure

Estimation of $\mathscr{K}(B)$ — anisotropic case

In the general anisotropic case $\lambda^2\mathscr{K}(B)$ can be estimated by $\kappa_{\rm st}(B)$, where

$$\kappa_{\rm st}(B) = \sum_{x,\, y\in\Phi\cap W}^{\neq} \frac{\mathbf{1}_B(y-x)}{\nu_d(W_x\cap W_y)} \tag{4.110}$$

defined for bounded Borel sets B such that $\nu_d(W\cap W_z)$ is positive for all z in B.

Proof of the unbiasedness of $\kappa_{\rm st}(B)$ (Ohser and Stoyan, 1981). Set $f(x_1,x_2)=\mathbf{1}_B(x_2-x_1)\mathbf{1}_W(x_1)\mathbf{1}_W(x_2)/\nu_2(W_{x_1}\cap W_{x_2})$ and apply (4.43) and (4.89):

$$\begin{aligned}\mathbf{E}\big(\kappa_{\rm st}(B)\big) &= \int f(x_1,x_2)\,\alpha^{(2)}\big(\mathrm{d}(x_1,x_2)\big)\\ &= \lambda^2\int\int f(x,x+h)\,\mathrm{d}x\,\mathscr{K}(\mathrm{d}h)\\ &= \lambda^2\int \mathbf{1}_B(h)\,\mathscr{K}(\mathrm{d}h) = \lambda^2\mathscr{K}(B).\end{aligned}$$

□

Estimation of $\mathscr{K}(B)$ and $K(r)$ — isotropic case

An unbiased estimator of $\lambda^2K(r)$ is $\kappa_{\rm iso}(r)$, defined by

$$\kappa_{\rm iso}(r) = \sum_{x,\, y\in\Phi\cap W} \frac{\mathbf{1}(0<\|x-y\|\le r)k(x,y)}{\nu_d(W^{(\|x-y\|)})} \qquad \text{for } 0\le r<r^*, \tag{4.111}$$

where

$$r^* = \sup\{r:\nu_d(W^{(r)})>0\} \qquad \text{and} \qquad W^{(r)}=\{x\in W:\partial(B(x,r))\cap W\neq\emptyset\},$$

and $k(x,y)$ is the inverse of the fraction of surface area in W of the sphere of radius $\|x-y\|$ centred at x. In the planar case it is $k(x,y)=2\pi/\alpha_{xy}$, where α_{xy} is the sum of all angles of the arcs in W of a circle centred at x with radius $\|x-y\|$. Figure 4.10 illustrates the definition of α_{xy}. If $\alpha_{xy}=0$, then $k(x,y)=0$.

The estimator $\kappa_{\rm iso}(r)$ was originally suggested by Ripley (1976); Formula (4.111) uses the modification due to Ohser (1983). The modification extends estimation to the case of large r,

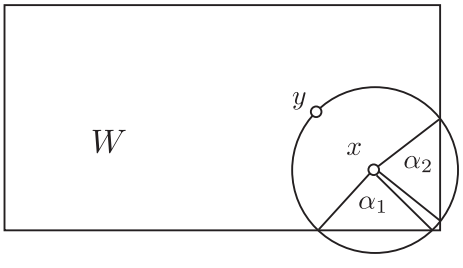

Figure 4.10 Definition of α_{xy} for estimation of $\kappa_{\rm iso}(r)$. In the example $\alpha_{xy}=2\pi-\alpha_1-\alpha_2$.

while Ripley's original form was unbiased only for

$$r \le r_0 = \inf\{t : \nu_2(W^{(t)}) < \nu_2(W)\}.$$

These and other estimators are discussed in Illian *et al.* (2008). As it seems it makes sense to use always $\kappa_{\rm st}(B)$ with $B = B(o, r)$, since in cases where the isotropy assumption is invalid, $\kappa_{\rm iso}(r)$ may behave badly.

Both estimators have edge-correction built in. For example, the naïve estimator corresponding to (4.111) is

$$\hat{\kappa}(r) = \frac{\text{number of point pairs in } W \text{ separated by distance less than } r}{\nu_d(W)} \tag{4.112}$$

and for small r (such that $W = W^{(r)}$) the edge-bias is corrected by a weighting factor $k(x, y)$,

$$\kappa_{\rm iso}(r) = \frac{1}{\nu_2(W)} \sum_{x,y\in\Phi\cap W} \mathbf{1}(0 < \|x - y\| \le r)k(x, y).$$

This is the form originally suggested by Ripley (1976). The principle is to correct the bias, by setting the weight $k(x, y)$ greater than 1 whenever one or both of the points x and y are near the boundary of W. Deep in W (when $B(x, \|x - y\|)$ and $B(y, \|x - y\|)$ lie completely in W) the weight is 1. The isotropy assumption is used to estimate the mean number of pairs missed out by the window at a particular location, and thus to calculate the weight. The full form of (4.111) makes a further correction which is necessary for large r.

Stoyan and Stoyan (2000) found that for estimating $K(r)$ the naïve approach of dividing estimators for $\lambda^2 K(r)$ by the squared $\hat{\lambda}$ of (4.108) can be improved by adapted intensity estimators which depend on r; see also Illian *et al.* (2008, p. 231).

Estimation of other second-order characteristics

For statistical purposes it is useful to stabilise variances of $\hat{K}(r)$. This can be done by using the L-function, estimating $L(r)$ by

$$\hat{L}(r) = \left(\frac{\hat{K}(r)}{b_d}\right)^{1/d}, \tag{4.113}$$

where $\hat{K}(r)$ is either $\kappa_{\rm st}(r)/\hat{\lambda}^2$ or $\kappa_{\rm iso}(r)/\hat{\lambda}^2$. Examples 2.1 and 2.3 illustrate the use of $L(r)$.

The product density $\varrho^{(2)}(r)$ and the pair correlation function $g(r)$ can be estimated by using an edge-corrected kernel estimator such as

$$\hat{\varrho}^{(2)}(r) = \sum_{x,y\in\Phi\cap W} \frac{\mathbf{k}(\|x - y\| - r)}{db_d r^{d-1}\nu_d(W_x \cap W_y)}, \tag{4.114}$$

where $\mathbf{k}$ is a kernel function; see Illian *et al.* (2008, pp. 230–1). An estimator of $g(r)$ is then

$$\hat{g}(r) = \hat{\varrho}^{(2)}(r)/\hat{\lambda^2}. \tag{4.115}$$

Edge-corrections for the estimation of mark correlation functions are analogous and discussed in Illian *et al.* (2008, Section 5.3.4).

4.7.5 Estimation of the spherical contact distribution and of the probability generating functional

The methods of estimating $H_s(r)$ for random closed sets (see Section 6.4.5) can be also used in the particular case of point processes. Since

$$H_s(r) = \mathbf{P}(o \in \Phi_{\oplus r}) \qquad \text{for } r \geq 0, \tag{4.116}$$

$H_s(r)$ is equal to the area fraction of the random closed set

$$\Xi = \Phi_{\oplus r} = \bigcup_{x \in \Phi} B(x, r). \tag{4.117}$$

An unbiased estimator is given by

$$\hat{H}_s(r) = \frac{v_d\left(W_{\ominus r} \cap \bigcup_{x \in \Phi} B(x, r)\right)}{v_d(W_{\ominus r})} \qquad \text{for } 0 \leq r \leq \frac{\operatorname{diam}(W)}{2}. \tag{4.118}$$

This is edge-correction by minus-sampling; see Figure 4.11.

Note that $\hat{H}_s(r)$ may be neither continuous nor monotonic, while the distribution function $H_s(r)$ always has a density function. Section 6.4.5 will present more sophisticated estimators which are free of these disadvantages. Their construction follows an idea of Hanisch, to be explained in the next section.

Estimation of the probability generating functional G follows a similar route. For example, take v in $\mathbf{V}$ with $v(x) = 1$ if x is not in $B(o, r)$. (In this case $G(v)$ is the same $1 - H_s(r)$, only in another expression.) Then

$$\hat{G}(v) = \frac{1}{n} \sum_{i=1}^{n} \prod_{x \in \Phi \cap W} v(x - z_i) \tag{4.119}$$

is an unbiased estimator of $G(v)$ if the points $z_1, \ldots, z_n$ form a grid in $W_{\ominus r}$. Again this estimator is of the minus-sampling type.

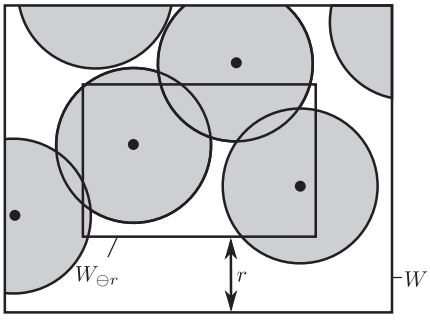

Figure 4.11 Edge-effects on the estimation of $H_s(r)$. Points outside of W contribute to $v_2(W \cap \Phi_{\oplus r})$. If their positions are not known, then only points falling in the eroded window $W_{\ominus r}$ should be used for unbiased estimation.

4.7.6 Estimation of the nearest-neighbour distance distribution function

Recall that the nearest-neighbour distribution function is formally defined for a stationary point process as

$$D(r) = 1 - P_o^!\big(\{\varphi \in \mathbb{N} : \varphi(B(o,r)) = 0\}\big) \qquad \text{for } r \geq 0. \tag{4.120}$$

An equivalent formulation uses the marked point process obtained by marking each point x of the process Φ with the distance $d(x)$ from x to its nearest neighbour $n(x)$. The resulting marked point process Ψ inherits the stationarity property from Φ. As can be seen from the mark distribution approach to Palm distributions on p. 129, the mark distribution function for Ψ is precisely $D(r)$. This formulation via Ψ clarifies the logic behind the estimators of $D(r)$. The following demonstrates two ways to handle the problems of edge-effects mentioned in Section 4.7.2.

The first estimator is in the spirit of minus-sampling. The so-called border estimator $\hat{D}_{\mathrm{b}}(r)$ was introduced by Ripley (1977). It is defined as

$$\hat{D}_{\mathrm{b}}(r) = \sum_{[x;\, d(x)]} \frac{\mathbf{1}_{W_{\ominus r}}(x)\mathbf{1}(0 < d(x) \leq r)}{\Phi(W_{\ominus r})} \qquad \text{for } r \geq 0. \tag{4.121}$$

The idea behind this estimator is quite simple: when estimating $D(r)$, only those points x in the window W are considered which have a distance larger than r from the window's boundary. For these points it is thus clear that there are no points closer than r outside the window and whether or not the nearest-neighbour distance $d(x)$ is not larger than r can be determined within W. These points lie in $W_{\ominus r}$. Since $\Phi(W_{\ominus r})$ is the number of those points, $\hat{D}_{\mathrm{b}}(r)$ is a natural ratio estimator for $D(r)$. Because numerator and denominator are random, the estimator is not unbiased, only ratio-unbiased. However, if Φ is ergodic, $\hat{D}_{\mathrm{b}}(r)$ is asymptotically unbiased.

Unfortunately, the estimator $\hat{D}_{\mathrm{b}}(r)$ has some disadvantages. It is not necessarily monotonically increasing in r, it can exceed 1 in value and it excludes many points from the estimation procedure.

For better understanding of the next estimator, rewrite $\hat{D}_{\mathrm{b}}(r)$ as follows. The quantity

$$\hat{\mathcal{D}}_{\mathrm{b}}(r) = \sum_{[x;\, d(x)]} \frac{\mathbf{1}_{W_{\ominus r}}(x)\mathbf{1}(0 < d(x) \leq r)}{\nu_d(W_{\ominus r})} \qquad \text{for } r \geq 0 \tag{4.122}$$

is clearly an unbiased estimator of $\lambda D(r)$, which can easily be shown by the Campbell theorem. Division by the adapted intensity estimator

$$\hat{\lambda}(r) = \Phi(W_{\ominus r})/\nu_d(W_{\ominus r}), \tag{4.123}$$

yields $\hat{D}_{\mathrm{b}}(r)$. The use of $\hat{\lambda}(r)$ reduces the estimation variance (while the use of the classical $\hat{\lambda}$ results in intolerable biases and large mean squared errors), but it leads to the non-monotonicity mentioned above.

Hanisch (1984c) suggested an unbiased estimator which outperforms the border estimator. It uses the so-called nearest-neighbour edge-correction (see Illian *et al.*, 2008, p. 187), which leads to the estimator

$$\hat{D}_{\mathrm{nn}}(r) = \hat{\mathcal{D}}_{\mathrm{nn}}(r) \big/ \hat{\lambda}_{\mathrm{nn}}(r) \qquad \text{for } 0 \leq r \leq R, \tag{4.124}$$

where

$$\hat{\mathcal{D}}_{\mathrm{nn}}(r) = \sum_{[x;\, d(x)]} \frac{\mathbf{1}_{W_{\ominus d(x)}}(x)\mathbf{1}(0 < d(x) \le r)}{\nu_d(W_{\ominus d(x)})} \tag{4.125}$$

with

$$\hat{\lambda}_{\mathrm{nn}} = \sum_{[x;d(x)]} \frac{\mathbf{1}_{W_{\ominus d(x)}}(x)}{\nu_d(W_{\ominus d(x)})} \tag{4.126}$$

and

$$R = \sup\{r > 0 : \nu_d(W_{\ominus r}) > 0\}. \tag{4.127}$$

The indicator $\mathbf{1}_{W_{\ominus d(x)}}(x)$ can be rewritten as $\mathbf{1}\big(d(x) < e(x)\big)$, where $e(x)$ is the distance of x to the boundary of W.

The principle underlying $\hat{D}_{\mathrm{nn}}(r)$ is simple: to use precisely those points x for which it is known that their nearest neighbours $n(x)$ are both within W and closer than r to x.

4.7.7 Estimation of Palm characteristics and mark distributions

The estimation procedures used for $K(r)$ and $D(r)$ above can be considered as prototypes for estimation of quantities related to Palm or mark distributions. The estimators given for these cases are complicated because of necessary edge-correction. If no edge-correction is necessary, all is much simpler.

Assume that the marks m of the points are given together with the points, as for example in the case where the points are tree positions and the marks their stem diameters. Then the mark distribution M can be estimated by

$$\widehat{\lambda M}(L) = \sum_{[x;\, m]\in\Psi} \frac{\mathbf{1}_W(x)\mathbf{1}_L(m)}{\nu_d(W)} \qquad \text{for } L \in \mathcal{M}, \tag{4.128}$$

and

$$\hat{M}(L) = \frac{\widehat{\lambda M}(L)}{\hat{\lambda}}. \tag{4.129}$$

For nonnegative marks, the mark distribution function $F_M(t)$ can be estimated by

$$\hat{F}_M(t) = \sum_{[x;\, m]\in\Psi} \frac{\mathbf{1}_W(x)\mathbf{1}(0 \le m \le t)}{\Phi(W)} \qquad \text{for } t \ge 0. \tag{4.130}$$

If the marks are 'constructed' marks (such as $d(x)$) that have to be determined from the given configuration of points, then edge-corrections are often necessary, namely when information from outside the window is required to construct the marks of some points, as in the cases of $D(r)$ and $K(r)$ considered above. If such information is indeed accessible, then possibly the *plus-sampling method* will be useful. Consider for example estimation of

$$\lambda P_f = \lambda E_o^!\big(f(\Psi)\big), \tag{4.131}$$

where f is a nonnegative measurable function on $\mathbb{N}$,

$$P_f = \int f(\varphi) P_o^!(\mathrm{d}\varphi). \tag{4.132}$$

Then

$$\widehat{\lambda P_f} = \sum_{x \in \Phi} \frac{\mathbf{1}_W(x) f(\Phi_{-x} \setminus \{o\})}{\nu_d(W)} \tag{4.133}$$

is unbiased for λP_f. It is natural to estimate P_f by $\widehat{\lambda P_f}/\hat{\lambda}$, but this estimator is only ratio-unbiased.

On the other hand, if only the information in W is available, then a universal method is *minus-sampling*. For example, suppose f in (4.131) is such that

$$f(\varphi) = f(\varphi \cap B)$$

for some set B (often but not necessarily $B = B(o, r)$ for some r), so that one only needs the information local to B. Then an unbiased estimator of λP_f is

$$\widehat{\lambda P_f} = \sum_{[x;m] \in \Psi} \frac{\mathbf{1}_{W \ominus \check{B}}(x) f(\Phi_{-x} \setminus \{o\})}{\nu_d(W \ominus \check{B})}. \tag{4.134}$$

Calculation of $\nu_2(W \ominus \check{B})$ is described in Matheron (1978) and Weil (1982b); see also Saxl (1989, p. 157).

An analogous procedure for estimation of $M(L)$ can be carried out if the mark m of point x can be determined when all points of Φ in B_x are known.

4.7.8 Parameter estimation

Maximum likelihood method

Maximum likelihood methods are widely used in classical statistics, and many statisticians believe that they should also be preferred in point process statistics. Readers familiar with estimation methods in classical statistics know that maximum likelihood estimation techniques can only be applied if the likelihood function – describing the probability (or probability density) of observing the data given the model – is known. This likelihood is maximised (with fixed data and variable parameters), yielding parameter estimators that best fit the data. However, often and particularly for stationary point processes, it is extremely difficult to find the likelihood function. Therefore, the maximum likelihood method can only be applied to specific classes of models. These are Poisson processes, Cox processes and finite Gibbs processes; see Diggle (2003), Møller and Waagepetersen (2004) and Gelfand *et al.* (2010).

It is possible to (heuristically) approximate likelihood functions when they are not known explicitly. One of these methods is the pseudo-likelihood method introduced by Besag (1975, 1978); see Møller and Waagepetersen (2004) and Gelfand *et al.* (2010). Another method, introduced by Tanaka *et al.* (2008), uses the pair correlation function (for which a formula must be known) and the constructed point process of difference points

$$\delta = x - y \qquad \text{for } x \neq y, \tag{4.135}$$

where x and y are points of the observed pattern (see also Illian *et al.*, 2008, pp. 449–50).

Method of moments

The method of moments has many applications in the context of spatial point processes, especially when the likelihood function is not available.

Note that the term 'method of moments' is used here somewhat loosely. The approaches described below are all based on the same general idea, which is applied to moments or moment-measure-related characteristics as well as to other summary characteristics that are not moment-related. The general idea is to find parameters that minimise the difference between a 'suitable' summary characteristic S that is known analytically (or from simulations) and its unbiased estimator $\hat{S}$ obtained from the data. It is important that S depends on the unknown parameter θ; to emphasise this dependence, the characteristic is denoted by S_θ. The value θ for which S_θ and $\hat{S}$ are 'as similar as possible' is used as an estimator. The term 'as similar as possible' means here 'similar in the sense of a specific approximation method', such as the least-squares approach.

As indicated by the vague expression 'suitable', different summary characteristics S may be used. Which of these is deemed suitable depends on the context. A first criterion for the choice of summary characteristics is often whether a formula for S_θ is known. However, if this is not the case, then simulation approaches may be used instead. Another criterion should be that S_θ is sensitive to variation in θ.

In many applications the method of moments is based on a functional summary characteristic, that is, the S_θ above is a function $S_\theta(r)$. Sometimes it is sufficient to simply plot $\hat{S}(r)$ and to identify specific points, for example cusp points or points where $S_\theta(r)$ becomes constant. This approach may be used to find estimators of particular distances such as the range of correlation. In the context of Neyman–Scott processes the radius of the clusters may be found in this way.

Typically, however, the method of moments for functional summary characteristics applies a least squares approach, which is a special case of the *minimum contrast method*. This is based on the simple idea of minimising

$$\Delta(\theta) = \int_{s_1}^{s_2} |\hat{S}(r) - S_\theta(r)|^\beta \mathrm{d}r \tag{4.136}$$

with respect to θ. The value of θ that minimises $\Delta(\theta)$ is the minimum contrast estimator $\hat{\theta}$.

Here, the values of s_1, s_2 and β as well as the summary characteristic $S_\theta(r)$ itself can in principle be chosen arbitrarily; often $\beta = 2$ (i.e., least squares) is used. In the literature most applications use either

$$S_\theta(r) = g(r), \qquad S_\theta(r) = H_s(r), \qquad \text{or} \qquad S_\theta(r) = D(r).$$

For more details see Illian *et al.* (2008) and Gelfand *et al.* (2010). Jolivet (1986) and Heinrich (1992a, 1993) investigated the statistical properties of $\hat{\theta}$ for special cases. In reasonable cases the minimum contrast estimator is strongly consistent, that is, it converges almost surely to the true value as the size of the window W increases, as a convex averaging sequence, to the whole space.

4.7.9 Representative windows, representative volume elements

An important question in spatial statistics in general is the issue of choosing an appropriate window size that ensures pre-defined statistical precision requirements. In this context the term *representative volume element* (RVE) is used.

Of course the RVE depends both on the variability of the point process of interest and on the summary characteristic considered. If the same summary statistic is considered, larger RVEs are necessary when cluster point processes are analysed as opposed to regular processes, where smaller RVEs are sufficient. On the other hand, for a precise estimation of the pair correlation function a larger RVE is necessary than for intensity estimation.

In classical statistics, sample size calculations require some prior knowledge on the nature of the data that will be analysed, such as their variation. In the context of spatial point processes this is similar. It is impossible to determine the RVE without any *a priori* knowledge of the distribution of the point process investigated. A straightforward approach to acquiring *a prior* knowledge is a *pilot study* consisting of a preliminary statistical analysis of a small window or a small number of windows if a series of windows has to be analysed. The expectation is that the pilot study yields a useful yet rough estimate of the intensity λ and fundamental information on the point processes type, that is, whether it is a regular or a clustered pattern. Based on this, estimates of RVEs can be obtained. This is sketched in Illian *et al.* (2008, Section 4.8.2) for estimation of intensity λ and pair correlation function $g(r)$. For the other summary characteristics similar methods can be used; see also Section 6.4.6.

4.7.10 Hypotheses testing

Statistical tests are considered as an important part of point processes statistics. In most cases these are goodness-of-fit tests, typically for proving that a model developed fits the data considered. As Baddeley (2010, p. 361) says: 'A general weakness of goodness-of-fit tests is that the alternative hypothesis is very broad (embracing all point processes other than the model specified in the null hypothesis), so that rejection of the null hypothesis is rather uninformative, and acceptance of the null hypothesis is unconvincing because of weak power against specific alternatives.'

Nevertheless, such tests are standard. It is typical for point process statistics that the distributions of the test characteristics are too complicated to derive analogues for the classical *t*-, χ^2- and *F*-tests. Therefore it is expedient to use *parametric bootstrap tests*, which are applied in point process statistics since Ripley (1977) and Besag and Diggle (1977). For an exposition of the theory of parametric bootstrap tests, see for example Davison and Hinkley (1997).

The following discusses a simple variant of goodness-of fit tests, so-called *deviation tests*, which have nature similar to that of the Kolmogorov–Smirnov and Cramér–von Mises tests.

A cursory description of a typical test runs as follows. Suppose the hypothesis to be considered is that a given point pattern φ, observed through a window W, is a sample of a point process Φ. One chooses a test statistic τ, such as

$$\tau = \max_{r_1 \le r \le r_2} |D(r) - \hat{D}(r)| \tag{4.137}$$

or

$$\tau = \int_{r_1}^{r_2} \left(D(r) - \hat{D}(r)\right)^2 \mathrm{d}r. \tag{4.138}$$

Here $D(r)$ is the theoretical model distribution function of nearest-neighbour distance and $\hat{D}(r)$ the corresponding empirical function. The deviation measure with L^∞-norm is of Kolmogorov–Smirnov type and that with the L^2-norm is of Cramér–von Mises type. The choice of which one to use may be a matter of taste, but *a priori* knowledge may help; see Section 3.4.2.

One calculates the value τ of the deviation measure for the given sample $\varphi \in W$, denoted by τ_{emp}. This value is compared with analogous τ-values obtained by simulating the specified model k times, always observing it through the window W.

If τ_{emp} takes an extreme position in the series of ordered τ-values, then the hypothesis may be cast in doubt. It is then possible to perform a test of the hypothesis at a significance level α. If no estimated parameter is used in the simulation of the specified model, the significance level is *exact*. One rejects the null hypothesis if τ_{emp}, when pooled together with all τ-values, has a rank larger than or equal to $(1-\alpha)(k+1)$. The p-value of the deviation test can be approximated by Formula (2.64). Loosmore and Ford (2006) discussed the variation of $\hat{p}$ as a function of k and suggested $k = 999$ for a good approximation of the p-value. However, for $\alpha = 0.05$, even using only $k = 99$ will not cause a serious loss of power (Davison and Hinkley, 1997, p. 156). For $\alpha = 0.05$ and $k = 999$ the critical ranks are 950 or above, for $\alpha = 0.01$ and $k = 999$ they are 990 or above.

Summary characteristics commonly chosen for constructing deviation measures are $L(r)$, $D(r)$ and $H_s(r)$, playing the rôle of $D(r)$ above. Often the theoretical functions for the specified model must be determined by simulation. Diggle (2003, p. 89) recommended using the k simulated functions and then averaging. In accordance with common practice in classical statistics the density functions $g(r)$, $d(r)$ and $h_s(r)$ are usually not used for testing, because the estimation of $L(r)$, $D(r)$ and $H_s(r)$ is more standardised than that of the pair correlation function $g(r)$ and of the density functions $d(r)$ and $h_s(r)$. A goodness-of-fit test should be based on a different summary characteristic from the one used in estimating the model parameters, for example, by the minimum contrast method.

Deviation tests for goodness-of-fit hypotheses also provide an *ultima ratio* in parameter estimation; see Ripley (1977) or Diggle (1979, 1983). (The minimum χ^2 estimation method of classical statistics parallels this.) A trial-and-error method chooses a set of parameters for a fixed point process model, and then conducts a goodness-of-fit test to see if the parameters are compatible with the empirical data. If they are, then the parameter set can be used as an estimate of the model parameters. If not, then the procedure is repeated for a new set of parameters, and so forth until a satisfactory set is found. This is a rather *ad hoc* procedure, but it can be refined using techniques involving iterative maximisation of likelihood; cf. Diggle and Gratton (1984).

4.8 General point processes

In previous sections, point processes are models for random point patterns in $\mathbb{R}^d$. This consideration is sufficient for most applications in stochastic geometry, but not for all. This chapter concludes with a briefly introduction to the general theory, which is presented in full detail in Daley and Vere-Jones (2008) and Schneider and Weil (2008, 2010).

Now the 'points' are elements of a general space E. To be mathematically precise, E is assumed to be a locally compact space with a countable base or a complete separable metric space. An example, the most important one for this book, is the set $\mathbb{F}$ of all closed subsets of $\mathbb{R}^d$ with the topology sketched in Section 6.1. Then a 'point' is a closed set, for example a line or a ball $B(x, r)$. In this case Schneider and Weil speak about *geometric processes*.

The needed σ-algebra is the Borel σ-algebra $\mathcal{E}$ of E, that is, the σ-algebra generated by the open subsets of E. This σ-algebra contains, for example, sets such as $\mathbb{F}_{B(x,r)}$, which denotes the set of all closed subsets of $\mathbb{R}^d$ which have a non-empty intersection with the ball $B(x, r)$.

A (simple) point process with points in E is then a random locally finite countable set of distinct points in E, or more mathematically, a measurable mapping from some probability space $[\Omega, \mathcal{A}, \mathbf{P}]$ to the set of all locally finite sequences of points in E. As above, the point process Φ denotes the random set of points as well as the corresponding counting measure, that is, $\Phi(A)$ for $A \in \mathcal{E}$ is the random number of points in the set $A \cap \Phi$.

Many ideas of the theory in previous sections can also be used in the general case. It is possible, for example, to define an intensity measure Λ, by the same equation as in the case of $\mathbb{R}^d$:

$$\Lambda(A) = \mathbf{E}\big(\Phi(A)\big) \qquad \text{for } A \in \mathcal{E},$$

and there is a corresponding Campbell theorem.

Poisson processes can also be defined, simply by requiring the two fundamental properties on p. 41 to be valid for Borel sets in E, instead of in $\mathbb{R}^d$.

With stationarity and isotropy it is a bit more complicated. These notions only make sense if for E there exist motions of the types 'translation' and 'rotation'. This is the case for geometric processes. Nevertheless, it is not a straightforward matter to define an intensity or point density, since in these spaces there is not necessarily a counterpart to the Lebesgue measure.

Finally, two important geometric processes should be mentioned here:

(a) Points = compact sets.

In this case Schneider and Weil speak about *particle processes*. A thorough study of these processes shows that these can be well described by marked point processes of the type $\{[x_n; K_n]\}$, where the x_n are points in $\mathbb{R}^d$ and the K_n are compact sets. This leads to the theory of germ–grain processes; see Section 6.5, and the theory of general point processes is not needed there.

(b) Points = flats.

This corresponds to what Schneider and Weil call *flat processes*. The points are here elements in the set $A(d, k)$ of all k-dimensional planes in $\mathbb{R}^d$. For $d = 2$ and $k = 1$, flats are lines in the plane, and for $d = 3$ and $k = 2$ planes in three-dimensional space.

Interpreting a flat process as a marked point process with points x_n in $\mathbb{R}^d$ is not suitable. If one would try to construct such a model, then the marks should be flats. The following scenario illustrates that such a construction would have too many flats. Consider the two-dimensional case and a stationary point process of points in the plane. Assume first that all 'mark' lines are parallel and orthogonal to the x_1-axis. Since in the strip $0 \le x_1 \le 1$ there are infinitely many points of the point process, the number of lines intersecting the interval $[0, 1]$ on the x_1-axis is infinite. The same can be expected for the case where the line directions are independent and uniform random. Thus, this approach does not work in general.

In contrast, Figure 8.7 shows a reasonable sample of a line process, constructed as a point process in $A(2, 1)$.

It can be said, extending the wording in Daley and Vere-Jones (2008, p. 484), that the structure of stationary isotropic line processes as well as of the more general stationary flat processes in $\mathbb{R}^d$ differs radically from that of stationary point processes in $\mathbb{R}^d$.

Line and flat processes are explained and discussed in Section 8.2.