Spatial Point Processes and their Applications

Adrian Baddeley

School of Mathematics & Statistics, University of Western Australia
Nedlands WA 6009, Australia
e-mail: adrian@maths.uwa.edu.au

A spatial point process is a random pattern of points in d-dimensional space (where usually $d = 2$ or $d = 3$ in applications). Spatial point processes are useful as statistical models in the analysis of observed patterns of points, where the points represent the locations of some object of study (e..g. trees in a forest, bird nests, disease cases, or petty crimes). Point processes play a special role in stochastic geometry, as the building blocks of more complicated random set models (such as the Boolean model), and as instructive simple examples of random sets.

These lectures introduce basic concepts of spatial point processes, with a view toward applications, and with a minimum of technical detail. They cover methods for constructing, manipulating and analysing spatial point processes, and for analysing spatial point pattern data. Each lecture ends with a set of practical computer exercises, which the reader can carry out by downloading a free software package.

Lecture 1 ('Point Processes') gives some motivation, defines point processes, explains how to construct point processes, and gives some important examples. Lecture 2 ('Moments') discusses means and higher moments for point processes, especially the intensity measure and the second moment measure, along with derived quantities such as the K-function and the pair correlation function. It covers the important Campbell formula for expectations. Lecture 3 ('Conditioning') explains how to condition on the event that the point process has a point at a specified location. This leads to the concept of the Palm distribution, and the related Campbell-Mecke formula. A dual concept is the conditional intensity, which provides many new results. Lecture 4 ('Modelling and Statistical Inference') covers the formulation of statistical models for point patterns, model-fitting methods, and statistical inference.

1 Point Processes

In this first lecture, we motivate and define point processes, construct examples (especially the **Poisson process** [28]), and analyse important properties of the Poisson process. There are different ways to mathematically construct and characterise a point process (using finite-dimensional distributions, vacancy probabilities, capacity functional, or generating function). An easier way to construct a point process is by transforming an existing point process (by thinning, superposition, or clustering) [43]. Finally we show how to use existing software to generate simulated realisations of many spatial point processes using these techniques, and analyse them using vacancy probabilities (or 'empty space functions').

1.1 Point Processes in 1D and 2D

A **point process** in one dimension ('time') is a useful model for the sequence of random times when a particular event occurs. For example, the random times when a hospital receives emergency calls may be modelled as a point process. Each emergency call happens at an instant, or point, of time. There will be a random number of such calls in any period of time, and they will occur at random instants of time.

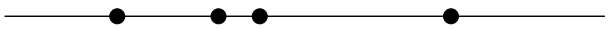

Fig. 1. A point process in time.

A **spatial point process** is a useful model for a random pattern of points in d-dimensional space, where $d \geq 2$. For example, if we make a map of the locations of all the people who called the emergency service during a particular day, this map constitutes a random pattern of points in two dimensions. There will be a random number of such points, and their locations are also random.

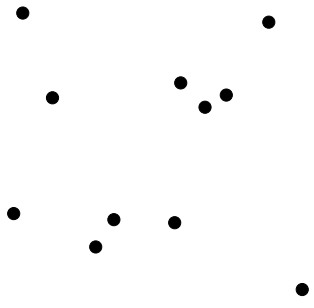

Fig. 2. A point process in two dimensions.

We may also record both the locations and the times of the emergency calls. This may be regarded as a point process in three dimensions (space × time), or alternatively, as a point process in two dimensions where each point (caller location) is labelled or **marked** by a number (the time of the call).

Spatial point processes can be used directly, to model and analyse data which take the form of a point pattern, such as maps of the locations of trees or bird nests ('statistical ecology' [16, 29]); the positions of stars and galaxies ('astrostatistics' [1]); the locations of point-like defects in a silicon crystal wafer (materials science [34]); the locations of neurons in brain tissue; or the home addresses of individuals diagnosed with a rare disease ('spatial epidemiology' [19]). Spatial point processes also serve as a basic model in random set theory [42] and image analysis [41]. For general surveys of applications of spatial point processes, see [16, 42, 43]. For general theory see [15].

1.2 Formulation of Point Processes

There are some differences between the theory of one-dimensional and higher-dimensional point processes, because one-dimensional time has a natural ordering which is absent in higher dimensions.

A one-dimensional point process can be handled mathematically in many different ways. We may study the **arrival times** $T_1 < T_2 < \dots$ where T_i is the time at which the ith point (emergency call) arrives. Using these random variables is the most direct way to handle the point pattern, but their use is complicated by the fact that they are strongly dependent, since $T_i < T_{i+1}$.

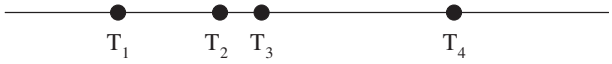

Fig. 3. Arrival times T_i.

Alternatively we may study the **inter-arrival times** $S_i = T_{i+1} - T_i$. These have the advantage that, for some special models (Poisson and renewal processes), the random variables $S_1, S_2, \dots$ are independent.

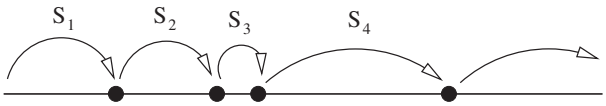

Fig. 4. Inter-arrival times S_i.

Alternatively it is common (especially in connection with martingale theory) to formulate a point process in terms of the cumulative **counting process**

$$\begin{aligned} N_t &= \text{number of points arriving up to time } t \\ &= \sum_{i=1}^{\infty} \mathbf{1}\{T_i \le t\}, \end{aligned}$$

for all $t \ge 0$, where $\mathbf{1}\{\ldots\}$ denotes the indicator function, equal to 1 if the statement "..." is true, and equal to 0 otherwise. This device has the advantage of converting the process to a random function of continuous time t, but has the disadvantage that the values N_t for different t are highly dependent.

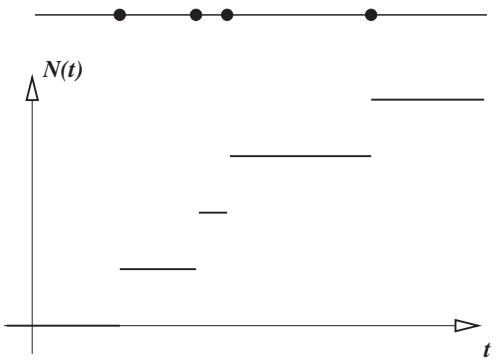

Fig. 5. The counting process N_t associated with a point process.

Alternatively one may use the interval counts

$$N(a,b] = N_b - N_a$$

for $0 \le a \le b$ which count the number of points arriving in the interval $(a,b]$. For some special processes (Poisson and independent-increments processes) the interval counts for **disjoint** intervals are stochastically independent.

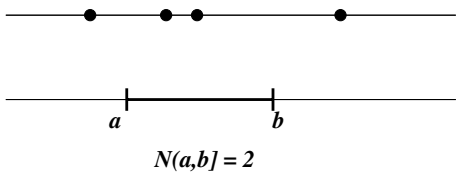

Fig. 6. Interval count $N(a,b]$ for a point process.

In higher dimensions, there is no natural ordering of the points, so that there is no *natural* analogue of the inter-arrival times S_i nor of the counting process N_t. Instead, the most useful way to handle a spatial point process is to generalise the interval counts $N(a,b]$ to the region counts

$$N(B) = \text{number of points falling in } B$$

defined for each bounded closed set $B \subset \mathbb{R}^d$.

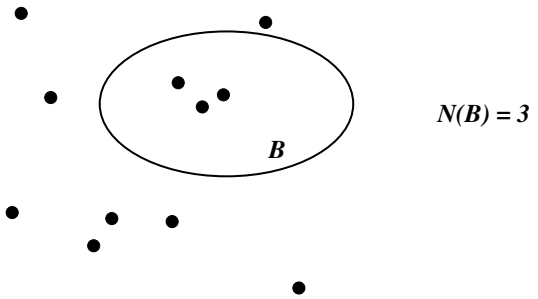

Fig. 7. Counting variables $N(B)$ for a spatial point process.

Rather surprisingly, it is often sufficient to study a point process using only the **vacancy indicators**

$$\begin{aligned} V(B) &= \mathbf{1}\{N(B) = 0\} \\ &= \mathbf{1}\{\text{there are no points falling in } B\}. \end{aligned}$$

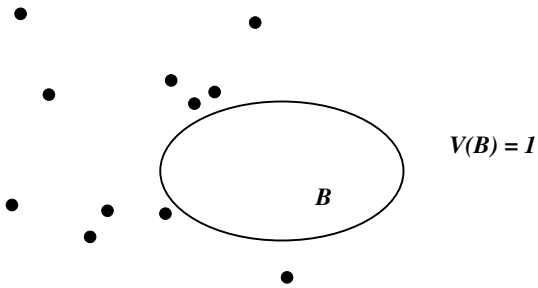

Fig. 8. Vacancy indicators $V(B)$ for a spatial point process.

The counting variables $N(B)$ are natural for exploring additive properties of a point process. For example, suppose we have two point processes, of 'red' and 'blue' points respectively, and we superimpose them (forming a single point process by discarding the colours). If $N_{\text{red}}(B)$ and $N_{\text{blue}}(B)$ are the counting variables for red and blue points respectively, then the counting variable for the superimposed process is $N(B) = N_{\text{red}}(B) + N_{\text{blue}}(B)$.

The vacancy indicators $V(B)$ are natural for exploring geometric and 'multiplicative' properties of a point process. If $V_{\text{red}}(B)$ and $V_{\text{blue}}(B)$ are the vacancy indicators for two point processes, then the vacancy indicator for the superimposed process is $V(B) = V_{\text{red}}(B)\, V_{\text{blue}}(B)$.

1.3 Example: Binomial Process

To take a very simple example, let us place a fixed number n of points at random locations inside a bounded region $W \subset \mathbb{R}^2$. Let $X_1, \ldots, X_n$ be i.i.d. (independent and identically distributed) random points which are uniformly distributed in W. Hence the probability density of each X_i is

$$f(x) = \begin{cases} 1/\lambda_2(W) & \text{if } x \in W \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda_2(W)$ denotes the area of W. A realisation of this process is shown in Figure 9.

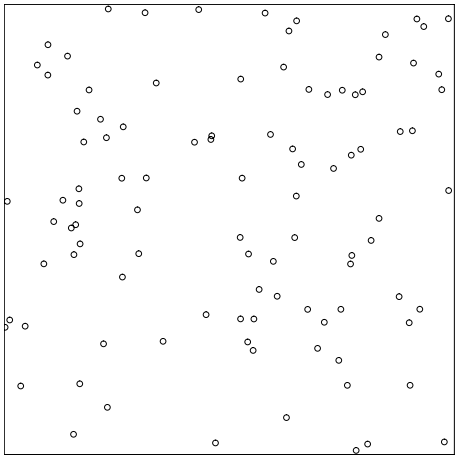

Fig. 9. Realisation of a binomial point process with $n = 100$ in the unit square.

Since each random point X_i is uniformly distributed in W, we have for any bounded set B in $\mathbb{R}^2$

$$\begin{aligned} \mathbb{P}(X_i \in B) &= \int_B f(x)\, \mathrm{d}x \\ &= \frac{\lambda_2(B \cap W)}{\lambda_2(W)}. \end{aligned}$$

The variables $N(B)$ and $V(B)$ may be represented explicitly as

$$\begin{aligned} N(B) &= \sum_{i=1}^{n} \mathbf{1}\{X_i \in B\} \\ V(B) &= \min_{i=1}^{n} \mathbf{1}\{X_i \notin B\} \end{aligned}$$

It follows easily that $N(B)$ has a binomial distribution with parameters n and $p = \lambda_2(B \cap W)/\lambda_2(W)$, hence the process is often called the binomial process.

Note that the counting variables $N(B)$ for different subsets B are not independent. If B_1 and B_2 are disjoint, then

$$N(B_1) + N(B_2) = N(B_1 \cup B_2) \leq n$$

so that $N(B_1)$ and $N(B_2)$ must be dependent. In fact, the joint distribution of $(N(B_1), N(B_2))$ is the multinomial distribution on n trials with success probabilities (p_1, p_2) where $p_i = \lambda_2(B_i \cap W)/\lambda_2(W)$.

1.4 Foundations

Foundations of the theory of point processes in $\mathbb{R}^d$ are expounded in detail in [15]. The following is a very brief and informal introduction.

Random Measure Formalism

The values of the counting variables $N(B)$ for all subsets B give us sufficient information to reconstruct completely the positions of all the points in the process. Indeed the points of the process are those locations x such that $N(\{x\}) > 0$. Hence we may as well *define* a point process as a collection of random variables $N(B)$ indexed by subsets B.

The counting variables $N(B)$ for different sets B satisfy certain relationships, including additivity

$$N(A \cup B) = N(A) + N(B)$$

whenever A and B are disjoint sets ($A \cap B = \emptyset$) and of course

$$N(\emptyset) = 0$$

where $\emptyset$ denotes the empty set. Furthermore, they are continuous in the sense that, if A_n is a decreasing sequence of closed, bounded sets ($A_n \supseteq A_{n+1}$) with limit $\bigcap_n A_n = A$, then we must have

$$N(A_n) \to N(A).$$

These properties must hold for each realisation of the point process, or at least, with probability 1. They amount to the requirement that N is a *measure* (or at least, that with probability 1, the values $N(B)$ can be extended to a measure). This is the concept of a **random measure** [26, 42].

Formally, then, a point process may be defined as a random measure in which the values $N(B)$ are nonnegative integers [15, 42]. We usually also assume that the point process is **locally finite**:

$$N(B) < \infty \quad \text{with probability 1}$$

for all bounded $B \subset \mathbb{R}^d$. That is, any bounded region contains only a finite number of points, with probability 1. We also assume that the point process is **simple**:

$$N(\{x\}) \le 1 \quad \text{for all } x \in \mathbb{R}^d$$

with probability 1. That is, with probability 1, no two points of the process are coincident. A simple point process can be regarded as a random set of points.

For example, the binomial process introduced in Section 1.3 is locally finite (since $N(B) \le n$ for all B) and it is simple because there is zero probability that two independent, uniformly distributed random points coincide:

$$\mathbb{P}(X_1 = X_2) = \mathbb{E}\left[\mathbb{P}\left(X_1 = X_2 \mid X_2\right)\right] = 0.$$

Hence the binomial process is a point process in the sense of this definition.

Random Set Formalism

A *simple* point process can be formulated in a completely different way since it may be regarded as a random set $\mathbf{X}$. Interestingly, the vacancy indicators $V(B)$ contain complete information about the process. If we know the value of $V(B)$ for all sets B, then we can determine the exact location of each point x in the (simple) point process $\mathbf{X}$. To do this, let G be the union of all open sets B such that $V(B) = 1$. The complement of G is a locally finite set of points, and this identifies the random set $\mathbf{X}$.

The vacancy indicators must satisfy

$$V(A \cup B) = \min\{V(A), V(B)\}$$

for any sets A, B, and have other properties analogous to those of the count variables $N(B)$. Thus we could alternatively define a simple point process as a random function V satisfying these properties almost surely. This approach is intimately related to the theory of random closed sets [27, 31, 32].

In the rest of these lectures, we shall often swap between the notation $\mathbf{X}$ (for a point process when it is considered as a random set) and N or $N_{\mathbf{X}}$ (for the counting variables associated with the same point process).

1.5 Poisson Processes

One-dimensional Poisson Process

Readers may be familiar with the concept of a **Poisson point process** in one-dimensional time (e.g. [28, 37]). Suppose we make the following assumptions:

1. The number of points which arrive in a given time interval has expected value proportional to the duration of the interval:

   $$\mathbb{E}N(a, b] = \beta(b - a)$$

   where $\beta > 0$ is the **rate** or **intensity** of the process;

2. Arrivals in disjoint intervals of time are independent: if $a_1 < b_1 < a_2 < b_2 < \ldots < a_m < b_m$ then the random variables $N(a_1, b_1], \ldots, N(a_m, b_m]$ are independent;
3. The probability of two or more arrivals in a given time interval is asymptotically of uniformly smaller order than the length of the interval:

$$\mathbb{P}(N(a, a+h] \geq 2) = o(h), \qquad h \downarrow 0.$$

For example these would be reasonable assumptions to make about the arrival of cosmic particles at a particle detector, or the occurrence of accidents in a large city.

From these assumptions it *follows* that the number of points arriving in a given time interval must have a Poisson distribution:

$$N(a, b] \sim \mathsf{Poisson}(\beta(b-a))$$

where $\mathsf{Poisson}(\mu)$ denotes the Poisson distribution with mean μ, defined by

$$\mathbb{P}(N = k) = e^{-\mu}\frac{\mu^k}{k!}, \qquad k = 0, 1, 2, \ldots \tag{1}$$

This conclusion follows by splitting the interval $(a, b]$ into a large number n of small intervals. The number of arrivals in each small interval is equal to 0 or 1, except for an event of small probability. Since $N(a, b]$ is the sum of these numbers, it has an approximately binomial distribution. Letting $n \to \infty$ we obtain that $N(a, b]$ must have a Poisson distribution.

Definition 1.1. *The* **one-dimensional Poisson process***, with uniform intensity $\beta > 0$, is a point process in $\mathbb{R}$ such that*

[PP1] *for every bounded interval $(a, b]$, the count $N(a, b]$ has a Poisson distribution with mean $\beta(b-a)$;*
[PP2] *if $(a_1, b_1], \ldots, (a_m, b_m]$ are disjoint bounded intervals, then the counts $N(a_1, b_1], \ldots, N(a_m, b_m]$ are independent random variables.*

Other properties of the one-dimensional Poisson process include

1. The inter-arrival times S_i have an exponential distribution with rate β:

$$\mathbb{P}(S_i \leq s) = 1 - e^{-\beta s}, \qquad s > 0.$$

2. The inter-arrival times S_i are independent.
3. The ith arrival time T_i has an Erlang or Gamma distribution with parameters $\alpha = i$ and β. The $\mathsf{Gamma}(\alpha, \beta)$ probability density is

$$f(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$$

for $t > 0$, and 0 otherwise.

Fig. 10. Realisation of the one-dimensional Poisson process with uniform intensity 1 in the time interval $[0, 30]$. Tick marks indicate the arrival times.

Properties 1 and 2 above suggest an easy way to generate simulated realisations of the Poisson process on $[0, \infty)$. We simply generate a sequence of independent, exponentially distributed, random variables $S_1, S_2, \ldots$ and take the arrival times to be $T_i = \sum_{1 \le j \le i} S_j$.

We may also study **inhomogeneous Poisson processes** in which the number of arrivals in $(a, b]$ is

$$\mathbb{E}\, N(a, b] = \int_a^b \beta(t)\, \mathrm{d}t$$

where $\beta(t) > 0$ is a function called the (instantaneous) **intensity function**. The probability that there will be a point of this process in an infinitesimal interval $[t, t + \mathrm{d}t]$ is $\beta(t)\, \mathrm{d}t$. Arrivals in disjoint time intervals are independent.

Spatial Poisson Process

The Poisson process can be generalised to two-dimensional space.

Definition 1.2. *The* **spatial Poisson process***, with uniform intensity $\beta > 0$, is a point process in $\mathbb{R}^2$ such that*

[PP1] *for every bounded closed set B, the count $N(B)$ has a Poisson distribution with mean $\beta\lambda_2(B)$;*
[PP2] *if $B_1, \ldots, B_m$ are disjoint regions, then $N(B_1), \ldots, N(B_m)$ are independent.*

Here $\lambda_2(B)$ again denotes the area of B.

It turns out that these two properties uniquely characterise the Poisson process. The constant β is the expected number of points per unit area. It has dimensions length^{-2} or "points per unit area".

As in the one-dimensional case, the spatial Poisson process can be derived by starting from a few reasonable assumptions: that $\mathbb{E}N(B) = \beta\lambda_2(B)$; that $\mathbb{P}(N(B) > 1) = o(\lambda_2(B))$ for small $\lambda_2(B)$; and that events in disjoint regions are independent.

An important fact about the Poisson process is the following.

Lemma 1.1 (Conditional Property). *Consider a Poisson point process in $\mathbb{R}^2$ with uniform intensity $\beta > 0$. Let $W \subset \mathbb{R}^2$ be any region with $0 < \lambda_2(W) < \infty$. Given that $N(W) = n$, the conditional distribution of $N(B)$ for $B \subseteq W$ is binomial:*

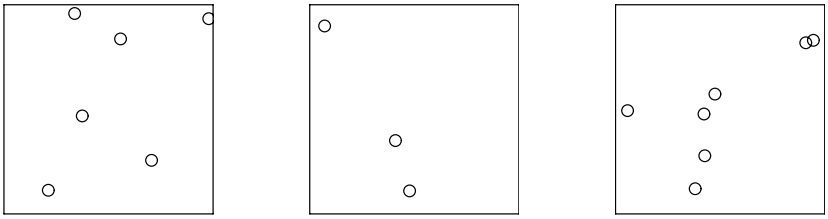

Fig. 11. Three different realisations of the Poisson process with uniform intensity 5 in the unit square.

$$\mathbb{P}\left(N(B) = k \mid N(W) = n\right) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $p = \lambda_2(B)/\lambda_2(W)$. *Furthermore the conditional joint distribution of* $N(B_1), \ldots, N(B_m)$ *for any* $B_1, \ldots, B_m \subseteq W$ *is the same as the joint distribution of these variables in a binomial process.*

In other words, *given that there are* n *points of the Poisson process in* W, *these* n *points are conditionally independent and uniformly distributed in* W.

Proof. Let $0 \leq k \leq n$. Then

$$\begin{aligned}\mathbb{P}\left(N(B) = k \mid N(W) = n\right) &= \frac{\mathbb{P}(N(B) = k,\ N(W) = n)}{\mathbb{P}(N(W) = n)} \\ &= \frac{\mathbb{P}(N(B) = k,\ N(W \setminus B) = n - k)}{\mathbb{P}(N(W) = n)}.\end{aligned}$$

By the independence property (PP2) the numerator can be rewritten

$$\mathbb{P}(N(B) = k,\ N(W \setminus B) = n - k) = \mathbb{P}(N(B) = k)\,\mathbb{P}(N(W \setminus B) = n - k)$$

We may then evaluate the numerator and denominator using (PP1) to give

$$\begin{aligned}\mathbb{P}\left(N(B) = k \mid N(W) = n\right) &= \frac{e^{-\beta\lambda_2(B)} \frac{(\beta\lambda_2(B))^k}{k!} e^{-\beta\lambda_2(W\setminus B)} \frac{(\beta\lambda_2(W\setminus B))^{n-k}}{(n-k)!}}{e^{-\beta\lambda_2(W)} \frac{(\beta\lambda_2(W))^n}{n!}} \\ &= \frac{n!}{k!\,(n-k)!} \left(\frac{\lambda_2(B)}{\lambda_2(W)}\right)^k \left(\frac{\lambda_2(W \setminus B)}{\lambda_2(W)}\right)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k}\end{aligned}$$

where $p = \lambda_2(B)/\lambda_2(W)$. □

Thus, for example, Figure 9 can also be taken as a realisation of a Poisson process in the unit square W, in which it happens that there are exactly 100 points in W. The only distinction between a binomial process and a Poisson

process in W is that *different realisations* of the Poisson process will consist of different numbers of points.

The conditional property also gives us a direct way to simulate Poisson processes. To generate a realisation of a Poisson process of intensity β in W, we first generate a random variable M with a Poisson distribution with mean $\beta\lambda_2(W)$. Given $M = m$, we then generate m independent uniform random points in W.

General Poisson Process

To define a uniform Poisson point process in $\mathbb{R}^d$, or an inhomogeneous Poisson process in $\mathbb{R}^d$, or a Poisson point process on some other space S, the following general definition can be used.

Definition 1.3. *Let S be a space, and Λ a measure on S. (We require S to be a locally compact metric space, and Λ a measure which is finite on every compact set and which has no atoms.)*

The **Poisson process** *on S with intensity measure Λ is a point process on S such that*

[PP1] *for every compact set $B \subset S$, the count $N(B)$ has a Poisson distribution with mean $\Lambda(B)$;*

[PP2] *if $B_1, \ldots, B_m$ are disjoint compact sets, then $N(B_1), \ldots, N(B_m)$ are independent.*

Example 1.1 (Poisson process in three dimensions). The uniform Poisson process on $\mathbb{R}^3$ with intensity $\beta > 0$ is defined by taking $S = \mathbb{R}^3$ and $\Lambda(B) = \beta\lambda_3(B)$.

Example 1.2 (Inhomogeneous Poisson process). The inhomogeneous Poisson process on $\mathbb{R}^2$ with intensity function $\beta(u)$, $u \in \mathbb{R}^2$ is defined by taking $S = \mathbb{R}^2$ and $\Lambda(B) = \int_B \beta(u)\, \mathrm{d}u$. See Figure 12.

Example 1.3 (Poisson process on the sphere). Take S to be the unit sphere (surface of the unit ball in three dimensions) and $\Lambda = \beta\mu$, where $\beta > 0$ and μ is the uniform area measure on S with total mass 4π. This yields the uniform Poisson point process on the unit sphere, with intensity β. This process has a finite number of points, almost surely. Indeed the total number of points $N(S)$ is a Poisson random variable with mean $\Lambda(S) = \beta\mu(S) = 4\pi\beta$. See Figure 13.

1.6 Distributional Characterisation

In Section 1.5 we discussed the fact that a Poisson process in a bounded region W, conditioned on the total number of points in W, is equivalent to a binomial process. This was expressed somewhat vaguely, because we do not yet have the tools needed to determine whether two point processes are 'equivalent' in distribution. We now develop such tools.

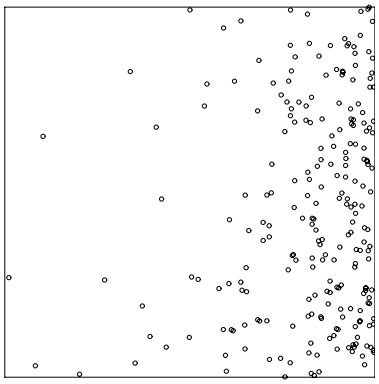

Fig. 12. Realisation of an inhomogeneous Poisson process in the unit square, with intensity function $\beta(x, y) = \exp(2 + 5x)$.

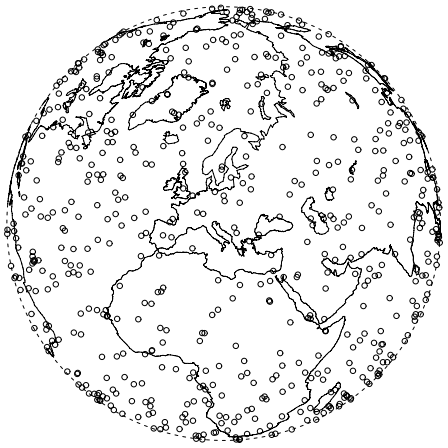

Fig. 13. Uniform Poisson point process on the surface of the Earth. Intensity is $\beta = 100$ points per solid radian; the expected total number of points is $4\pi \times 100 = 1256.6$. Orthogonal projection from a position directly above Martina Franca.

Space of Outcomes

Like any random phenomenon, a point process can be described in statistical terms by defining the space of possible outcomes and then specifying the probabilities of different *events* (an event is a subset of all possible outcomes).

The space of realisations of a point process in $\mathbb{R}^d$ is N, the set of all counting measures on $\mathbb{R}^d$, where a **counting measure** is a nonnegative integer valued measure which has a finite value on every compact set.

A basic event about the point process is the event that there are exactly k points in the region B,

$$E_{B,k} = \{N(B) = k\} = \{N \in \mathsf{N} : \ N(B) = k\}$$

for compact $B \subset \mathbb{R}^d$ and integer $k = 0, 1, 2, \ldots$.

Definition 1.4. *Let* N *be the set of all counting measures on* $\mathbb{R}^d$. *Let* $\mathcal{N}$ *be the* σ*-field of subsets of* N *generated by all events of the form* $E_{B,k}$. *The space* N *equipped with its* σ*-field* $\mathcal{N}$ *is called the* **canonical space** *or* **outcome space** *for a point process in* $\mathbb{R}^d$.

The σ-field $\mathcal{N}$ includes events such as

$$E_{B_1,k_1} \cap \ldots \cap E_{B_m,k_m} = \{N \in \mathsf{N} : \ N(B_1) = k_1, \ldots, N(B_m) = k_m\},$$

i.e. the event that there are exactly k_i points in region B_i for $i = 1, \ldots, m$. It also includes, for example, the event that the point process has no points at all,

$$\{N \equiv 0\} = \{N \in \mathsf{N} : \ N(B) = 0 \text{ for all } B\}$$

since this event can be represented as the intersection of the countable sequence of events $E_{b(0,n),0}$ for $n = 1, 2, \ldots$. Here $b(0, r)$ denotes the ball of radius r and centre 0 in $\mathbb{R}^d$.

A point process $\mathbf{X}$ may now be defined formally, using its counting measure $N = N_{\mathbf{X}}$, as a measurable map $N : \Omega \to \mathsf{N}$ from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to the outcome space $(\mathsf{N}, \mathcal{N})$. Thus, each elementary outcome $\omega \in \Omega$ determines an outcome $N^\omega \in \mathsf{N}$ for the entire point process. Measurability is the requirement that, for any event $E \in \mathcal{N}$, the event

$$\{N \in E\} = \{\omega \in \Omega : \ N^\omega \in E\}$$

belongs to $\mathcal{A}$. This implies that any such event has a well-defined probability $\mathbb{P}(N \in E)$. For example, the probability that the point process is empty, $\mathbb{P}(N \equiv 0)$, is well defined.

The construction of $\mathcal{N}$ guarantees that, if N is a point process on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then the variables $N(B)$ for each compact set B are random variables on the same probability space. In fact $\mathcal{N}$ is the minimal σ-field on N which guarantees this.

Definition 1.5. *The* **distribution** *of a point process* $\mathbf{X}$ *is the probability measure* $\mathbf{P}_{\mathbf{X}}$, *on the outcome space* $(\mathsf{N}, \mathcal{N})$, *defined by*

$$\mathbf{P}_{\mathbf{X}}(A) = \mathbb{P}(N_{\mathbf{X}} \in A), \qquad A \in \mathcal{N}.$$

For example, the distribution of a point process specifies the values of joint probabilities

$$\mathbb{P}(N(B) = k \text{ and } N(B') = k')$$

for two sets B, B' and integers k, k'; it also specifies the probability that the entire point process is empty,

$$\mathbb{P}(N \equiv 0) = \mathbb{P}(\mathbf{X} = \emptyset).$$

Characterisations of a Point Process Distribution

The distribution of a point process may be characterised using either the joint distributions of the variables $N(B)$, or the marginal distributions of the variables $V(B)$. First we consider the count variables $N(B)$.

Definition 1.6. *The* **finite-dimensional distributions** *or* **fidi***s of a point process are the joint probability distributions of*

$$(N(B_1), \ldots, N(B_m))$$

for all finite integers $m > 0$ and all compact $B_1, B_2, \ldots$.

Equivalently, the fidis specify the probabilities of all events of the form

$$\{N(B_1) = k_1, \ldots, N(B_m) = k_m\}$$

involving finitely many regions.

Clearly the fidis of a point process convey only a subset of the information conveyed in its distribution. Probabilities of events such as $\{\mathbf{X} = \emptyset\}$ are not specified in the fidis, since they cannot be expressed in terms of a finite number of compact regions. However, it turns out that the fidis are sufficient to characterise the entire distribution.

Theorem 1.1. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two point processes. If the fidis of* $\mathbf{X}$ *and of* $\mathbf{Y}$ *coincide, then* $\mathbf{X}$ *and* $\mathbf{Y}$ *have the same distribution.*

Corollary 1.1. *If* $\mathbf{X}$ *is a point process satisfying axioms (PP1) and (PP2) then* $\mathbf{X}$ *is a Poisson process.*

A *simple* point process (Section 1.4) can be regarded as a random set of points. In this case the vacancy probabilities are useful. The **capacity functional** of a simple point process $\mathbf{X}$ is the functional

$$T(K) = \mathbb{P}(N(K) > 0), \qquad K \text{ compact.}$$

This is a very small subset of the information conveyed by the fidis, since $T(K) = 1 - \mathbb{P}(E_{K,0})$. However, surprisingly, it turns out that the capacity functional is sufficient to determine the entire distribution.

Theorem 1.2. *Suppose* $\mathbf{X}$ *and* $\mathbf{Y}$ *are two simple point processes whose capacity functionals are identical. Then their distributions are identical.*

Corollary 1.2. *A simple point process is a uniform Poisson process of intensity β if and only if its capacity functional is*

$$T(K) = 1 - \exp\{-\beta\lambda_d(K)\}$$

for all compact $K \subset \mathbb{R}^d$.

Corollary 1.3. *A simple point process is a binomial process (of n points in W) if and only if its capacity functional is*

$$T(K) = 1 - \left(1 - \frac{\lambda_d(K \cap W)}{\lambda_d(W)}\right)^n$$

for all compact $K \subset \mathbb{R}^d$.

This characterisation of the binomial process now makes it easy to prove the conditional property of the Poisson process described in the last section.

Note that the results above do not provide a simple way to construct a point process *ab initio*. Theorem 1.1 does not say that any given choice of finite dimensional distributions will automatically determine a point process distribution. On the contrary, the fidis must satisfy a suite of conditions (self-consistency, continuity) if they are to correspond to a point process. Hence, the fidis are not a very practical route to the *construction* of point processes. More practical methods of construction are described in Section 1.7.

The concept of a **stationary** point process plays an important role.

Definition 1.7. *A point process $\mathbf{X}$ in $\mathbb{R}^d$ is called* **stationary** *if, for any fixed vector $v \in \mathbb{R}^d$, the distribution of the shifted point process $\mathbf{X} + v$ (obtained by shifting each point $x \in \mathbf{X}$ to $x + v$) is identical to the distribution of $\mathbf{X}$.*

Lemma 1.2. *A point process is stationary if and only if its capacity functional is invariant under translations, $T(K) = T(K+v)$ for all compact sets $K \subset \mathbb{R}^d$ and all $v \in \mathbb{R}^d$.*

For example, the uniform Poisson process is stationary, since its capacity functional $T(K)$ is clearly invariant under translation.

Similarly, a point process is called **isotropic** if its distribution is invariant under all rotations of $\mathbb{R}^d$. The uniform Poisson process is isotropic.

1.7 Transforming a Point Process

One pragmatic way to construct a new point process is by transforming or changing an existing point process. Convenient transformations include **mapping**, **thinning**, **superposition**, and **clustering**.

Mapping

Figure 14 sketches in one dimension the concept of **mapping** a point process $\mathbf{X}$ to another point process by applying a fixed transformation $s : \mathbb{R}^d \to \mathbb{R}^d$ to each individual point of $\mathbf{X}$. The resulting point process is thus $\mathbf{Y} = \bigcup_{x \in \mathbf{X}} s(x)$. For example, the mapping $s(x) = ax$ where $a > 0$ would rescale the entire point process by the constant scale factor a.

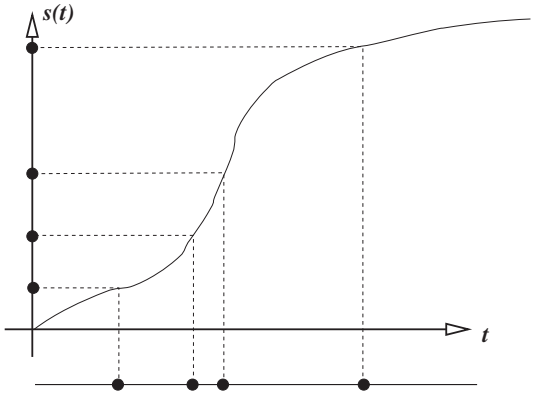

Fig. 14. Application of a transformation s to each individual point in a point process

A vector translation $s(x) = x + v$, where $v \in \mathbb{R}^d$ is fixed, shifts all points of $\mathbf{X}$ by the same vector v. If the original process $\mathbf{X}$ is a uniform Poisson process, then the translated point process $\mathbf{Y}$ is also a uniform Poisson process with the same intensity, as we saw above.

Any mapping s which has a continuous inverse, or at least which satisfies

$$0 < \lambda_d(s^{-1}(B)) < \infty \text{ whenever } B \text{ is compact} \tag{2}$$

transforms a uniform Poisson process into another Poisson process, generally an inhomogeneous one.

An important caution is that, if the transformation s does not satisfy (2), then in general we cannot even be sure that the transformed point process $\mathbf{Y}$ is well defined, since the points of $\mathbf{Y}$ may not be locally finite. For example, consider the projection of the cartesian plane onto the x-axis, $s(x, y) = x$. If $\mathbf{X}$ is a uniform Poisson process in $\mathbb{R}^2$ then the projection onto the x-axis is everywhere dense: there are infinitely many projected points in any open interval (a, b) in the x-axis, almost surely, since $s^{-1}((a, b)) = (a, b) \times \mathbb{R}$. Hence, the projection of $\mathbf{X}$ onto the x-axis is not a well-defined point process.

Thinning

Figure 15 sketches the operation of **thinning** a point process $\mathbf{X}$, by which some of the points of $\mathbf{X}$ are deleted. The remaining, undeleted points form the thinned point process $\mathbf{Y}$. We may formalise the thinning procedure by supposing that each point $x \in \mathbf{X}$ is labelled with an indicator random variable I_x taking the value 1 if the point x is to be retained, and 0 if it is to be deleted. Then the thinned process consists of those points $x \in \mathbf{X}$ with $I_x = 1$.

Independent thinning is the case where the indicators I_x are independent. If a uniform Poisson process is subjected to independent thinning, the resulting thinned process is also Poisson.

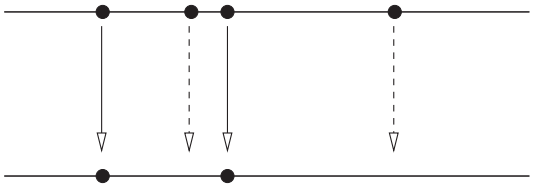

Fig. 15. Thinning a point process. Points of the original process (above) are either retained (solid lines) or deleted (dotted lines) to yield a thinned process (below).

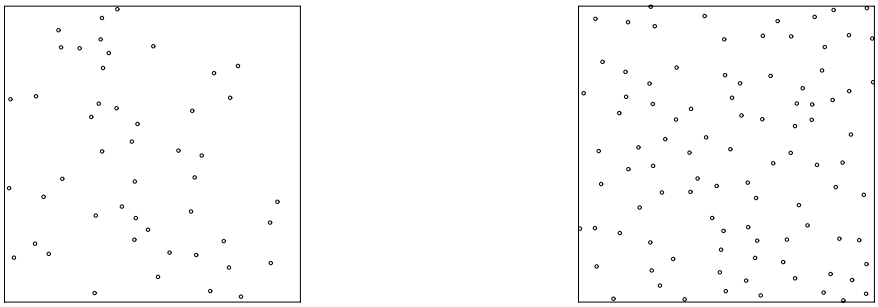

Fig. 16. Dependent thinning: simulated realisations of Matern's Model I (left) and Model II (right). Both are derived from a Poisson process of intensity 200 in the unit square, and have the same inhibition radius $r = 0.05$.

Examples of **dependent thinning** are the two models of Matérn [30] for spatial inhibition between points. In **Model I**, we start with a uniform Poisson process $\mathbf{X}$ in $\mathbb{R}^2$, and delete any point which has a close neighbour (closer than a distance r, say). Thus $I_x = 1$ if $||x - x'|| \leq r$ for any $x' \in \mathbf{X}$. In **Model II**, we start with a uniform Poisson process $\mathbf{X}$ in $\mathbb{R}^2 \times [0, 1]$, interpreting this as a process of two-dimensional points $x \in \mathbb{R}^2$ with 'arrival times' $t \in [0, 1]$. Then we delete any point which has a close neighbour whose arrival time was earlier than the point in question. Thus $I_{(x,t)} = 1$ if $||x - x'|| \leq r$ and $t > t'$ for any $(x', t') \in \mathbf{X}$. The arrival times are then discarded to give us a point process in $\mathbb{R}^2$. Simulated realisations of these two models are shown in Figure 16.

Superposition

Figure 17 sketches the **superposition** of two point processes $\mathbf{X}$ and $\mathbf{Y}$ which consists of all points in the union $\mathbf{X} \cup \mathbf{Y}$. If we denote by $N_{\mathbf{X}}(B)$ and $N_{\mathbf{Y}}(B)$ the numbers of points of $\mathbf{X}$ and $\mathbf{Y}$ respectively in a region $B \subset \mathbb{R}^d$, then the superposition has $N_{\mathbf{X}\cup\mathbf{Y}}(B) = N_{\mathbf{X}}(B) + N_{\mathbf{Y}}(B)$ assuming there are no coincident points. Superposition can thus be viewed either as the union of sets or as the sum of measures.

If $\mathbf{X}$ and $\mathbf{Y}$ are *independent,* with capacity functionals $T_{\mathbf{X}}, T_{\mathbf{Y}}$, then the superposition has capacity functional $T_{\mathbf{X}\cup\mathbf{Y}}(K) = 1-(1-T_{\mathbf{X}}(K))(1-T_{\mathbf{Y}}(K))$.

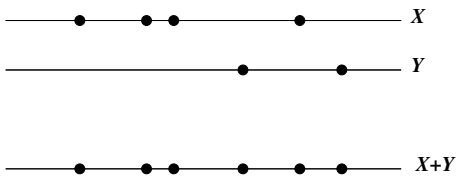

Fig. 17. Superposition of two point processes

The superposition of two *independent* Poisson processes $\mathbf{X}$ and $\mathbf{Y}$, say uniform Poisson processes of intensity μ and ν respectively, is a uniform Poisson process of intensity $\mu + \nu$.

Cluster Formation

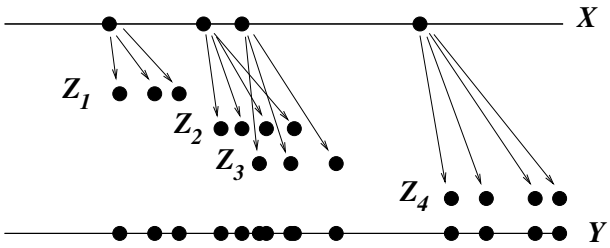

Fig. 18. Schematic concept of the formation of a cluster process.

Finally, in a **cluster process**, we start with a point process $\mathbf{X}$ and replace each point $x \in \mathbf{X}$ by a random finite set of points Z_x called the cluster associated with x. The superposition of all clusters yields the process $\mathbf{Y} = \bigcup_{x \in \mathbf{X}} Z_x$. See Figure 18.

Usually it is assumed that the clusters Z_x for different parent points x are independent processes. A simple example is the **Matérn cluster process** in which the 'parent' process $\mathbf{X}$ is a uniform Poisson process in $\mathbb{R}^2$, and each cluster Z_x consists of a random number M_x of points, where $M_x \sim \mathsf{Poisson}(\mu)$, independently and uniformly distributed in the disc $b(x, r)$ of radius r centred on x. Simulated realisations of this process are shown in Figure 19.

1.8 Marked Point Processes

Earlier we mentioned the idea that the points of a point process might be labelled with extra information called **marks**. For example, in a map of the locations of emergency calls, each point might carry a label stating the time of the call and the nature of the emergency.

A marked point can be formalised as a pair (x, m) where x is the point location and m is the mark attached to it.

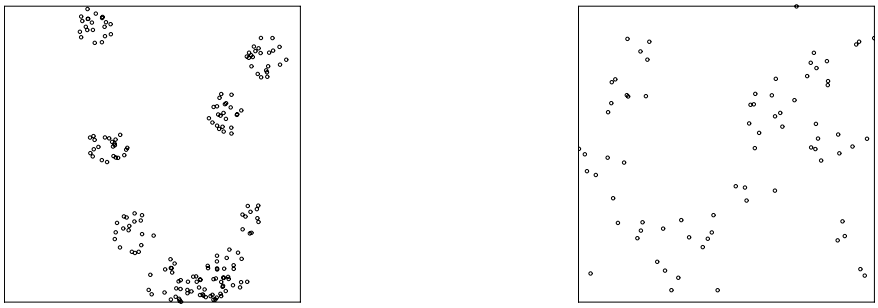

Fig. 19. Simulated realisations of the Matérn cluster process in the unit square. *Left:* parent intensity $\beta = 5$, mean cluster size $\mu = 20$, cluster radius $r = 0.07$. *Right:* $\beta = 50$, $\mu = 2$, $r = 0.07$. Both processes have an average of 100 points in the square.

Definition 1.8. *A* **marked point process** *on a space S with marks in a space M is a point process $\mathbf{Y}$ on $S \times M$ such that $N_Y(K \times M) < \infty$ a.s. for all compact $K \subset S$. That is, the corresponding projected process (of points without marks) is locally finite.*

Note that the space of marks M can be very general. It may be a finite set, a continuous interval of real numbers, or a more complicated space such as the set of all convex polygons.

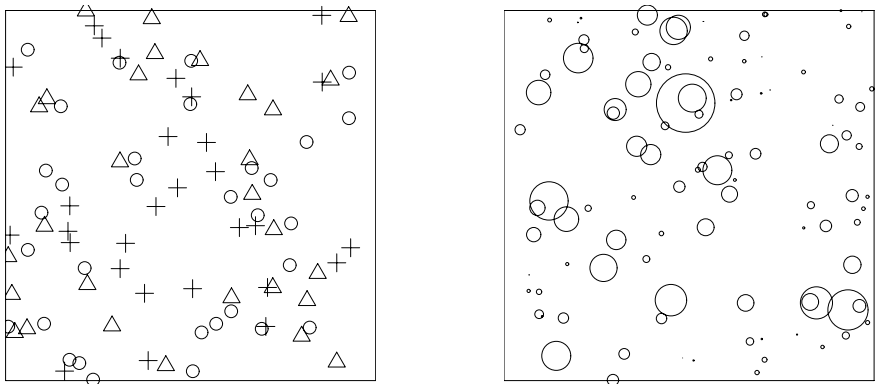

Fig. 20. Realisations of marked point processes in the unit square. *Left:* finite mark space $M = \{a, b, c\}$, marks plotted as symbols $\triangle, O, +$. *Right:* continuous mark space $M = [0, \infty)$, marks plotted as radii of circles.

Example 1.4. Let $\mathbf{Y}$ be a uniform Poisson process in $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$. This *cannot* be interpreted as a marked point process in $\mathbb{R}^2$ with marks in $\mathbb{R}$, because the finiteness condition fails. The set of marked points (x, m) which project into

a given compact set $K \subset \mathbb{R}^2$ is the solid region $K \times \mathbb{R}$, which has infinite volume, and hence contains infinitely many marked points, almost surely.

Example 1.5. Let $\mathbf{Y}$ be a uniform Poisson process on the three-dimensional slab $R^2 \times [0, a]$ with intensity β. This can be interpreted as a marked point process on $\mathbb{R}^2$ with marks in $M = [0, a]$. The finiteness condition is clearly satisfied. The projected point process (i.e. obtained by ignoring the marks) is a uniform Poisson process in $\mathbb{R}^2$ with intensity βa. By properties of the uniform distribution, the marks attached to different points are independent and uniformly distributed in $[0, a]$.

A marked point process formed by attaching independent random marks to a Poisson process of locations, is equivalent to a Poisson process in the product space.

Theorem 1.3. *Let* $\mathbf{Y}$ *be a marked point process on* S *with marks in* M*. Let* $\mathbf{X}$ *be the projected process in* S *(of points without marks). Then the following are equivalent:*

1. $\mathbf{X}$ *is a Poisson process in* S *with intensity* μ*, and given* $\mathbf{X}$*, the marks attached to the points of* $\mathbf{X}$ *are independent and identically distributed with common distribution* Q *on* M*;*
2. $\mathbf{Y}$ *is a Poisson process in* $S \times M$ *with intensity measure* $\mu \otimes Q$*.*

See e.g. [28]. This result can be obtained by comparing the capacity functionals of the two processes.

Marked point processes are also used in the formal description of operations like thinning and clustering. For example, thinning a point process $\mathbf{X}$ is formalised by construct a marked point process with marks in $\{0, 1\}$. The mark I_x attached to each point x indicates whether the point is to be retained (1) or deleted (0).

1.9 Distances in Point Processes

One simple way to analyse a point process is in terms of the distances between points. If $\mathbf{X}$ is a point process, let $\mathsf{dist}(u, \mathbf{X})$ for $u \in \mathbb{R}^d$ denote the shortest distance from the given location u to the nearest point of $\mathbf{X}$. This is sometimes called the **contact distance**. Note the key fact that

$$\mathsf{dist}(u, \mathbf{X}) \leq r \text{ if and only if } N(b(u, r)) > 0$$

where $b(u, r)$ is the disc of radius r centred at x. Since $N(b(u, r))$ is a random variable for fixed u and r, the event $\{N(b(u, r)) > 0\}$ is measurable, so the event $\{\mathsf{dist}(u, \mathbf{X}) \leq r\}$ is measurable for all r, which implies that the contact distance $\mathsf{dist}(u, \mathbf{X})$ is a well-defined random variable.

If $\mathbf{X}$ is a uniform Poisson process in $\mathbb{R}^d$ of intensity β, then this insight also gives us the distribution of $\mathsf{dist}(u, \mathbf{X})$:

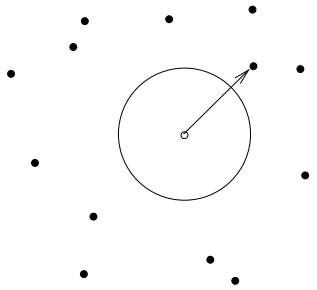

Fig. 21. The contact distance $\mathsf{dist}(u, \mathbf{X})$ from a fixed location ($\circ$) to the nearest random point ($\bullet$) satisfies $\mathsf{dist}(u, \mathbf{X}) > r$ if and only if there are no random points in the disc of radius r centred on the fixed location.

$$\begin{aligned}\mathbb{P}(\mathsf{dist}(u, \mathbf{X}) \le r) &= \mathbb{P}(N(b(u, r)) > 0) \\ &= 1 - \exp(-\beta \lambda_d(b(u, r))) \\ &= 1 - \exp(-\beta \kappa_d r^d)\end{aligned}$$

where $\kappa_d = \lambda_d(b(0, 1))$ is the volume of the unit ball in $\mathbb{R}^d$.

One interesting way to rephrase this is that $V = \kappa_d \mathsf{dist}(u, \mathbf{X})^d$ has an exponential distribution with rate β,

$$\mathbb{P}(V \le v) = 1 - \exp(-\beta v).$$

Notice that V is the volume of the ball of random radius $\mathsf{dist}(u, \mathbf{X})$, or equivalently, the volume of the largest ball centred on u that contains no points of $\mathbf{X}$.

Definition 1.9. *Let* $\mathbf{X}$ *be a stationary point process in* $\mathbb{R}^d$. *The* **contact distribution function** *or* **empty space function** F *is the cumulative distribution function of the distance*

$$R = \mathsf{dist}(u, \mathbf{X})$$

from a fixed point u *to the nearest point of* $\mathbf{X}$. *That is*

$$\begin{aligned}F(r) &= \mathbb{P}(\mathsf{dist}(u, \mathbf{X}) \le r) \\ &= \mathbb{P}(N(b(u, r)) > 0).\end{aligned}$$

By stationarity this does not depend on u.

Notice that $F(r) = T(b(0, r)) = T(b(u, r))$, where T is the capacity functional of $\mathbf{X}$. Thus the empty space function F gives us the values of the capacity functional $T(K)$ for all discs K. This does not fully determine T, and hence does not fully characterise $\mathbf{X}$. However, F gives us a lot of qualitative information about $\mathbf{X}$. The empty space function is a simple property of the point process that is useful in data analysis.

1.10 Estimation from Data

In applications, spatial point pattern data usually take the form of a finite configuration of points $\mathbf{x} = \{x_1, \ldots, x_n\}$ in a region (**window**) W, where $x_i \in W$ and where $n = n(\mathbf{x}) \geq 0$ is not fixed. The data would often be treated as a realisation of a stationary point process $\mathbf{X}$ inside W. It is then important to estimate properties of the process $\mathbf{X}$.

An unbiased estimator of F is

$$\widehat{F}(r) = \frac{1}{\lambda_d(W)} \int_W \mathbf{1}\{\mathsf{dist}(u, \mathbf{X}) \leq r\} \, \mathrm{d}u. \tag{3}$$

This is an unbiased estimator of $F(r)$, for each fixed value of r, since

$$\begin{aligned}
\mathbb{E}\left[\widehat{F}(r)\right] &= \frac{1}{\lambda_d(W)} \mathbb{E}\left[\int_W \mathbf{1}\{\mathsf{dist}(u, \mathbf{X}) \leq r\} \, \mathrm{d}u\right] \\
&= \frac{1}{\lambda_d(W)} \int_W \mathbb{E}\mathbf{1}\{\mathsf{dist}(u, \mathbf{X}) \leq r\} \, \mathrm{d}u \\
&= \frac{1}{\lambda_d(W)} \int_W \mathbb{P}(\mathsf{dist}(u, \mathbf{X}) \leq r) \, \mathrm{d}u \\
&= \frac{1}{\lambda_d(W)} \int_W F(r) \, \mathrm{d}u \\
&= F(r)
\end{aligned}$$

where the penultimate line follows by the stationarity of $\mathbf{X}$.

A practical problem is that, if we only observe $\mathbf{X} \cap W$, the integrand in (3) is not observable. When u is a point close to the boundary of the window W, the point of $\mathbf{X}$ nearest to u may lie outside W. More precisely, we have $\mathsf{dist}(u, \mathbf{X}) \leq r$ if and only if $n(\mathbf{X} \cap b(u, r)) > 0$. But our data are a realisation of $\mathbf{X} \cap W$, so we can only evaluate $n(\mathbf{X} \cap W \cap b(u, r))$.

It was once a common mistake to ignore this, and simply to replace $\mathbf{X}$ by $\mathbf{X} \cap W$ in (3). But this results in a negatively biased estimator of F. Call the estimator $\widehat{F}_W(r)$. Since $n(\mathbf{X} \cap W \cap b(u, r)) \leq n(\mathbf{X} \cap b(u, r))$, we have

$$\mathbf{1}\{n(\mathbf{X} \cap W \cap b(u, r)) > 0\} \leq \mathbf{1}\{n(\mathbf{X} \cap b(u, r)) > 0\}$$

so that $\mathbb{E}\widehat{F}_W(r) \leq F(r)$. This is called a **bias due to edge effects**.

One simple strategy for eliminating the edge effect bias is the **border method**. When estimating $F(r)$, we replace W in equation (3) by the erosion

$$W_{-r} = W \ominus b(0, r) = \{x \in W : \, \mathsf{dist}(x, \partial W) \geq r\}$$

consisting of all points of W that are at least r units away from the boundary ∂W. Clearly, $u \in W_{-r}$ if and only if $b(u, r) \subset W$. Thus, $n(\mathbf{x} \cap b(u, r))$ is observable when $u \in W_{-r}$. Thus we estimate $F(r)$ by

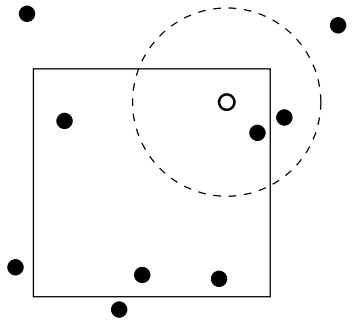

Fig. 22. Edge effect problem for estimation of the empty space function F. If we can only observe the points of $\mathbf{X}$ inside a window W (bold rectangle), then for some reference points u in W (open circle) it cannot be determined whether there is a point of $\mathbf{X}$ within a distance r of u. This problem occurs if u is closer than distance r to the boundary of W.

$$\widehat{F}_b(r) = \frac{1}{\lambda_2(W_{-r})} \int_{W_{-r}} \mathbf{1}\{\mathsf{dist}(u, \mathbf{x}) \le r\} \, \mathrm{d}u. \tag{4}$$

This is observable, and by the previous argument, it is an unbiased estimator of $F(r)$.

For a survey of corrections for edge effects, see [2].

1.11 Computer Exercises

Software is available for generating simulated realisations of point processes as shown above. The user needs access to the statistical package R, which can be downloaded free from the R website [13] and is very easy to install. Introductions to R are available at [23, 38].

We have written a library `spatstat` in the R language for performing point pattern data analysis and simulation. See [8] for an introduction. The `spatstat` library should also be downloaded from the R website [13], and installed in R.

The following commands in R will then generate and plot simulations of the point processes shown in Figures 9, 11, 12, 16, 19 and 20 above.

```
library(spatstat)
X <- runifpoint(100)
plot(X)
X <- rpoispp(5)
plot(X)
X <- rpoispp(function(x, y) { exp( 2 + 5 * x) })
plot(X)
plot(rMaternI(200, 0.05))
plot(rMaternII(200, 0.05))
```

```
plot(rMatClust(5,  0.07, 20))
plot(rMatClust(50, 0.07, 2))
X <- rpoispp(100)
M <- sample(1:3, X$n, replace=TRUE)
plot(X %mark% M)
M <- rexp(X$n)
plot(X %mark% M)
```

Further information on each *command* can be obtained by typing `help(`*command*`)` in R.

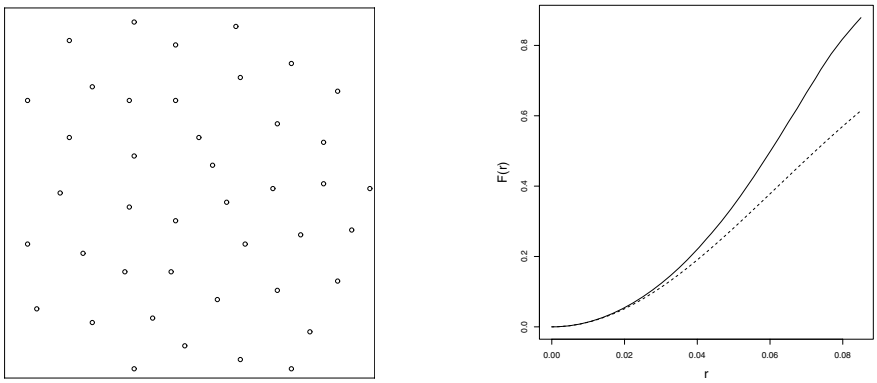

Fig. 23. *Left:* the `cells` point pattern dataset. *Right:* estimated empty space function $F(r)$ plotted against r (solid lines) together with the empty space function of a Poisson process (dotted lines).

The `spatstat` library also contains point pattern datasets and techniques for analysing them. In particular the function `Fest` will estimate the contact distribution function or empty space function F (defined in Section 1.9) from an observed realisation of a stationary point process. The following commands access the `cells` point pattern dataset, plot the data, then compute an estimate of F and plot this function.

```
data(cells)
plot(cells)
Fc <- Fest(cells)
plot(Fc)
```

The resulting plots are shown in Figure 23. There is a striking discrepancy between the estimated function F and the function expected for a Poisson process, indicating that the data cannot be treated as Poisson.

2 Moments and Summary Statistics

In this lecture we describe the analogue, for point processes, of the moments (expected value, variance and higher moments) of a random variable. These quantities are useful in theoretical study of point processes and in statistical inference about point patterns.

The **intensity** or first moment of a point process is the analogue of the expected value of a random variable. **Campbell's formula** is an important result for the intensity. The 'second moment measure' is related to the variance or covariance of random variables. The **K function** and **pair correlation** are derived second-moment properties which have many applications in the statistical analysis of spatial point patterns [16, 43]. The second-moment properties of some point processes will be found here. In the computer exercises we will compute statistical estimates of the K function from spatial point pattern data sets.

2.1 Intensity

Definition 2.1. *Let* $\mathbf{X}$ *be a point process on* $S = \mathbb{R}^d$ *(or on any locally compact metric space* S*). Writing*

$$\nu(B) = \mathbb{E}[N_X(B)], \qquad B \subset S,$$

defines a measure ν *on* S*, called the* **intensity measure** *of* $\mathbf{X}$*, provided* $\nu(B) < \infty$ *for all compact* B*.*

Example 2.1 (Binomial process). The binomial point process (Section 1.3) of n points in a region $W \subset \mathbb{R}^d$ has $N(B) \sim \mathsf{binomial}(n, p)$ where $p = \lambda_d(B \cap W)/\lambda_d(W)$ so

$$\nu(B) = \mathbb{E}N(B) = np = n\frac{\lambda_d(B \cap W)}{\lambda_d(W)}.$$

Thus $\nu(B)$ is proportional to the volume of $B \cap W$.

Example 2.2 (Poisson process). The uniform Poisson process of intensity $\beta > 0$ has $N(B) \sim \mathsf{Poisson}(\beta\lambda_d(B))$ so

$$\nu(B) = \beta\lambda_d(B).$$

Thus $\nu(B)$ is proportional to the volume of B.

Example 2.3 (Translated grid). Suppose U_1, U_2 are independent random variables uniformly distributed in $[0, s]$. Let $\mathbf{X}$ be the point process consisting of all points with coordinates $(U_1 + ms, U_2 + ns)$ for all integers m, n. A realisation of this process is a square grid of points in $\mathbb{R}^2$, with grid spacing s, which has been randomly translated. See Figure 24. It is easy to show that

$$\nu(B) = \mathbb{E}N(B) = \frac{1}{s^2}\,\lambda_2(B)$$

for any set B in $\mathbb{R}^2$ of finite area. This principle is important in applications to **stereology** [4].

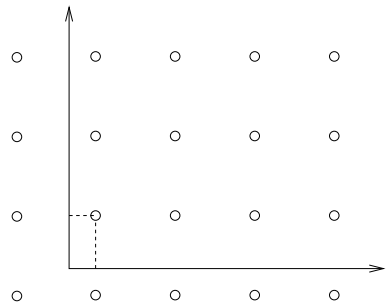

Fig. 24. A randomly translated square grid.

If $\mathbf{X}$ is a stationary point process in $\mathbb{R}^d$, then

$$\nu(B + v) = \mathbb{E}N(B + v) = \mathbb{E}N(B) = \nu(B)$$

for all $v \in \mathbb{R}^d$. That is, the intensity measure of a stationary point process is invariant under translations. But we know that the only such measures are multiples of Lebesgue measure:

Theorem 2.1. *If ν is a translation-invariant measure on $\mathbb{R}^d$ then $\nu(B) = c\lambda_d(B)$ for some $c \geq 0$.*

Corollary 1 *If $\mathbf{X}$ is a stationary point process in $\mathbb{R}^d$, then its intensity measure ν is a constant multiple of Lebesgue measure λ_d.*

The constant c in Corollary 1 is often called the **intensity** of $\mathbf{X}$.

Definition 2.2. *Suppose the intensity measure ν of a point process $\mathbf{X}$ in $\mathbb{R}^d$ satisfies*

$$\nu(B) = \int_B \beta(u)\,\mathrm{d}u$$

for some function β. Then we call β the **intensity function** *of $\mathbf{X}$.*

If it exists, the intensity function has the interpretation that in a small region $\mathrm{d}x \subset \mathbb{R}^d$

$$\mathbb{P}(N(\mathrm{d}x) > 0) \sim \mathbb{E}N(\mathrm{d}x) \sim \beta(x)\,\mathrm{d}x.$$

For the uniform Poisson process with intensity $\beta > 0$, the intensity function is obviously $\beta(u) \equiv \beta$. The randomly translated square grid (Example 2.3) is a stationary process with intensity measure $\nu(B) = \beta\lambda_2(B)$, so it has an intensity function, $\beta(u) \equiv 1/s^2$.

Theorem 2.2 (Campbell's Formula). *Let* $\mathbf{X}$ *be a point process on* S *and let* $f : S \to \mathbb{R}$ *be a measurable function. Then the random sum*

$$T = \sum_{x \in \mathbf{X}} f(x)$$

is a random variable, with expected value

$$\mathbb{E}\left[\sum_{x \in \mathbf{X}} f(x)\right] = \int_S f(x)\, \nu(\mathrm{d}x). \tag{5}$$

In the special case where $\mathbf{X}$ is a point process on $\mathbb{R}^d$ with an intensity function β, Campbell's Formula becomes

$$\mathbb{E}\left[\sum_{x \in \mathbf{X}} f(x)\right] = \int_{\mathbb{R}^d} f(x)\beta(x)\, \mathrm{d}x.$$

Campbell's Formula applies even to non-simple point processes (i.e. where points may have a multiplicity greater than 1) if the terms in the sum in (5) are counted with their multiplicity.

Proof. The result (5) is true when f is a step function, i.e. a function of the form

$$f = \sum_{i=1}^{m} c_i 1_{B_i}$$

for $B_i \subset S$ compact and $c_i \in \mathbb{R}$, because in that case

$$T = \sum_{x \in \mathbf{X}} f(x) = \sum_x \sum_i c_i 1_{B_i}(x) = \sum_i c_i N_{\mathbf{X}}(B_i)$$

so

$$\mathbb{E}T = \mathbb{E}\left[\sum_i c_i N_{\mathbf{X}}(B_i)\right] = \sum_i c_i \mathbb{E}N(B_i) = \sum_i c_i \nu(B_i) = \int_S f(x)\, \nu(\mathrm{d}x).$$

The result for general f follows by monotone approximation. □

Example 2.4 (Monte Carlo integration). Suppose we want to compute the integral

$$I = \int_W f(x)\, \mathrm{d}x$$

where $W \subset \mathbb{R}^d$ and f is a nonnegative, integrable, real-valued function. Take any point process $\mathbf{X}$ with intensity

$$\lambda(x) = \begin{cases} c \text{ if } x \in W \\ 0 \text{ if } x \notin W \end{cases}$$

Evaluate the function f at the random points of $\mathbf{X}$, and estimate the integral I by the discrete sum approximation

$$\widehat{I} = \frac{1}{c} \sum_{x \in \mathbf{X}} f(x).$$

Then Campbell's formula (5) gives

$$\mathbb{E}[\widehat{I}] = \frac{1}{c} \mathbb{E}\left[\sum_{x \in \mathbf{X}} f(x) \right] = \frac{1}{c} \int_{\mathbb{R}^d} f(x) \lambda(x) \, \mathrm{d}x = \int_W f(x) \, \mathrm{d}x = I$$

so that $\widehat{I}$ is an *unbiased estimator* of I.

Example 2.5 (Olbers' Paradox). In 1826, the astronomer Heinrich Olbers pointed out a physical paradox in the fact that the sky is dark at night. Suppose we make the following assumptions: (i) the universe exists in 3-dimensional Euclidean space $\mathbb{R}^3$; (ii) the stars currently visible from Earth (with a given absolute magnitude) constitute a stationary point process in $\mathbb{R}^3$; and (iii) the observed brilliance of the light reaching Earth from a star at location $x \in \mathbb{R}^3$ is $a/||x||^2$ where a is constant (the inverse square law).

Then the expected total brilliance of the night sky is infinite:

$$\mathbb{E}\left[\sum_{x \in \mathbf{X} \setminus \mathrm{Earth}} \frac{a}{||x||^2} \right] = \lambda \int_{\mathbb{R}^3 \setminus \mathrm{Earth}} \frac{a}{||x||^2} \, \mathrm{d}x = \infty.$$

By this argument, 19th century physicists realized that, in a stable, infinite universe with an even distribution of stars, the entire universe should gradually heat up. The paradox led to a review of the theory of thermodynamics.

Example 2.6. Suppose $\mathbf{X}$ consists of a fixed, finite number of random points in $\mathbb{R}^d$, say $\mathbf{X} = \{X_1, \ldots, X_n\}$. Assume X_i has a marginal probability density $f_i(u), u \in \mathbb{R}^d$. Then $\mathbf{X}$ has intensity function $\beta(u) = \sum_{i=1}^n f_i(u)$.

Example 2.7. Consider a Poisson cluster process $\mathbf{Y}$ (Section 1.7). This is formed by taking a uniform Poisson process $\mathbf{X}$ of parent points, with intensity α, and replacing each $x \in \mathbf{X}$ by a random cluster Z_x which is a finite point process.

Suppose Z_x has intensity function $f(u \mid x)$. Then *conditional on* $\mathbf{X}$, the process $\mathbf{Y}$ has intensity function

$$\beta_{\mathbf{Y}|\mathbf{X}}(u) = \sum_{x \in \mathbf{X}} f(u \mid x)$$

It is not hard to show that the (unconditional) intensity function β of $\mathbf{Y}$ is the expectation with respect to $\mathbf{X}$,

$$\begin{aligned}\beta(u) &= \mathbb{E}\left[\beta_{\mathbf{Y}|\mathbf{X}}(u)\right] \\ &= \mathbb{E}\sum_{x\in\mathbf{X}} f(u \mid x) \\ &= \alpha \int_{\mathbb{R}^d} f(u \mid x)\, \mathrm{d}x\end{aligned}$$

by Campbell's formula.

For example, in Matérn's cluster process, a cluster Z_x consists of a $\mathsf{Poisson}(\mu)$ random number of points, uniformly distributed in the disc $b(x,r)$ of radius r centred on x. This has intensity $f(u \mid x) = \mu/(\pi r^2)$ if $u \in b(x,r)$ and 0 otherwise. Now

$$\begin{aligned}\int_{\mathbb{R}^d} f(u \mid x)\, \mathrm{d}x &= \frac{\mu}{\pi r^2}\int_{\mathbb{R}^d} \mathbf{1}\{u \in b(x,r)\}\, \mathrm{d}x \\ &= \frac{\mu}{\pi r^2}\int_{\mathbb{R}^d} \mathbf{1}\{x \in b(u,r)\}\, \mathrm{d}x \\ &= \mu.\end{aligned}$$

Hence Matérn's cluster process has intensity $\beta(u) = \alpha\mu$.

2.2 Intensity for Marked Point Processes

Marked point processes were introduced in Section 1.8. Let $\mathbf{Y}$ be a marked point process on the space S with marks in a space M. Viewing $\mathbf{Y}$ as a point process on $S \times M$, we may extend the definition of intensity measure to marked point processes without further work.

The intensity measure of $\mathbf{Y}$ is (by Definition 2.1) a measure ν on $S \times M$ defined by

$$\nu(U) = \mathbb{E}N_{\mathbf{Y}}(U), \qquad U \subset S \times M.$$

It is completely determined by the values

$$\begin{aligned}\nu(B \times C) &= \mathbb{E}N_{\mathbf{Y}}(B \times C) \\ &= \mathbb{E}\sum_{(x,m)\in\mathbf{Y}} \mathbf{1}\{x \in B\}\mathbf{1}\{m \in C\}\end{aligned}$$

for all compact $B \subset S$ and measurable $C \subset M$.

For marked point processes, Campbell's Formula takes the form

$$\mathbb{E}\sum_{(x,m)\in\mathbf{Y}} f(x,m) = \int_{S\times M} f(x,m)\, \nu(\mathrm{d}x, \mathrm{d}m) \tag{6}$$

where $f : S \times M \to \mathbb{R}$ is a measurable function.

Differences between marked and unmarked point processes arise with regard to the concept of stationarity.

Definition 2.3. *A marked point process on $\mathbb{R}^d$ with marks in M is* **stationary** *if its distribution is invariant under shifts of $\mathbb{R}^d$* only

$$(x, m) \mapsto (x + v, m)$$

for all $v \in \mathbb{R}^d$.

Note that the shift operation changes the location of a point but does not alter the mark attached to it.

Theorem 2.3. *Let Y be a stationary marked point process in $\mathbb{R}^d$. Assume the corresponding process of unmarked points has finite intensity (that is $\mathbb{E}N_Y(K \times M) < \infty$ for all compact $K \subset \mathbb{R}^d$).*

Then the intensity measure ν of Y takes the form

$$\nu(A \times B) = \beta \lambda_d(A) Q(B) \tag{7}$$

for all $A \subset \mathbb{R}^d$, $B \subset M$, where $\beta \geq 0$ is the intensity (expected number of points per unit volume), and Q is a probability measure on M called the **distribution of the typical mark**.

As a simple example of (7), consider a point process consisting of points of three colours. This may be formalised as a marked point process in $\mathbb{R}^2$ in which the marks are colours, $M = \{\text{red}, \text{green}, \text{blue}\}$. For a region $A \subset \mathbb{R}^2$, the quantity $\nu(A \times \{\text{red}\})$ is the expected number of red points in A, and by equation (7), this is equal to $\beta\lambda_2(A)Q(\{\text{red}\})$, a constant times the area of A times the probability of the colour red.

Proof. Since Y is stationary, ν is invariant under shifts of $\mathbb{R}^d$,

$$\nu(A \times B) = \nu((A + v) \times B)$$

for all $A \subset \mathbb{R}^d$, $B \subset M$ and all translation vectors $v \in \mathbb{R}^d$. If we fix B and define

$$\mu_B(A) = \nu(A \times B)$$

for all $A \subset \mathbb{R}^d$, then μ_B is a measure on $\mathbb{R}^d$ which is invariant under translations. It follows from Theorem 2.1 that, for fixed B,

$$\nu(A \times B) = c_B \, \lambda_d(A)$$

for all $A \subset \mathbb{R}^d$, where c_B is a constant depending on B.

On the other hand, if we fix A to be the unit cube, and define $\kappa(B) = \nu(A \times B) = c_B\lambda_d(A) = c_B$, then κ is a measure on M satisfying $\kappa(M) = \nu(A \times M) = \mathbb{E}N_Y(A \times M) < \infty$ by assumption. Letting $\beta = \kappa(M)$ and $Q(B) = \kappa(B)/\beta$ yields the result. □

The argument we have just seen is often called **factorisation** or **disintegration**. It exploits the property that the intensity measure is invariant with respect to translations on the first factor of the product $\mathbb{R}^d \times M$. We shall have occasion to use the same argument many times.

For a stationary marked point process, Campbell's formula becomes

$$\mathbb{E}\left[\sum_{(x,m)\in\mathbf{Y}} f(x,m)\right] = \beta\mathbb{E}_Q\left[\int_{\mathbb{R}^d} f(x,K)\right] \tag{8}$$

where K denotes a random mark (a random element of M) with distribution Q. As an exercise, the reader may like to use this to prove Olbers' Paradox (Example 2.5) in greater generality, treating the stars in the universe as a stationary marked point process in $\mathbb{R}^3$, with the marks indicating the absolute brightness of each star.

2.3 Second Moment Measures

Let $\mathbf{X}$ be a point process. We are interested in the variance of the count $N(B)$,

$$\mathsf{var}\, N(B) = \mathbb{E}\left[N(B)^2\right] - [\mathbb{E}N(B)]^2$$

and the covariance of two such counts,

$$\mathsf{cov}[N(B_1), N(B_2)] = \mathbb{E}\left[N(B_1)N(B_2)\right] - [\mathbb{E}N(B_1)]\,[\mathbb{E}N(B_2)].$$

A key observation is that $N(B_1)N(B_2)$ is equal to the number of ordered pairs (x, x') of points in the process $\mathbf{X}$ such that $x \in B_1$ and $x' \in B_2$.

Definition 2.4. *Let $\mathbf{X}$ be a point process on a space S. Then $\mathbf{X}\times\mathbf{X}$ is a point process on $S\times S$ consisting of all ordered pairs (x,x') of points $x, x' \in \mathbf{X}$. The intensity measure ν_2 of $\mathbf{X}\times\mathbf{X}$ is a measure on $S\times S$ satisfying*

$$\nu_2(A\times B) = \mathbb{E}\left[N_{\mathbf{X}}(A)N_{\mathbf{X}}(B)\right].$$

This measure ν_2 is called the **second moment measure** *of $\mathbf{X}$.*

Clearly, the second moment measure contains all information about the variances and covariances of the variables $N_{\mathbf{X}}(A)$. Campbell's formula applied to $\mathbf{X}\times\mathbf{X}$ becomes

$$\mathbb{E}\left[\sum_{x\in\mathbf{X}}\sum_{y\in\mathbf{X}} f(x,y)\right] = \int_S\int_S f(x,y)\;\nu_2(\mathrm{d}x,\,\mathrm{d}y)$$

for a measurable function $f : S\times S \to \mathbb{R}$.

Example 2.8. For the uniform Poisson point process of intensity $\beta > 0$ in $\mathbb{R}^d$, the second moment measure satisfies

$$\nu_2(A \times B) = \beta^2 \lambda_d(A)\, \lambda_d(B) + \beta \lambda_d(A \cap B).$$

Geometrically this means that the measure ν_2 consists of two components: there is a constant density β^2 on all of $\mathbb{R}^d \times \mathbb{R}^d$, plus a positive mass on the diagonal $\Delta = \{(x, x) : x \in \mathbb{R}^d\}$. The mass on the diagonal arises from the fact that $\mathbf{X} \times \mathbf{X}$ includes pairs (x, x) of identical points. We could write the second moment measure informally as

$$\nu_2(\mathrm{d}x,\, \mathrm{d}y) = \beta^2\, \mathrm{d}x\, \mathrm{d}y + \beta.\delta(x - y)\, \mathrm{d}x$$

where δ is the delta function. More formally

$$\nu_2 = \beta^2 \lambda_d \otimes \lambda_d + \beta \mathsf{diag}^{-1} \lambda_d$$

where $\mathsf{diag}(x, x) = x$.

To remove the mass on the diagonal, and also to simplify the calculation of certain moments, we introduce the **second factorial moment measure**

$$\nu_{[2]}(A \times B) = \mathbb{E}[N(A)N(B)] - \mathbb{E}[N(A \cap B)].$$

This is the intensity measure of the process $\mathbf{X} * \mathbf{X}$ of all ordered pairs of *distinct* points of $\mathbf{X}$. It satisfies

$$\mathbb{E}\left[\sum_{x \in \mathbf{X}} \sum_{y \in \mathbf{X},\ y \neq x} f(x, y)\right] = \int_S \int_S f(x, y)\, \nu_{[2]}(\mathrm{d}x,\, \mathrm{d}y).$$

The name 'factorial' is derived from

$$\begin{aligned} \nu_{[2]}(A \times A) &= \mathbb{E}\left[N(A)^2\right] - \mathbb{E}[N(A)] \\ &= \mathbb{E}\left[N(A)[N(A) - 1]\right]. \end{aligned}$$

For example, for the uniform Poisson process of intensity β, the second factorial moment measure is $\nu_{[2]} = \beta^2 \lambda_d \otimes \lambda_d$.

Definition 2.5. *A point process* $\mathbf{X}$ *on* $\mathbb{R}^d$ *is said to have* **second moment density** g_2 *if*

$$\nu_{[2]}(C) = \int_C g_2(x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{9}$$

for any compact $C \subset \mathbb{R}^d \times \mathbb{R}^d$.

Informally, $g_2(x, y)$ gives the joint probability that there will be points of $\mathbf{X}$ at two specified locations x and y:

$$\mathbb{P}(N(\,\mathrm{d}x) > 0,\ N(\,\mathrm{d}y) > 0) \sim g_2(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

For example, the uniform Poisson process has second moment density $g_2(x,y) = \beta^2$. The binomial process of n points in W has

$$g_2(x,y) = \frac{n(n-1)}{\lambda_d(W)^2}$$

if $x, y \in W$, and zero otherwise.

Definition 2.6. *Suppose* $\mathbf{X}$ *is a point process on* $\mathbb{R}^d$ *which has an intensity function* $\beta(x)$ *and a second moment density* $g_2(x,y)$*. Then we define the* **pair correlation function** *of* $\mathbf{X}$ *by*

$$\rho_2(x,y) = \frac{g_2(x,y)}{\beta(x)\beta(y)}.$$

Example 2.9. For a uniform Poisson process of intensity β, we have $\beta(x) \equiv \beta$ and $g_2 \equiv \beta^2$, so that

$$\rho_2(x,y) \equiv 1.$$

Example 2.10. For a binomial process of n points in a region W, we have

$$\rho_2(x,y) \equiv 1 - \frac{1}{n}.$$

Note that the pair correlation function always satisfies $\rho_2(x,y) \geq 0$. It should be regarded as a 'non-centred' analogue of the usual correlation of random variables. The value $\rho_2 = 1$ corresponds to a lack of correlation in the usual statistical sense: if $\rho_2 \equiv 1$ then $\mathsf{cov}[N(B), N(B')] = 0$ for disjoint sets B, B'.

Example 2.11. Continuing Example 2.6, suppose $\mathbf{X}$ consists of a fixed, finite number of random points in $\mathbb{R}^d$, say $\mathbf{X} = \{X_1, \ldots, X_n\}$. Let $f_i(u), u \in \mathbb{R}^d$ be the marginal probability density of X_i, and $f_{ij}(u,v), u, v \in \mathbb{R}^d$ the marginal joint density of (X_i, X_j). Then $\mathbf{X}$ has second moment density

$$g_2(x,y) = \sum_{i \neq j} f_{ij}(x,y)$$

and pair correlation function

$$\rho_2(x,y) = \frac{\sum_{i \neq j} f_{ij}(x,y)}{\left(\sum_i f_i(x)\right)\left(\sum_j f_j(y)\right)}.$$

Example 2.12. Continuing Example 2.7, consider a Poisson cluster process $\mathbf{Y}$, formed from a Poisson process $\mathbf{X}$ of parent points with intensity α. The clusters Z_x for different x are independent processes.

Suppose Z_u has intensity function $f(u \mid x)$ and second moment density $h(u, v \mid x)$. It is not hard to show, by first conditioning on X, that the second moment density of Y is

$$g_2(u, v) = \beta(u)\beta(v) + \alpha \int_{\mathbb{R}^d} h(u, v \mid x)\, \mathrm{d}x,$$

where $\beta(u) = \int_{\mathbb{R}^d} f(u \mid x)\, \mathrm{d}x$ is the intensity of $\mathbf{Y}$. The integral term arises from pairs of points in $\mathbf{Y}$ that come from the same cluster Z_x.

For example, in the Matérn cluster process, the second moment density of a cluster Z_x is (by a simple extension of Example 2.11) $h(u, v \mid x) = \mu^2/(\pi^2 r^4)$ if $u, v \in b(x, r)$, and 0 otherwise. We have

$$\begin{aligned}\int_{\mathbb{R}^d} \mathbf{1}\{u, v \in b(x, r)\}\, \mathrm{d}x &= \int_{\mathbb{R}^d} \mathbf{1}\{u \in b(x, r)\}\mathbf{1}\{v \in b(x, r)\}\, \mathrm{d}x \\ &= \int_{\mathbb{R}^d} \mathbf{1}\{x \in b(u, r)\}\mathbf{1}\{x \in b(v, r)\}\, \mathrm{d}x \\ &= \lambda_2(b(u, r) \cap b(v, r)).\end{aligned}$$

Hence the second moment density of the Matérn cluster process is

$$g_2(u, v) = \alpha^2\mu^2 + \alpha \frac{\mu^2}{\pi^2 r^4} \lambda_2(b(u, r) \cap b(v, r)).$$

2.4 Second Moments for Stationary Processes

For a *stationary* point process in $\mathbb{R}^d$, there is a 'disintegration' of the second moment measure. Stationarity implies

$$\mathbb{E}\,[N(A + v)N(B + v)] = \mathbb{E}\,[N(A)N(B)]$$

for all $v \in \mathbb{R}^d$. Thus $\nu_2, \nu_{[2]}$ are invariant under **simultaneous shifts**

$$(x, y) \mapsto (x + v, y + v).$$

See the left panel in Figure 25.

Let us transform the problem by mapping each pair of points (x, y) to the pair $(x, y - x)$. Thus the first element of the image is the first point x, and the second element $y - x$ is the vector from x to y. This transforms $\mathbb{R}^d \times \mathbb{R}^d$ onto itself by $\Psi(x, y) = (x, y - x)$. Under this transformation, the simultaneous shift $(x, y) \mapsto (x + v, y + v)$ becomes a shift of the first coordinate

$$(s, t) \mapsto (s + v, t).$$

See the right panel in Figure 25.

The image of $\nu_{[2]}$ under Ψ is a measure μ on $\mathbb{R}^d$ which is invariant under translations of the first coordinate

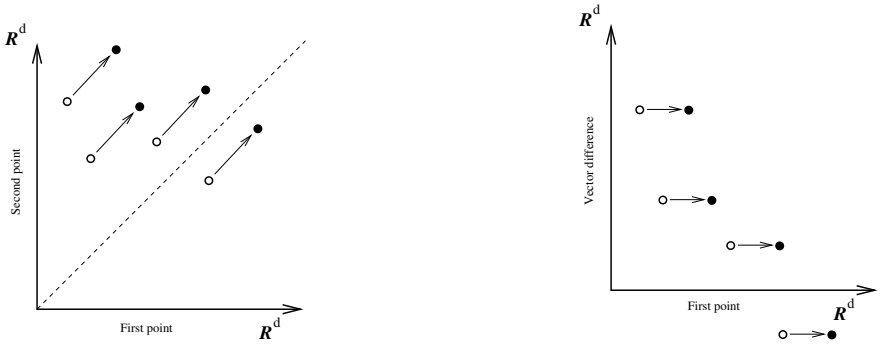

Fig. 25. Disintegration of the second moment measure of a stationary point process. *Left:* The second moment measure is invariant under shifts parallel to the diagonal in $\mathbb{R}^d \times \mathbb{R}^d$. *Right:* If we transform the problem by mapping (x, y) to $(x, y - x)$, the image of the second moment measure is invariant under shifts of the first coordinate. This factorises as a product measure.

$$(s, t) \mapsto (s + v, t)$$

for all $v \in \mathbb{R}^d$. By Theorem 2.1 it follows that

$$\mu = \beta \lambda_d \otimes \mathcal{K}$$

where β is the intensity of the process $\mathbf{X}$, and $\mathcal{K}$ is a measure on $\mathbb{R}^d$ called the **reduced second moment measure** of $\mathbf{X}$.

Retracing our steps and using Campbell's Formula, we find that for an arbitrary integrand f,

$$\begin{aligned}
\mathbb{E}\left[\sum_{x\in\mathbf{X}} \sum_{y\in\mathbf{X},\, y\neq x} f(x,y)\right] &= \int\int f(x,y)\; \nu_{[2]}(\mathrm{d}x,\, \mathrm{d}y) \\
&= \int\int f(x, x+u)\; \mu(\mathrm{d}x,\, \mathrm{d}u) \\
&= \beta \int\int f(x, x+u)\; \mathcal{K}(\mathrm{d}u)\; \mathrm{d}x.
\end{aligned}$$

Theorem 2.4. *Let* $\mathbf{X}$ *be a stationary point process on* $\mathbb{R}^d$ *with intensity* β. *Then there is a measure* $\mathcal{K}$ *on* $\mathbb{R}^d$ *such that, for a general integrand* f,

$$\mathbb{E}\left[\sum_{x\in\mathbf{X}} \sum_{y\in\mathbf{X},\, y\neq x} f(x,y)\right] = \beta \int\int f(x, x+u)\; \mathcal{K}(\mathrm{d}u)\; \mathrm{d}x. \tag{10}$$

$\mathcal{K}$ *is called the* **reduced second moment measure** *of* $\mathbf{X}$.

To understand the measure $\mathcal{K}$, we notice that for $A, B \subset \mathbb{R}^d$

$$\begin{aligned}\beta\lambda_d(A)\mathcal{K}(B) &= \mu(A \times B)\\ &= \int\int \mathbf{1}\{s \in A\}\mathbf{1}\{t \in B\}\mu(s,t)\\ &= \int\int \mathbf{1}\{x \in A\}\mathbf{1}\{y - x \in B\}\ \nu_{[2]}(\mathrm{d}x,\, \mathrm{d}y)\\ &= \mathbb{E}\left[\sum_{x\in\mathbf{X}}\sum_{y\in\mathbf{X},\ y\neq x} \mathbf{1}\{x \in A\}\mathbf{1}\{y - x \in B\}\right]\end{aligned}$$

This may also be obtained directly from (10) by taking $f(x,y) = \mathbf{1}\{x \in A\}\mathbf{1}\{y - x \in B\}$. Since $\beta\lambda_d(A) = \mathbb{E}N(A)$, we have

$$\mathcal{K}(B) = \frac{\mathbb{E}\sum_{x\in\mathbf{X}\cap A} N((B+x)\setminus x)}{\mathbb{E}N(A)} \tag{11}$$

The right hand side of (11) may be interpreted as the average, over all points x of the process, of the number of other points y of the process such that $y - x \in B$.

Example 2.13. For the uniform Poisson process,

$$\begin{aligned}\nu_{[2]} &= \beta^2\ \lambda_d \otimes \lambda_d\\ \mu &= \beta^2\ \lambda_d \otimes \lambda_d\\ \mathcal{K} &= \beta\ \lambda_d\end{aligned}$$

Example 2.14. Suppose $\mathbf{X}$ is a stationary process on $\mathbb{R}^d$ which has a second moment density function g_2. Then by comparing (9) with (10) we can see that $g_2(x,y)$ depends only on $y - x$, say

$$g_2(x,y) = g(y - x),$$

for some function g, and we can write

$$\mathcal{K}(B) = \frac{1}{\beta}\int_B g(u)\ \mathrm{d}u.$$

Example 2.15. The randomly translated square grid was introduced in Example 2.3. This is a stationary process. Following through the derivation above, we find that the reduced second moment measure $\mathcal{K}$ puts mass 1 at each integer point (ns, ms) for all integers n, m, except that there is no atom at $(0,0)$.

Intuitively this reflects the fact that, if we know there is a point of $\mathbf{X}$ at the origin, then this determines the position of the entire grid of points, and we know there will be a point of $\mathbf{X}$ at each location (ns, ms).

This point process does not have a second moment density g_2.

2.5 The K-function

Second moment properties are important in the statistical analysis of spatial point pattern data, just as the sample variance is important in classical statistics.

The reduced second moment measure $\mathcal{K}$ carries important information about the dependence or **interaction** between different points of the process. For practical data analysis, we need some simplification of the measure $\mathcal{K}$. Ripley [39] suggested the function

$$K(t) = \frac{1}{\beta}\mathcal{K}(b(0,t)), \quad t \geq 0. \tag{12}$$

See also Ornstein & Zernike [35].

Using (11) with $B = b(0,t)$, we see that $\beta K(t)$ is the expected number of points y of the process that satisfy $0 < ||y - x|| \leq t$ for a given point x of the process. In other words, $\beta K(t)$ is the expected number of points close to a given point of the process, where 'close' means 'within a distance t'.

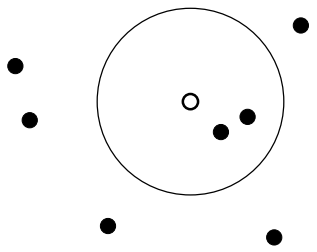

Fig. 26. Concept of the K-function. The value $\beta K(t)$ is the expected number of other points within a circle of radius t centred on a typical point of the process.

Example 2.16. For a uniform Poisson process in $\mathbb{R}^d$,

$$K(t) = \kappa_d t^d, \quad t \geq 0$$

where κ_d is the volume of the unit ball in $\mathbb{R}^d$.

The factor $1/\beta$ in (12) normalises the K-function, making it independent of the intensity β in the Poisson case.

Example 2.17. For a stationary point process in $\mathbb{R}^d$ which has a second moment density, Example 2.14 gives

$$K(t) = \frac{1}{\beta^2}\int_{b(0,t)} g_2(0,x)\,\mathrm{d}x = \int_{b(0,t)} \rho_2(0,x)\,\mathrm{d}x.$$

Lemma 2.1. *Suppose* $\mathbf{X}$ *is a stationary and isotropic point process in* $\mathbb{R}^2$ *which possesses a second moment density* g_2 *and pair correlation function* ρ. *Then* $g_2(x,y)$ *and* $\rho(x,y)$ *depend only on* $||y-x||$, *say*

$$g_2(x,y) = g(||x-y||) \tag{13}$$
$$\rho_2(x,y) = \rho(||x-y||) \tag{14}$$

and the pair correlation can be recovered from the K*-function by*

$$\rho_2(t) = \frac{\frac{\mathrm{d}}{\mathrm{d}t}K(t)}{2\pi t} \tag{15}$$

Example 2.18. For the uniformly randomly translated grid (Examples 2.3 and 2.15) the K-function is $K(t) = M(t/s) - 1$, where $M(r)$ is the number of points of the integer grid $\mathbb{Z}^2$ inside the disc $b(0,r)$. The function M is studied closely in Prof. Baranyi's lectures in this volume.

Lemma 2.2 (Invariance of K **under thinning).** *Suppose* $\mathbf{X}$ *is a stationary point process, and* $\mathbf{Y}$ *is obtained from* $\mathbf{X}$ *by random thinning (each point of* X *is deleted or retained, independently of other points, with retention probability* p*). Then the* K*-functions of* $\mathbf{X}$ *and* $\mathbf{Y}$ *are identical.*

The proof is an exercise.

2.6 Estimation from Data

Assume again that we have observed data in the form of a finite configuration of points $\mathbf{x} = \{x_1, \ldots, x_n\}$ in a window W, where $x_i \in W$ and where $n = n(\mathbf{x}) \geq 0$ is not fixed.

In order to estimate the K-function, consider the identity

$$K(t) = \frac{\mathbb{E}\sum_{x \in \mathbf{X} \cap W} N(b(x,t) \setminus x)}{\beta\, \mathbb{E}N(W)}. \tag{16}$$

Again we have an edge effect problem in applying this identity. If we only observe $\mathbf{X} \cap W$, the random variable in the numerator of (16) is not observable. When x is a point close to the boundary of the window W, the disc $b(x,t)$ may extend outside W. Since the process $\mathbf{X}$ is not observed outside W, the number of points of $\mathbf{X}$ in $b(x,t)$ is not observable.

It is a common mistake to ignore this problem, and estimate the numerator of (16) by

$$\sum_{i=1}^{n} n(\mathbf{x} \cap b(x_i,t) \setminus x_i) = \sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1}\{||x_i - x_j|| \leq t\}. \tag{17}$$

The right hand side of (17) is proportional to the empirical distribution function of the distances $s_{ij} = ||x_i - x_j||$ between all pairs of points. But this is a

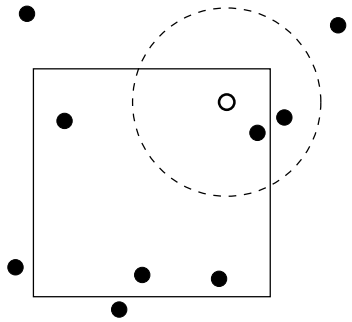

Fig. 27. Edge effect problem for estimation of the K function. If we can only observe the points inside a window W (bold rectangle), then the number of points inside a circle of radius t, centred on a point of the process inside W, is not observable if the circle extends outside W.

biased estimator: the expectation of (17) is less than the numerator of (16), because the observable quantity $n(\mathbf{X} \cap W \cap b(u,t))$ is less than or equal to the desired quantity $n(\mathbf{X} \cap b(u,t))$. This is a bias due to edge effects.

One simple strategy for eliminating the edge effect bias is the border method, introduced in Section 1.10. When estimating $K(t)$, we replace W in equation (16) by the erosion

$$W_{-t} = W \ominus b(0,t) = \{x \in W : \ \mathsf{dist}(x, \partial W) \geq t\}$$

consisting of all points of W that are at least t units away from the boundary ∂W. Clearly, $u \in W_{-t}$ if and only if $b(u,t) \subset W$. Thus, $n(\mathbf{x} \cap b(x_i,t) \setminus x_i)$ is observable when $x_i \in W_{-t}$. Thus we estimate $K(t)$ by

$$\begin{aligned} \widehat{K}(t) &= \frac{\sum_{x \in W_{-t}} N_{\mathbf{X}}(b(x,t) \setminus x)}{\widehat{\beta} n(\mathbf{x} \cap W_{-t})} \\ &= \frac{\sum_{i=1}^{n} \sum_{j \neq i} \mathbf{1}\{||x_i - x_j|| \leq t\}}{\widehat{\beta} n(\mathbf{x} \cap W_{-t})} \end{aligned} \tag{18}$$

where $\widehat{\beta}$ is usually $n(\mathbf{x})/\lambda_2(W)$. This is called the **border method** of edge correction. More sophisticated edge corrections with better performance are discussed in [42, 2].

2.7 Exercises

We again make use of the package `spatstat` described in section 1.11.

The function `Kest` computes estimates of the K function from spatial point pattern data.

```
library(spatstat)
data(cells)
```

```
Kc <- Kest(cells)
plot(Kc)
data(redwood)
plot(Kest(redwood))
```

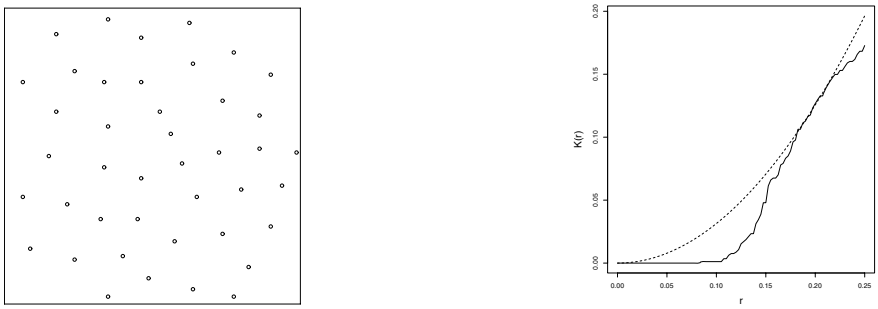

Fig. 28. *Left:* the `cells` point pattern dataset. *Right:* estimated K function plotted against r, together with the theoretical K function for a Poisson process with the same (estimated) intensity.

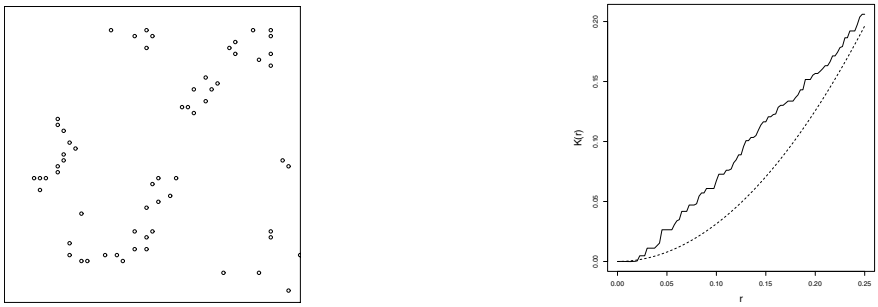

Fig. 29. *Left:* the `redwood` point pattern dataset. *Right:* estimated K plotted against r, together with the empty space function of a Poisson process.

The function `Kmeasure` computes an estimate of (a kernel-smoothed density of) the reduced second moment measure $\mathcal{K}$.

```
KMc <- Kmeasure(cells, sigma=0.03)
plot(KMc)
```

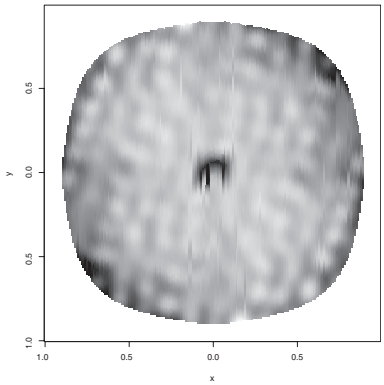

Fig. 30. Kernel-smoothed density estimate of the second moment measure $\mathcal{K}$ of the `cells` dataset. Lighter greys indicate higher estimated densities.

3 Conditioning

In the study of a point process we are often interested in properties relating to a **typical point** of the process. This requires the calculation of conditional probabilities of events given that there is a point of the process at a specified location. It leads to the concept of the **Palm distribution** of the point process, and the related **Campbell-Mecke formula** [42]. These tools allow us to define new characteristics of a point process, such as the nearest neighbour distance distribution function G. A dual concept is the **conditional intensity** which provides many new results about point processes. In the computer exercises we compute statistical estimates of the function G from spatial point pattern data sets.

3.1 Motivation

One simple question about a point process $\mathbf{X}$ is: what is the probability distribution of the distance from a point of $\mathbf{X}$ to its nearest neighbour (the nearest other point of $\mathbf{X}$)?

Note that this is different from the empty space function F introduced in Section 1.9, which is the distribution of the distance $\mathsf{dist}(u, \mathbf{X})$ from a *fixed location* u to the nearest point of $\mathbf{X}$. Here we are asking about the distance from a *point of the process* $\mathbf{X}$ to the nearest other point of the process.

If x is known to be a point of $\mathbf{X}$, then the nearest neighbour distance is $R_x = \mathsf{dist}(x, \mathbf{X} \setminus x)$, and we seek the 'conditional probability'

$$\mathbb{P}\left(R_x \le r \mid x \in \mathbf{X}\right).$$

The problem is that this is not a conditional probability in the elementary sense, because the event $\{x \in \mathbf{X}\}$ typically has probability zero.

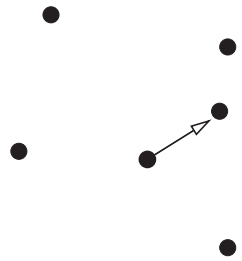

Fig. 31. Concept of nearest neighbour distance.

For some basic examples of point processes, this question can be resolved using classical methods.

Example 3.1. Consider the binomial process (Section 1.3)

$$\mathbf{X} = \{X_1, \ldots, X_n\}$$

where $X_1, \ldots, X_n$ are i.i.d. random points, uniformly distributed in $W \subset \mathbb{R}^2$. For each $i = 1, \ldots, n$ the conditional probability

$$\mathbb{P}\left(R_x \leq r \mid X_i = x\right) = \mathbb{P}\left(\mathsf{dist}(x, \mathbf{X} \setminus X_i) \leq r \mid X_i = x\right)$$

is well-defined (using the classical definition of conditional probability) and equal to

$$\begin{aligned}\mathbb{P}\left(R_x \leq r \mid X_i = x\right) &= 1 - \mathbb{P}\left(R_x > r \mid X_i = x\right) \\ &= 1 - \mathbb{P}(\mathbf{X}' \cap b(x, r) = \emptyset)\end{aligned}$$

where

$$\mathbf{X}' = \mathbf{X} \setminus X_i$$

is a binomial process with $n - 1$ points. Thus

$$\mathbb{P}\left(R_x \leq r \mid X_i = x\right) = 1 - \left[\frac{\lambda_2(b(x, r) \cap W)}{\lambda_2(W)}\right]^{n-1}.$$

The same quantity is obtained for each i, as we might have expected given the exchangeability of $X_1, \ldots, X_n$. Hence it seems reasonable to interpret this to be the value of $\mathbb{P}\left(R_x \leq r \mid x \in \mathbf{X}\right)$.

A similar argument can be used for other point processes which contain a finite number of points, almost surely.

For a stationary point process $\mathbf{X}$, another argument must be used. It is sufficient to consider $x = 0$, that is, to condition on the event that there is a point of $\mathbf{X}$ at the origin 0. One simple way to define and calculate such probabilities would be to condition on the event that there is a point of $\mathbf{X}$ in a small neighbourhood U of the origin 0, and then take the limit as U shrinks down to $\{0\}$.

Example 3.2. Suppose $\mathbf{X}$ is a Poisson process in $\mathbb{R}^2$ with intensity β. For $\epsilon > 0$, let $U = b(0, \epsilon)$ and define $R_{(\epsilon)} = \mathsf{dist}(0, \mathbf{X} \setminus U)$, the distance from 0 to the nearest point of X outside U. Clearly, $R_{(\epsilon)} > r$ iff $\mathbf{X}$ has no points in $b(0, r) \setminus U$. Since U and $b(0, r) \setminus U$ are disjoint,

$$\mathbb{P}\left(R_{(\epsilon)} > r \mid N(U) > 0\right) = \mathbb{P}(R_{(\epsilon)} > r) = \exp\{-\beta\pi(r^2 - \epsilon^2)\}.$$

As $\epsilon \downarrow 0$, this conditional probability converges to the limit $\exp\{-\beta\pi r^2\}$. Note also that

$$\mathbb{P}\left(N(U) = 1 \mid N(U) > 0\right) \to 1 \quad \text{as } \epsilon \downarrow 0$$

so that, for small ϵ, we may effectively assume there is at most one point in U. Additionally, if $\mathbf{X} \cap U = \{x\}$, then

$$\left|\mathsf{dist}(x, \mathbf{X} \setminus x) - \mathsf{dist}(0, \mathbf{X} \setminus U)\right| \leq \epsilon$$

so we have some confidence in formally writing

$$\mathbb{P}\left(R_0 \leq r \mid 0 \in \mathbf{X}\right) = \exp\{-\beta\pi r^2\}.$$

3.2 Palm Distribution

The **Palm distribution** formalises the concept of conditioning on a point of the process. It was developed by C. Palm (1907-1951) for the study of telephone traffic [36].

Palm Probabilities

The Palm probability $\mathbb{P}^x(A)$ of an event A at a location x is, intuitively speaking, the conditional probability that the event A will occur, given $x \in \mathbf{X}$: that is, given that there is a point of the process X at the specified location x.

An elegant way to define $\mathbb{P}^x(A)$ is the following. Let $(\Omega, \mathcal{A}, P)$ be the underlying probability space. Define the Campbell measure C on $S \times \Omega$ by

$$C(B \times A) = \mathbb{E}\left[N(B)\mathbf{1}_A\right]$$

for all $A \in \mathcal{A}$ and $B \in \mathcal{B}(S)$, then by extension to $\mathcal{A} \otimes \mathcal{B}\,(S)$. Here $\mathbf{1}_A$ is the indicator random variable of the event A (equal to 1 if the event A occurs and 0 if not) and $\mathcal{B}(S)$ is the Borel σ-field of S. Notice that

$$C(B \times A) \leq \mathbb{E}N(B) = \nu(B)$$

where ν is the intensity measure of X (assumed to exist and to be locally finite).

For any fixed A, let $\mu_A(B) = C(B \times A)$ for all B. Then μ_A is a measure, and $\mu_A \leq \nu$, so certainly $\mu_A << \nu$. By the Radon-Nikodým Theorem,

$$\mu_A(B) = \int_B f_A(x)\, \nu(\mathrm{d}x)$$

where $f_A : S \to \mathbb{R}_+$ is measurable (and unique up to equality almost everywhere). We shall interpret $f_A(x)$ as the Palm probability $\mathbb{P}^x(A)$.

Under conditions on $(\Omega, \mathcal{A})$, there exist **regular conditional probabilities** $\mathbb{P}^x(A)$ such that

- for all A, the function $x \mapsto \mathbb{P}^x(A)$ is a version of f_A, i.e.

$$\int_B \mathbb{P}^x(A)\, \nu(\mathrm{d}x) = C(B \times A) = \mathbb{E}\left[N(B)1_A\right]$$

- for almost all x, the map $A \mapsto \mathbb{P}^x(A)$ is a probability measure on $(\Omega, \mathcal{A})$.

Then $\mathbb{P}^x$ is called the **Palm probability measure** associated with the point process $\mathbf{X}$ at the location x. We write $\mathbb{E}^x$ for the expectation with respect to $\mathbb{P}^x$.

Example 3.3 (Poisson process). Let $\mathbf{X}$ be a uniform Poisson process in $\mathbb{R}^d$. Consider the event

$$A = \{N(K) = 0\}$$

where $K \subset \mathbb{R}^d$ is compact. For any closed U disjoint from K we have, by independence properties of the Poisson process,

$$C(U \times A) = \mathbb{E}\left[N(U)1_A\right] = \mathbb{E}[N(U)]\mathbb{P}(A) = \nu(U)\mathbb{P}(A).$$

It follows that $\mathbb{P}^x(A) = \mathbb{P}(A)$ for almost all $x \in \mathbb{R}^d \setminus K$. On the other hand, for $U \subseteq K$ we have $N(U) \leq N(K)$ so that

$$C(U \times A) = \mathbb{E}\left[N(U)1_A\right] = 0$$

so that $\mathbb{P}^x(A) = 0$ for almost all $x \in K$.

Now holding x fixed and varying K, and taking the complementary probabilities, we have $\mathbb{P}^x(N(K) > 0) = \mathbb{P}(N(K) > 0)$ if $x \notin K$, and $\mathbb{P}^x(N(K) > 0) = 1$ if $x \in K$. But this is the capacity functional of

$$\mathbf{X} \cup \{x\},$$

the Poisson process $\mathbf{X}$ augmented by a fixed point at the location x.

In other words, under the Palm distribution $\mathbb{P}^x$, the process behaves as if it were a Poisson process superimposed with a fixed point at the location x.

Note that $\mathbb{P}^x$ is a probability measure on the original space $(\Omega, \mathcal{A})$, giving a probability $\mathbb{P}^x(A)$ for any event $A \in \mathcal{A}$, and not just for events defined by the process $\mathbf{X}$.

Example 3.4 (Mixed Poisson process). Suppose Γ is a nonnegative real random variable defined on Ω and that, given $\Gamma = \gamma$, the point process $\mathbf{X}$ is Poisson with intensity γ. The intensity measure of this mixture process is

$$\mathbb{E}N(B) = \mathbb{E}\left[\mathbb{E}[N(B) \mid \Gamma]\right] = \mathbb{E}[\Gamma]\,\lambda_d(B).$$

Let $A = \{\Gamma \le \gamma\}$ for some fixed $\gamma \ge 0$. Then for $B \subset \mathbb{R}^d$ we have

$$\begin{aligned} C(B \times A) &= \mathbb{E}\left[N(B)1_A\right] \\ &= \mathbb{E}\left(\mathbb{E}\left[N(B)1_A \mid \Gamma\right]\right) \\ &= \mathbb{E}\left[\Gamma\lambda_d(B)\mathbf{1}\{\Gamma \le \gamma\}\right] \\ &= \mathbb{E}\left[\Gamma\mathbf{1}\{\Gamma \le \gamma\}\right]\lambda_d(B) \end{aligned}$$

so that

$$\mathbb{P}^x(\Gamma \le \gamma) = \frac{\mathbb{E}[\Gamma\mathbf{1}\{\{\Gamma \le \gamma\}\}]}{\mathbb{E}[\Gamma]}.$$

Thus, the distribution of Γ under $\mathbb{P}^x$ is the Γ-weighted counterpart of its original distribution.

Palm Distribution of Point Process

Many writers consider only the Palm distribution of the point process $\mathbf{X}$ itself, that is, the distribution $\mathbf{P}^x$ on N defined by

$$\mathbf{P}^x(A) = \mathbb{P}^x(\mathbf{X} \in A)$$

for $A \in \mathcal{N}$. Note the distinction between $\mathbf{P}^x$, a point process distribution in the sense of Definition 1.5, and $\mathbb{P}^x$, a probability measure on the original probability space Ω. We sometimes denote the Palm distribution of $\mathbf{X}$ by $\mathbf{P}^x_{\mathbf{X}}$.

When $\mathbf{X}$ is a homogeneous Poisson process, we have just shown in Example 3.3 that the Palm distribution satisfies

$$\mathbf{P}^x = \mathbf{P} * \Delta_x$$

where $\mathbf{P}$ is the distribution of the original Poisson process, $*$ denotes convolution (superposition of two point processes), and Δ_x is the distribution of the point process consisting of a single point at x.

We sometimes write $\mathbf{X}^x$ for the process governed by the Palm distribution $\mathbf{P}^x$, so that the last equation can be expressed as

$$\mathbf{X}^x \stackrel{\mathrm{d}}{=} \mathbf{X} \cup \{x\}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equivalence in distribution. In fact, this property is characteristic of Poisson processes.

Theorem 3.1 (Slivnyak's Theorem). *Let* $\mathbf{X}$ *be a point process with locally finite intensity measure* ν*. Suppose the distribution* $\mathbf{P} = \mathbf{P}_{\mathbf{X}}$ *and the Palm distribution* $\mathbf{P}^x = \mathbf{P}^x_{\mathbf{X}}$ *of* $\mathbf{X}$ *are related by*

$$\mathbf{P}^x = \mathbf{P} * \Delta_x.$$

Then $\mathbf{X}$ *is a Poisson process with intensity measure* ν*.*

It is often convenient to remove the point x from consideration.

Definition 3.1. *The* **reduced Palm distribution** $\mathbf{P}^{!x}$ *of a point process* $\mathbf{X}$ *is the distribution of* $\mathbf{X} \setminus x$ *under* $\mathbf{P}^x$*:*

$$\mathbf{P}^{!x}(A) = \mathbf{P}^x(\mathbf{X} \setminus x \in A)$$

for $A \in \mathcal{N}$*.*

Thus Slivnyak's Theorem states that $\mathbf{X}$ is a Poisson point process if and only if $\mathbf{P}^{!x}_{\mathbf{X}} = \mathbf{P}_{\mathbf{X}}$.

Example 3.5 (Binomial process). Let $\mathbf{Y}_{(n)}$ be the binomial process (Section 1.3) consisting of n independent random points $X_1, \ldots, X_n$ uniformly distributed in a domain W. It is easy to show that the reduced Palm distribution of $\mathbf{Y}^{(n)}$ is identical to the distribution of $\mathbf{Y}^{(n-1)}$.

Example 3.6 (Palm distribution of mixed Poisson process). Let $\mathbf{X}$ be the mixed Poisson process described in Example 3.4. Consider the event

$$A = \{N(K) = 0\}$$

where $K \subset \mathbb{R}^d$ is compact. Following the argument in Example 3.3 we find that if $x \in K$ then $\mathbb{P}^x(A) = 0$, while if $x \notin K$, then

$$\begin{aligned}
\mathbb{P}^x(A) &= \frac{\mathbb{E}[\Gamma \mathbf{1}\{N(K) = 0\}]}{\mathbb{E}[\Gamma]} \\
&= \frac{\mathbb{E}[\Gamma \mathbb{P}(N(K) = 0 \mid \Gamma)]}{\mathbb{E}[\Gamma]} \\
&= \frac{\mathbb{E}[\Gamma \exp(-\Gamma \lambda_d(K))]}{\mathbb{E}[\Gamma]}.
\end{aligned}$$

Hence, the capacity functional of $\mathbf{X}^{!x}$ is the Γ-weighted mean of the capacity functional of a Poisson process with intensity Γ. This is different from the capacity functional of the original process $\mathbf{X}$, which is the unweighted mean $T(K) = \mathbb{E}\left[\exp(-\Gamma \lambda_d(K))\right]$.

The distribution $\mathbf{P}$ and reduced Palm distribution $\mathbf{P}^{!x}$ of $\mathbf{X}$ satisfy, for all events $A \in \mathcal{N}$,

$$\mathbf{P}(A) = \mathbb{E}[\pi_{\Gamma}(A)] \tag{19}$$

$$\mathbf{P}^{!x}(A) = \frac{\mathbb{E}[\Gamma\, \pi_{\Gamma}(A)]}{\mathbb{E}[\Gamma]} \tag{20}$$

where π_γ denotes the distribution of the uniform Poisson process with intensity γ.

To put it another way, let $F(t) = \mathbb{P}(\Gamma \le t)$ be the cumulative distribution function of the random intensity of the original process. Define the weighted c.d.f.

$$F_1(t) = \frac{1}{\mathbb{E}[\Gamma]} \int_0^t s\, F(\mathrm{d}s).$$

Then $\mathbf{X}^{!x}$ is a mixed Poisson process whose random intensity has the weighted c.d.f. F_1.

An intuitive explanation for the last example is the following. Points of the process are generated with greater intensity when Γ is larger. Hence, by Bayes' Theorem, given that a point was observed to occur, the posterior probability distribution of Γ favours larger values of Γ.

Theorem 3.2 (Campbell-Mecke formula). *For any function $Y : S \times \Omega \mapsto \mathbb{R}_+$ that is integrable with respect to the Campbell measure,*

$$\mathbb{E} \sum_{x \in \mathbf{X}} Y(x) = \int_S \mathbb{E}^x[Y(x)]\, \nu(\mathrm{d}x) \tag{21}$$

In particular, if $Y(x) = f(x, \mathbf{X})$, that is, $Y(x, \omega) = f(x, \mathbf{X}(\omega))$, we get

$$\mathbb{E}\left[\sum_{x \in \mathbf{X}} f(x, \mathbf{X})\right] = \int_S \mathbb{E}^x[f(x, \mathbf{X})]\, \nu(\mathrm{d}x). \tag{22}$$

Example 3.7 (dependent thinning). We shall determine the intensity of Matérn's Model I, which was described in Section 1.7. Let $\mathbf{X}$ denote the original Poisson process, of intensity β in $\mathbb{R}^2$, and $\mathbf{Y}$ the thinned process obtained by deleting any point $x \in \mathbf{X}$ such that $\mathsf{dist}(x, \mathbf{X} \setminus x) \le r$, that is, deleting any point which has a neighbour closer than r units. For any $B \subset \mathbb{R}^d$ let

$$\begin{aligned} f(x, \mathbf{X}) &= \mathbf{1}\{x \in B\}\mathbf{1}\{\mathsf{dist}(x, \mathbf{X} \setminus x) \le r\} \\ &= \mathbf{1}\{x \in B\}\mathbf{1}\{x \in \mathbf{Y}\}. \end{aligned}$$

Since $\mathbf{X}$ is a Poisson process we have

$$\mathbb{P}^x(\mathsf{dist}(x, \mathbf{X} \setminus x) \le r) = \mathbb{P}(\mathsf{dist}(x, \mathbf{X}) \le r) = 1 - \exp\{-\beta\pi r^2\}.$$

Hence

$$\begin{aligned}\mathbb{E}\,[n(\mathbf{Y}\cap B)] &= \mathbb{E}\sum_{x\in\mathbf{X}} f(x,\mathbf{X})\\ &= \beta\int_{\mathbb{R}^2}\mathbb{E}^x[f(x,\mathbf{X})]\,\mathrm{d}x\\ &= \beta\lambda_2(B)(1-\exp\{-\beta\pi r^2\}).\end{aligned}$$

It follows that $\mathbf{Y}$ has intensity $\beta(1-\exp\{-\beta\pi r^2\})$.

Example 3.8 (Boundary length of Boolean model). Consider the union

$$Z = \bigcup_{x\in\mathbf{X}} b(x,r)$$

where $\mathbf{X}$ is a homogeneous Poisson process of intensity β in $\mathbb{R}^2$, and $r>0$ is fixed. We want to find

$$\mathbb{E}\,[\mathsf{length}(W\cap\partial Z)]$$

where ∂ denotes boundary. Write

$$\mathsf{length}(W\cap\partial Z) = \sum_{x\in\mathbf{X}} Y(x)$$

where

$$Y(x) = \mathsf{length}(W\cap\partial b(x,r)\setminus Z_{-x})$$

and

$$Z_{-x} = \bigcup_{y\in\mathbf{X}\setminus\{x\}} b(x,r).$$

Under the Palm probability measure $\mathbb{P}^x$, this random set Z_{-x} is a Boolean model with the same distribution as Z. Hence

$$\mathbb{E}^x[Y(x)] = (1-p)\,\mathsf{length}(W\cap\partial b(x,r))$$

where $p = 1-\exp(-\beta\pi r^2)$ is the coverage probability of Z. Hence by Campbell-Mecke

$$\begin{aligned}\mathbb{E}\,[\mathsf{length}(W\cap\partial Z)] &= \int_{\mathbb{R}^2}(1-p)\,\mathsf{length}(W\cap\partial b(x,r))\beta\,\mathrm{d}x\\ &= 2\pi\beta r\exp(-\beta\pi r^2)\lambda_2(W).\end{aligned}$$

3.3 Palm Distribution for Stationary Processes

In the case of a stationary point process, the Palm distributions $\mathbf{P}^x$ at different locations x are equivalent under translation.

Lemma 3.1. *If* $\mathbf{X}$ *is a stationary point process in* $\mathbb{R}^d$*, then*

$$\mathbf{X}^x \stackrel{d}{=} \mathbf{X}^0 + x$$

where $\mathbf{X}^x$ *again denotes a process governed by the Palm probability measure* $\mathbb{P}^x$*.*

More formally, let T_x denote the effect of translation by a vector $x \in \mathbb{R}^d$ on a counting measure N,

$$T_x N(B) = N(B - x), \qquad B \subset \mathbb{R}^d$$

and correspondingly for events $E \in \mathcal{N}$

$$T_x E = \{N \in \mathsf{N} : \ T_x N \in E\}$$

and for any point process distribution $\mathbf{Q}$ define $T_x\mathbf{Q}$ by

$$T_x\mathbf{Q}(E) = Q(T_x E), \qquad E \in \mathcal{N}.$$

Then Lemma 3.1 states that

$$\mathbf{P}^x_{\mathbf{X}} = T_x \, \mathbf{P}^0_{\mathbf{X}} \tag{23}$$

for any stationary point process $\mathbf{X}$.

Proof. Apply the Campbell-Mecke formula to functions of the form

$$f(x, \mathbf{X}) = \mathbf{1}\{x \in B\}\mathbf{1}\{\mathbf{X} - x \in A\}$$

where $A \in \mathcal{N}$ is an event, $B \subset \mathbb{R}^d$, and $\mathbf{X} - x = \mathbf{X} + (-x)$ is the result of shifting $\mathbf{X}$ by the vector $-x$. This yields

$$\mathbb{E}\left[\sum_{x \in \mathbf{X} \cap B} \mathbf{1}\{\mathbf{X} - x \in A\}\right] = \beta \int_B \mathbb{P}^x(\mathbf{X} - x \in A) \, \mathrm{d}x.$$

Since $\mathbf{X}$ is stationary, $\mathbf{X}$ has the same distribution as $\mathbf{X} + v$ for any vector v, so

$$\begin{aligned}\mathbb{E}\left[\sum_{x \in \mathbf{X} \cap B} \mathbf{1}\{\mathbf{X} - x \in A\}\right] &= \mathbb{E}\left[\sum_{x \in (\mathbf{X}+v) \cap B} \mathbf{1}\{(\mathbf{X} + v) - x \in A\}\right] \\ &= \mathbb{E}\left[\sum_{x \in \mathbf{X} \cap T_{-v}B} \mathbf{1}\{\mathbf{X} - x \in A\}\right].\end{aligned}$$

Thus

$$\beta \int_B \mathbb{P}^x(\mathbf{X} - x \in A) \, \mathrm{d}x = \beta \int_{T_{-v}B} \mathbb{P}^x(\mathbf{X} - x \in A) \, \mathrm{d}x$$

which implies that for all B

$$\int_B \mathbb{P}^x(\mathbf{X} - x \in A)\, \mathrm{d}x = c\, \lambda_2(B)$$

for some constant c, and hence that $\mathbb{P}^x(\mathbf{X} - x \in A)$ is constant. This proves (23). □

One way to interpret this result is to construct a marked point process $\mathbf{Y}$ on $\mathbb{R}^d$ with marks in N by attaching to each point $x \in \mathbf{X}$ the mark $\mathbf{X} - x$. That is, the mark attached to the point x is a copy of the entire realisation of the point process, translated so that x is shifted to the origin. The result shows that $\mathbf{Y}$ is a *stationary* marked point process. Hence the intensity measure of $\mathbf{Y}$ factorises,

$$\nu(B \times A) = \beta\, \lambda_d(B)\, Q(A)$$

for $B \subset \mathbb{R}^d$, $A \in \mathcal{N}$ where Q is the **mark distribution**. Clearly Q can be interpreted as the Palm distribution given there is a point at 0. That is, $\mathbf{P}^0 = Q$, and $\mathbf{P}^x = T_{-x}Q$.

This gives us a direct interpretation of the Palm distribution $\mathbf{P}^0$ (but not the Palm probability measure $\mathbb{P}^0$) for a stationary point process in $\mathbb{R}^d$. We have

$$\mathbb{E}\left[\sum_{x \in B} \mathbf{1}\{\mathbf{X} - x \in A\}\right] = \beta\, \lambda_d(B)\, \mathbf{P}^0(A) \tag{24}$$

for $B \subset \mathbb{R}^d$, $A \in \mathcal{N}$. Thus

$$\mathbf{P}^0(A) = \frac{\mathbb{E}\left[\sum_{x \in B} \mathbf{1}\{\mathbf{X} - x \in A\}\right]}{\mathbb{E}N(B)} \tag{25}$$

for all $B \subset \mathbb{R}^d$ such that $0 < \lambda_d(B) < \infty$. On the right side of (25), the denominator is the expected number of terms in the numerator, so we can interpret $\mathbf{P}^0(A)$ as the 'average' fraction of points x satisfying $\mathbf{X} - x \in A$.

3.4 Nearest Neighbour Function

In Definition 1.9 we defined the empty space function F of a stationary point process. The function F can be estimated from data, and provides a simple summary of the process. It can be useful in statistical analysis of point patterns.

A related concept is the nearest neighbour distance distribution.

Definition 3.2. *Let* $\mathbf{X}$ *be a stationary point process in* $\mathbb{R}^d$. *The* **nearest neighbour function** G *is the cumulative distribution function of the distance*

$$R' = \mathsf{dist}(x, \mathbf{X} \setminus x)$$

from a typical point $x \in \mathbf{X}$ *to the nearest other point of* $\mathbf{X}$. *That is*

$$\begin{aligned} G(r) &= \mathbb{P}^x(\mathsf{dist}(x, \mathbf{X} \setminus x) \le r) \\ &= \mathbb{P}^x(N(b(x,r) \setminus x) > 0). \end{aligned}$$

By stationarity, this does not depend on x.

Example 3.9. For a stationary Poisson process in $\mathbb{R}^d$, since $\mathbf{X}^x \equiv \mathbf{X} \cup \{x\}$, we have

$$\begin{aligned} G(r) &= \mathbb{P}^x(\mathsf{dist}(x, \mathbf{X} \setminus x) \le r) \\ &= \mathbb{P}(\mathsf{dist}(x, \mathbf{X}) \le r) \\ &= 1 - \exp(-\beta \kappa_d r^d). \end{aligned}$$

In this case $G(r) \equiv F(r)$.

Estimation of the function G from observed point pattern data is hampered by edge effects, similar to those affecting the estimation of F (see Section 1.10). The basic identity for estimation is (25), or more specifically

$$G(r) = \frac{\mathbb{E}\left[\sum_{x \in B} \mathbf{1}\{\mathsf{dist}(x, \mathbf{X} \setminus x) \le r\}\right]}{\mathbb{E}N(B)}. \tag{26}$$

The simplest strategy for avoiding edge effects is an adaptation of the border method. When estimating $G(r)$ we set $B = W_{-r}$ in equation (26) so that the quantities on the right hand side of (26) are observable. This yields an estimator

$$\widehat{G}_b(r) = \frac{\sum_{x \in W_{-r}} \mathbf{1}\{\mathsf{dist}(x, \mathbf{x} \setminus x) \le r\}}{n(\mathbf{x} \cap W_{-r})}. \tag{27}$$

If we write for each $x_i \in \mathbf{x}$

$$\begin{aligned} d_i &= \mathsf{dist}(x_i, \mathbf{x} \setminus x_i) \\ b_i &= \mathsf{dist}(x_i, \partial W) \end{aligned}$$

so that d_i is the observed nearest-neighbour distance and b_i is the distance to the boundary of the observation window, then the estimator can be rewritten

$$\widehat{G}_b(r) = \frac{\sum_i \mathbf{1}\{d_i \le r,\ b_i \ge r\}}{\sum_i \mathbf{1}\{b_i \ge r\}}. \tag{28}$$

Further discussion of edge corrections can be found in [42, 2].

3.5 Conditional Intensity

Return for a moment to the heuristic definition of the Palm probability $\mathbb{P}^x(A)$ as the limit of $\mathbb{P}(A \mid N(U) > 0)$ as $U \downarrow \{x\}$, where U is an open neighbourhood of x in $\mathbb{R}^d$, and A is an event in $\mathcal{A}$. Applying Bayes' Theorem

$$\mathbb{P}\left(N(U) > 0 \mid A\right) = \frac{\mathbb{P}(N(U) > 0)}{\mathbb{P}(A)} \mathbb{P}\left(A \mid N(U) > 0\right)$$

so that, as $U \downarrow \{x\}$,

$$\frac{\mathbb{P}\left(N(U) > 0 \mid A\right)}{\mathbb{P}(N(U) > 0)} \to \frac{\mathbb{P}^x(A)}{\mathbb{P}(A)}.$$

Suppose $\mathbf{X}$ has an intensity function $\beta(u)$, $u \in \mathbb{R}^d$ which is continuous at x. Then asymptotically

$$\mathbb{P}(N(U) > 0) \sim \mathbb{E}[N(U)] = \int_U \beta(u)\, \mathrm{d}u \sim \beta(x)\lambda_d(U)$$

so that

$$\frac{\mathbb{P}\left(N(U) > 0 \mid A\right)}{\lambda_d(U)} \to \beta(x) \frac{\mathbb{P}^x(A)}{\mathbb{P}(A)}. \tag{29}$$

Since we can also write

$$\frac{\mathbb{P}(N(U) > 0)}{\lambda_d(U)} \to \beta(x),$$

then (29) can be interpreted as a conditional analogue of the intensity $\beta(x)$ given the event A.

This motivates the following definitions.

Definition 3.3. *Let* $\mathbf{X}$ *be a point process on a space* S*, The* **reduced Campbell measure** *of* $\mathbf{X}$ *is the measure* $C^!$ *on* $S \times \mathsf{N}$ *such that*

$$C^![B \times A] = \mathbb{E}\left[\sum_{x \in \mathbf{X}} \mathbf{1}\{x \in B\}\mathbf{1}\{\mathbf{X} \setminus x \in A\}\right]$$

for $B \subset \mathbb{R}^d$ *and* $A \in \mathcal{N}$*.*

Definition 3.4. *Let* $\mathbf{X}$ *be a point process on* $\mathbb{R}^d$*, and suppose its reduced Campbell measure* $C^!$ *is absolutely continuous with respect to* $\lambda_d \otimes \mathbf{P}$ *(where* $\mathbf{P}$ *is the distribution of* $\mathbf{X}$*).*

Then the Radon-Nikodým derivative $\beta^* : \mathbb{R}^d \times \mathcal{N} \to \mathbb{R}_+$ *of* $C^!$ *with respect to* $\lambda_d \otimes \mathbf{P}$ *is called the* **conditional intensity** *of* $\mathbf{X}$*. It is defined to satisfy*

$$C^![B \times A] = \int_B \mathbb{E}\left[\beta^*(u, \mathbf{X})\mathbf{1}\{\mathbf{X} \in A\}\right]\, \mathrm{d}u \tag{30}$$

for $B \subset \mathbb{R}^d$ *and* $A \in \mathcal{N}$*.*

If $\mathbf{X}$ has a conditional intensity then, by extension of the last equation, for any integrable $g : \mathbb{R}^d \times \mathsf{N} \to \mathbb{R}_+$,

$$\mathbb{E}\left[\sum_{x\in\mathbf{X}} g(x, \mathbf{X}\setminus x)\right] = \int_{\mathbb{R}^d} \mathbb{E}\left[\beta^*(x,\mathbf{X})g(x,\mathbf{X})\right]\,\mathrm{d}x. \tag{31}$$

If $\mathbf{X}$ has an intensity function $\beta(u)$, $u \in \mathbb{R}^d$, then the Campbell-Mecke formula gives

$$\mathbb{E}\left[\sum_{x\in\mathbf{X}} g(x, \mathbf{X}\setminus x)\right] = \int_{\mathbb{R}^d} \mathbb{E}^{!x}[g(x,\mathbf{X})]\beta(x)\,\mathrm{d}x \tag{32}$$

writing $\mathbb{E}^{!x}$ for the expectation with respect to $\mathbf{P}^{!x}$. Comparing the right sides of (31) and (32) shows that, for almost all $x \in \mathbb{R}^d$,

$$\mathbb{E}^{!x}[g(x,\mathbf{X})] = \mathbb{E}\left[\frac{\beta^*(x,\mathbf{X})}{\beta(x)} g(x,\mathbf{X})\right]. \tag{33}$$

Thus, $\beta^*(x,\mathbf{X})/\beta(x)$ is the Radon-Nikodým density of $\mathbf{P}^{!x}$ with respect to $\mathbf{P}$. In particular taking $g \equiv 1$

$$\beta(x) = \mathbb{E}[\beta^*(x, X)]. \tag{34}$$

Example 3.10. If $\mathbf{X}$ is a Poisson process on $\mathbb{R}^d$ with intensity function $\beta(x)$, then we have $\mathbf{P}^{!x} = \mathbf{P}$ for all x, so $\beta^*(x,\mathbf{X})/\beta(x)$ is identically equal to 1, and the conditional intensity is $\beta^*(x,\mathbf{X}) = \beta(x)$.

Our heuristic argument says that

$$\beta^*(x,\mathbf{X})\,\mathrm{d}x = \mathbb{P}\left(N(\,\mathrm{d}x) > 0 \mid \mathbf{X}\setminus x\right);$$

that is, roughly speaking, $\beta^*(x,\mathbf{X})\,\mathrm{d}x$ is the conditional probability that there will be a point of $\mathbf{X}$ in an infinitesimal neighbourhood of x, given the location of all points of $\mathbf{X}$ outside this neighbourhood.

Example 3.11. The binomial process $\mathbf{Y}^{(n)}$ consists of n independent random points, uniformly distributed in a domain W. We saw above that the reduced Palm distribution of $\mathbf{Y}^{(n)}$ is identical to the distribution of $\mathbf{Y}^{(n-1)}$. In this case, $\mathbf{P}^{!x}$ and $\mathbf{P}$ are mutually singular (since, for example, the event that there are exactly n points in the process has probability 1 under $\mathbf{P}$ and has probability 0 under $\mathbf{P}^{!x}$). Hence, this process does not have a conditional intensity.

Example 3.12 (Mixture of binomial processes). Suppose $\mathbf{X}$ consists of a random number N of points, where $\mathbb{P}(N = n) = p(n)$, and that given $N = n$, the points are independent and uniformly distributed in a domain W. The intensity of $\mathbf{X}$ is $\beta(u) = \mathbb{E}[N]/\lambda_d(W)$ for $u \in W$, and $\beta(u) = 0$ for $u \notin W$. If Q_n denotes the distribution of the binomial process with n points, then the distribution of $\mathbf{X}$ is

$$\mathbf{P}(A) = \sum_{n=0}^{\infty} p(n) Q_n(A).$$

It is fairly easy to show that the distribution of $\mathbf{X}^{!x}$ is

$$\mathbf{P}^{!x}(A) = \frac{1}{\mathbb{E}[N]} \sum_{n=0}^{\infty} np(n) Q_{n-1}(A).$$

We saw above that Q_n and Q_m are mutually singular for $n \neq m$. Assume that $p(n) > 0$ implies $p(n-1) > 0$ for any n. Then we must have

$$\frac{\beta^*(x, \mathbf{X})}{\beta(x)} = \frac{\mathrm{d}\mathbf{P}^{!x}}{\mathrm{d}\mathbf{P}}(\mathbf{X}) = \frac{r(n(\mathbf{X}))}{\mathbb{E}[N]}$$

where

$$r(n) = \frac{(n+1)p(n+1)}{p(n)}.$$

Hence

$$\beta^*(x, \mathbf{X}) = \frac{r(n(\mathbf{X}))}{\lambda_d(W)}.$$

For example, if $\mathbf{X}$ is a uniform Poisson process with intensity α in W, then $N \sim \mathsf{Poisson}(\alpha\lambda_d(W))$, and we get $r(n) = \alpha\lambda_d(W)$, yielding $\beta^*(x, \mathbf{X}) = \alpha$.

Example 3.13. The randomly translated grid in $\mathbb{R}^2$ was studied in Example 2.3. Intuitively, if we know that there is a point of the grid at the location x, this determines the position of the entire grid. The Palm distribution $\mathbf{P}^x$ for this process is completely deterministic: with probability one, $\mathbf{X}^x$ consists of points at the locations $x + (ks, ms)$ for all integers k, m. This can be proved using (25).

It follows that $\mathbf{P}^x$ is not absolutely continuous with respect to $\mathbf{P}$, so this process does not possess a conditional intensity.

3.6 *J*-function

An interesting combination of the empty space function F and the nearest neighbour function G is the following [47].

Definition 3.5. *Let* $\mathbf{X}$ *be a stationary point process in* $\mathbb{R}^d$. *The* ***J*-function** *of* $\mathbf{X}$ *is*

$$J(r) = \frac{1 - G(r)}{1 - F(r)}$$

for all $r \geq 0$ *such that* $F(r) < 1$.

For a uniform Poisson process, we know that $F(r) \equiv G(r)$ and hence $J(r) \equiv 1$. The J-function of a stationary process can be written explicitly in terms of the conditional intensity:

$$\begin{aligned} J(r) &= \frac{\mathbb{P}^0(\mathsf{dist}(0, \mathbf{X} \setminus 0) > r)}{\mathbb{P}(\mathsf{dist}(0, \mathbf{X}) > r)} \\ &= \frac{\mathbf{P}^{!0}(\mathsf{dist}(0, \mathbf{X}) > r)}{\mathbb{P}(\mathsf{dist}(0, \mathbf{X}) > r)} \\ &= \frac{\mathbb{E}\left[\frac{\beta^*(0,\mathbf{X})}{\beta(0)} \mathbf{1}\{\mathsf{dist}(0, \mathbf{X}) > r\}\right]}{\mathbb{P}(\mathsf{dist}(0, \mathbf{X}) > r)} \\ &= \mathbb{E}\left[\frac{\beta^*(0, \mathbf{X})}{\beta(0)} \mid \mathsf{dist}(0, \mathbf{X}) > r\right]. \end{aligned}$$

This representation can often be evaluated, while F and G often cannot be evaluated explicitly.

The J-function has good properties with respect to many operations on point processes. For example, suppose $\mathbf{X}$ and $\mathbf{Y}$ are independent stationary point processes, with intensities $\alpha_{\mathbf{X}}, \alpha_{\mathbf{Y}}$ and J-functions $J_{\mathbf{X}}, J_{\mathbf{Y}}$. Then the superposition $\mathbf{X} \cup \mathbf{Y}$ has J-function

$$J_{\mathbf{X}\cup\mathbf{Y}}(r) = \frac{\alpha_{\mathbf{X}}}{\alpha_{\mathbf{X}} + \alpha_{\mathbf{Y}}} J_{\mathbf{X}}(r) + \frac{\alpha_{\mathbf{Y}}}{\alpha_{\mathbf{X}} + \alpha_{\mathbf{Y}}} J_{\mathbf{Y}}(r).$$

3.7 Exercises

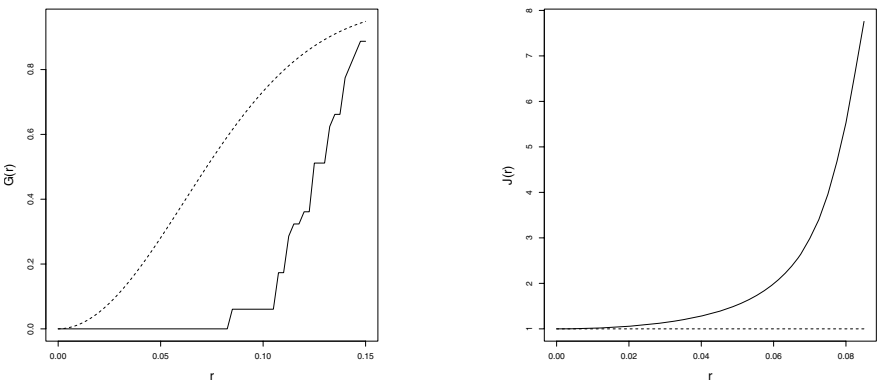

Fig. 32. Estimates of the nearest neighbour function G (Left) and the J-function (Right) for the `cells` dataset. The functions are plotted (as solid lines) against the distance argument r, together with the theoretical expected value for a Poisson process (dotted lines).

Again we use the package `spatstat` described in Sections 1.11 and 2.7. The commands `Gest` and `Jest` compute estimates of the nearest neighbour function G and the J-function, respectively, from spatial point pattern data.

For the `cells` data introduced in Section 1.11, Figure 32 shows an analysis using the G and J functions, produced by the following code.

```
library(spatstat)
data(cells)
G <- Gest(cells)
plot(G)
plot(Jest(cells))
```

Inspection of the plot of $G(r)$ shows that the `cells` data have strong inhibition between points: no point has a nearest neighbour closer than 0.08 units.

4 Modelling and Statistical Inference

This lecture is concerned with statistical models for point pattern data. Point processes in a bounded region of space are considered: in this case it is possible to construct many point process models by writing down their probability densities. We then describe techniques for fitting such models to data [16].

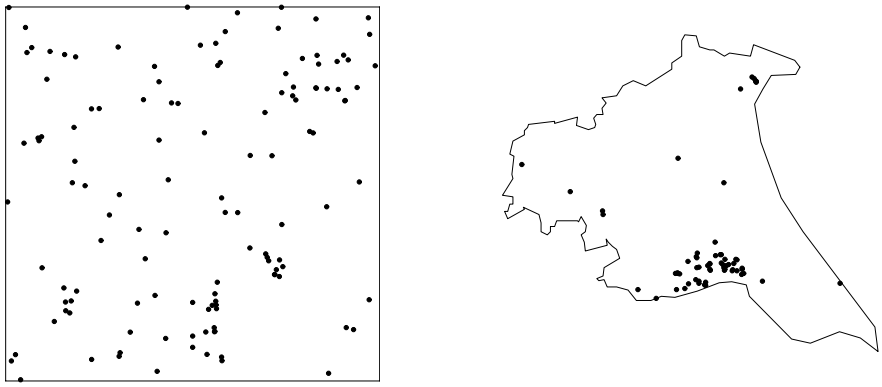

Fig. 33. Examples of point pattern data. *Left:* Locations of 126 pine saplings in a Finnish forest (kindly supplied by S. Kellomaki and A. Penttinen). *Right:* Home address locations of 62 cases of childhood leukaemia and lymphoma in North Humberside, England (from [14]).

4.1 Motivation

Suppose we have observed a point pattern dataset

$$\mathbf{x} = \{x_1, \ldots, x_n\}, \quad n \geq 0, \quad x_i \in W$$

consisting of a finite number of points in a bounded window $W \subset \mathbb{R}^d$ where typically $d = 2, 3$ or 4.

After an initial exploratory phase of data analysis, in which we use tools such as the F, G and K functions to gain qualitative information about the spatial pattern, we may feel confident enough to proceed to a more formal style of statistical analysis. This involves formulating a statistical model, fitting the model to the data, and deciding whether the model is a good fit to the data.

A simple example of a statistical model is the uniform Poisson process with unknown intensity β. This often serves as a reference model for **'complete spatial randomness'** [16], since the points of a uniform Poisson process are stochastically independent of each other, and are uniformly distributed over the region of observation. Alternative models may describe various kinds of departures from a completely random pattern.

4.2 Parametric Modelling and Inference

In this lecture we will adopt the approach of *parametric* statistical modelling. Our statistical model will be a spatial point process $\mathbf{X}$. Our model will completely specify the probability distribution of $\mathbf{X}$, except for the unknown value of a parameter θ. (An example would be the uniform Poisson point process with unknown intensity β). The parameter θ ranges over some set Θ of valid parameter values, typically a subset of $\mathbb{R}^k$ for some $k \geq 1$. Write $\mathbb{P}_\theta$ for the probability distribution, and $\mathbb{E}_\theta$ for the expectation operator, under the model. Then we shall estimate θ from the data $\mathbf{x}$ giving an estimate $\widehat{\theta} = \widehat{\theta}(\mathbf{x})$.

Three generic methods for parameter estimation are surveyed below.

Method of Moments

In the Method of Moments, we choose a statistic $T(\mathbf{x})$, and take our estimate $\widehat{\theta}$ to be the solution of

$$\mathbb{E}_\theta[T(\mathbf{X})] = T(\mathbf{x}). \tag{35}$$

That is, we choose θ so that the theoretical expected value $\mathbb{E}_\theta[T(\mathbf{X})]$ matches the observed value $T(\mathbf{x})$.

Example 4.1. Suppose the model is the uniform Poisson process with intensity β. This can be fitted by the method of moments. Let the statistic T be $T(\mathbf{x}) = n(\mathbf{x})$, the number of points in $\mathbf{x}$. Then we have $\mathbb{E}_\beta[T(\mathbf{X})] = \beta\ \lambda_d(W)$. The solution to (35) is

$$\widehat{\beta} = n(\mathbf{x})/\lambda_d(W).$$

The method of moments works well when we have an analytic expression for $\mathbb{E}_\theta[T(\mathbf{X})]$ as a function of θ, when the equation (35) is guaranteed to have a solution, and when the solution is always unique.

A disadvantage of the method of moments in applications to spatial statistics is that the moments of interesting point processes are often difficult to calculate. Usually we have to resort to simulation. That is, for each possible

value of θ, we generate a large number of simulated realisations $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$ from the model with parameter θ, and estimate $\mathbb{E}_\theta[T(\mathbf{X})]$ by the sample mean $\overline{T} = \frac{1}{N}\sum_i T(\mathbf{x}^{(i)})$. Since the simulations have to be performed again for each new value of θ under consideration, this computation can be expensive.

As we consider more complex models involving several parameters, the statistic T must also become more complex. If θ is a k-dimensional vector, then the values of $T(\mathbf{x})$ must also be k-dimensional (or higher) in order that the solution of (35) may be unique.

Minimum Contrast Estimation

In the Method of Minimum Contrast, our estimate is

$$\widehat{\theta} = \mathsf{argmin}_\theta \;\; D(T(\mathbf{x}), \mathbb{E}_\theta[T(\mathbf{X})]). \tag{36}$$

where $T(\mathbf{x})$ is a chosen statistic, as above, and D is a metric. That is, $\widehat{\theta}$ is the parameter value giving an expectation $\mathbb{E}_\theta[T(\mathbf{X})]$ that is closest (as measured by D) to the observed value $T(\mathbf{x})$. This usually avoids difficulties arising when the method of moments estimating equation (35) does not have a solution. It also allows us to make use of a statistic $T(\mathbf{x})$ which takes values in an arbitrary space.

Diggle & Gratton [18] proposed the method of minimum contrast in combination with the K-function. Suppose that the K-function for the model is known analytically as a function of θ, say $K_\theta(r)$. Given point pattern data $\mathbf{x}$, we may first estimate the K-function using the non-parametric estimator $\widehat{K}(t)$ described in Section 2.6. Then we choose θ to minimise

$$\int_a^b |K_\theta(r) - \widehat{K}(r)|^p \, \mathrm{d}r$$

where $0 \le a < b$ and $p > 0$ are chosen values. This is the minimum contrast method based on $T(\mathbf{x}) = (\widehat{K}(r), a \le r \le b)$, a function-valued statistic, and the metric D is an L^p distance.

Example 4.2. The Thomas process is a cluster process (Section 1.7) formed by taking a Poisson process of 'parents' with intensity α, and replacing each parent x by several offspring, where the number of offspring is Poisson with mean μ, and the offspring of a parent x are i.i.d. random points $y_i = x + e_i$ where the displacement vector e_i has coordinates which are independent Normal $N(0, \sigma^2)$ random variables. The K-function for this process is known analytically:

$$K(r) = \pi r^2 + \frac{1}{\alpha}\left(1 - \exp\left(-\frac{r^2}{4\sigma^2}\right)\right).$$

The intensity of the process is $\beta = \alpha\mu$. Diggle & Gratton [18] proposed fitting this model by the method of minimum contrast. See also [16]. Given point

pattern data $\mathbf{x}$ in a window W, we first estimate the K-function using the non-parametric estimator $\widehat{K}(t)$ described in Section 2.6. The intensity $\beta = \alpha\mu$ is also estimated by the method of moments, $\widehat{\beta} = n(\mathbf{x})/\lambda_2(W)$. Then we find the values (α, μ, r) which minimise

$$\int_a^b |K_\theta(t) - \widehat{K}(t)|^p \, \mathrm{d}t$$

subject to the constraint $\alpha\mu = n(\mathbf{x})/\lambda_2(W)$.

However, there are very few point processes for which we have analytic expressions for K_θ. Usually one would estimate K_θ by simulation.

It is well known that the K function does not completely characterise a stationary point process. A counterexample is given in [5]. Thus, if we are not careful, our model may be **unidentifiable** from the K-function, in the sense that there are two distinct parameter values θ_1, θ_2 such that $K_{\theta_1} \equiv K_{\theta_2}$.

The method of minimum contrast avoids some of the problems of the method of moments. If D has certain convexity properties, then the minimisation (36) has a unique solution. However the statistical properties of the solution are not well understood in general. Additionally, the numerical behaviour of the algorithm used to find the minimum in (36) may also cause difficulties.

Maximum Likelihood

In the Method of Maximum Likelihood, first articulated by R.A. Fisher, we define the likelihood function

$$L(\theta; \mathbf{x}) = \frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(\mathbf{x})$$

where ν is some reference measure. After observing data $\mathbf{x}$, we treat L as a function of θ only, and choose $\widehat{\theta}$ to maximise the likelihood:

$$\widehat{\theta} = \mathsf{argmax}_\theta L(\theta; \mathbf{x})$$

is the value of θ for which the likelihood is maximised. There may be difficulties with non-unique maxima.

Maximum likelihood has some good statistical properties, including asymptotic (large-sample) normality and optimality. Hence a major goal in spatial statistics is to apply maximum likelihood methods.

The likelihood is a probability density. Hence we need to *define/construct point process distributions in terms of their probability densities.*

4.3 Finite Point Processes

A point process $\mathbf{X}$ on S with $N(S) < \infty$ a.s. is called a **finite point process**. The binomial process (Section 1.3) is a simple example in which the total number of points is fixed. In general, the total number of points $N(S)$ is a random variable. The distribution of the process can be specified by giving the probability distribution of $N(S)$, and given $N(S) = n$, the conditional joint distribution of the n points.

Example 4.3. Consider a Poisson process with intensity measure μ that is totally finite ($\mu(S) < \infty$). This is equivalent to choosing a random number $K \sim \mathsf{Poisson}(\mu(S))$, then given $K = k$, generating k i.i.d. random points with common distribution $Q(B) = \mu(B)/\mu(S)$.

Realisations of a finite point process $\mathbf{X}$ belong to the space

$$\mathsf{N}^f = \{N \in \mathsf{N} : \ N(S) < \infty\}$$

of totally finite, simple, counting measures on S. This may be decomposed into subspaces according to the total number of points:

$$\mathsf{N}^f = \mathsf{N}_0 \cup \mathsf{N}_1 \cup \mathsf{N}_2 \cup \ldots$$

where for each $k = 0, 1, 2, \ldots$

$$\mathsf{N}_k = \{N \in \mathsf{N} : \ N(S) = k\}$$

is the set of all counting measures with total mass k, that is, effectively the set of all configurations of k points. This can be represented more explicitly by introducing the space of ordered k-tuples

$$S^{!k} = \{(x_1, \ldots, x_k) : \ x_i \in S, \ x_i \neq x_j \text{ for all } i \neq j\}.$$

Define a mapping $I_k : S^{!k} \to \mathsf{N}_k$ by

$$I_k(x_1, \ldots, x_k) = \delta_{x_1} + \ldots + \delta_{x_k}$$

This gives

$$\mathsf{N}_k \equiv S^{!k}/\sim$$

where $\sim$ is the equivalence relation under permutation, i.e.

$$(x_1, \ldots, x_k) \sim (y_1, \ldots, y_k) \quad \Leftrightarrow \quad \{x_1, \ldots, x_k\} = \{y_1, \ldots, y_k\}$$

Using this representation we can give explicit formulae for point process distributions.

Example 4.4 (binomial process). Let $X_1, \ldots, X_n$ be i.i.d. random points uniformly distributed in W. Set $\mathbf{X} = I_n(X_1, \ldots, X_n)$. The distribution of $\mathbf{X}$ is the probability measure $P_{\mathbf{X}}$ on N defined by

$$\begin{aligned} P_{\mathbf{X}}(A) &= \mathbb{P}(I_n(X_1, \ldots, X_n) \in A) \\ &= \frac{1}{|W|^n} \int_W \ldots \int_W \mathbf{1}\{I_n(x_1, \ldots, x_n) \in A\} \, \mathrm{d}x_1 \ldots \, \mathrm{d}x_n \end{aligned}$$

Example 4.5 (finite Poisson process). Let $\mathbf{X}$ be the Poisson process on S with totally finite intensity measure μ. We know that $N(S) \sim \mathsf{Poisson}(\mu(S))$ and that, given $N(S) = n$, the distribution of $\mathbf{X}$ is that of a binomial process of n points i.i.d. with common distribution $Q(B) = \mu(B)/\mu(S)$. Thus

$$\begin{aligned} P_{\mathbf{X}}(A) &= \sum_{n=0}^{n} \mathbb{P}(N(S) = n) \mathbb{P}(I_n(X_1, \ldots, X_n) \in A) \\ &= \sum_{n=0}^{\infty} e^{-\mu(S)} \frac{\mu(S)^n}{n!} \int_S \ldots \int_S \mathbf{1}\{I_n(x_1, \ldots, x_n) \in A\} \, Q(\mathrm{d}x_1) \ldots \, Q(\mathrm{d}x_n) \\ &= e^{-\mu(S)} \sum_{n=0}^{\infty} \frac{1}{n!} \int_S \ldots \int_S \mathbf{1}\{I_n(x_1, \ldots, x_n) \in A\} \, \mu(\mathrm{d}x_1) \ldots \, \mu(\mathrm{d}x_n). \end{aligned} \tag{37}$$

The term for $n = 0$ in the sum should be interpreted as $\mathbf{1}\{\mathbf{0} \in A\}$ (where $\mathbf{0}$ is the zero measure, corresponding to an empty configuration.)

4.4 Point Process Densities

Henceforth we fix a standard measure μ on S. Typically μ is Lebesgue measure on a bounded set W in $\mathbb{R}^d$. Let π_μ denote the distribution of the Poisson process with intensity measure μ.

Definition 4.1. *Let $f : \mathsf{N}^f \to \mathbb{R}_+$ be a measurable function satisfying $\int_{\mathsf{N}} f(\mathbf{x}) \, \pi_\mu(\mathrm{d}\mathbf{x}) = 1$. Define*

$$\mathbf{P}(A) = \int_A f(\mathbf{x}) \, \pi_\mu(\mathrm{d}\mathbf{x}).$$

for any event $A \in \mathcal{N}$. Then $\mathbf{P}$ is a point process distribution in the sense of Definition 1.5. The function f is said to be the **probability density** *of the point process with distribution $\mathbf{P}$.*

Lemma 4.1. *For a point process $\mathbf{X}$ with probability density f we have*

$$\begin{aligned} \mathbb{P}(X \in A) = e^{-\mu(S)} \sum_{n=0}^{\infty} \frac{1}{n!} \int_S \ldots \int_S \mathbf{1}\{I_n(x_1, \ldots, x_n) \in A\} \\ f(I_n(x_1, \ldots, x_n)) \, \mu(\mathrm{d}x_1) \ldots \, \mu(\mathrm{d}x_n) \end{aligned} \tag{38}$$

for any event $A \in \mathcal{N}$, and

$$\mathbb{E}[g(\mathbf{X})] = e^{-\mu(S)} \sum_{n=0}^{\infty} \frac{1}{n!} \int_S \dots \int_S g(I_n(x_1, \dots, x_n)) \\ f(I_n(x_1, \dots, x_n))\, \mu(\mathrm{d}x_1) \dots\, \mu(\mathrm{d}x_n) \tag{39}$$

for any integrable function $g : \mathsf{N} \to \mathbb{R}_+$.

We can also rewrite these identities as

$$\mathbb{P}(X \in A) = \mathbb{E}\left[f(\Pi) 1_A(\Pi)\right] \tag{40}$$

$$\mathbb{E}[g(\mathbf{X})] = \mathbb{E}[g(\Pi) f(\Pi)] \tag{41}$$

where Π is the Poisson process with intensity measure μ.

For some elementary point processes, it is possible to determine the probability density directly.

Example 4.6 (uniform Poisson process). Let $\beta > 0$. Set

$$f(\mathbf{x}) = \alpha\, \beta^{n(\mathbf{x})}$$

where α is a normalising constant and $n(\mathbf{x}) = \mathbf{x}(S)$ =number of points in $\mathbf{x}$. Then

$$\mathbf{P}(A) = \alpha e^{-\mu(S)} \sum_{n=0}^{\infty} \frac{1}{n!} \int_S \dots \int_S \mathbf{1}\{I_n(x_1, \dots, x_n) \in A\} \beta^n\, \mu(\mathrm{d}x_1) \dots\, \mu(\mathrm{d}x_n).$$

But this is the distribution of the Poisson process with intensity β. The normalising constant must be $\alpha = e^{(1-\beta)\mu(S)}$. Thus, the uniform Poisson process with intensity β has probability density

$$f(\mathbf{x}) = \beta^{n(\mathbf{x})} e^{(1-\beta)\mu(S)}. \tag{42}$$

Example 4.7 (Hard core process). Fix $r > 0$. Let

$$H_n = \left\{(x_1, \dots, x_n) \in S^{!n} : \; ||x_i - x_j|| \geq r \text{ for all } i \neq j\right\}$$

and

$$H = \bigcup_{n=0}^{\infty} I_n(H_n).$$

Thus H is the subset of N consisting of all point patterns $\mathbf{x}$ with the property that every pair of distinct points in $\mathbf{x}$ is at least r units apart. Now define the probability density

$$f(\mathbf{x}) = \alpha \mathbf{1}\{\mathbf{x} \in H\}$$

where α is the normalising constant. Then we have, for any event $A \in \mathcal{N}$,

$$\begin{aligned}\mathbf{P}(A) &= \mathbb{E}\left[f(\Pi)\mathbf{1}\{\Pi \in A\}\right] \\ &= \alpha\mathbb{E}\left[\mathbf{1}\{\Pi \in H\}\mathbf{1}\{\Pi \in A\}\right] \\ &= \alpha\mathbb{P}(\Pi \in H \cap A).\end{aligned}$$

It follows that

$$\alpha = 1/\mathbb{P}(\Pi \in H)$$

and $\mathbf{P}$ is the conditional distribution of the Poisson process Π given that $\Pi \in H$. In other words, the process $\mathbf{X}$ with probability density f is equivalent to a Poisson process conditioned on the event that there are no pairs of points closer than r units apart.

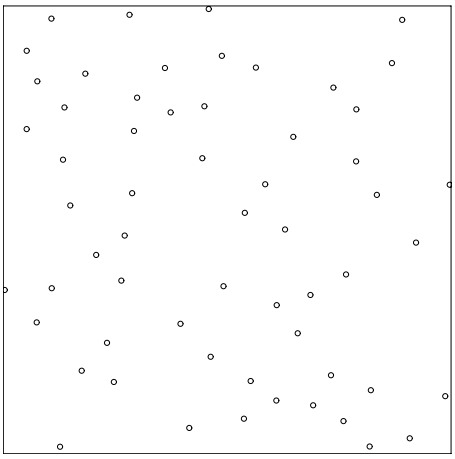

Fig. 34. Realisation of the hard core process with $\beta = 200$ and $r = 0.07$ in the unit square.

Figure 34 shows a realisation of the hard core process with $\beta = 200$ and $r = 0.07$ in the unit square. One simple way to generate such a picture is by the rejection method: we generate a sequence of realisations of the uniform Poisson process with intensity 200, and plot the first realisation which satisfies the constraint H. More efficient simulation methods are described in [33, 46].

4.5 Conditional Intensity

Consider a finite point process $\mathbf{X}$ in a compact set $W \subset \mathbb{R}^d$. Recall that the conditional intensity $\beta^*(u, \mathbf{X})$, if it exists, satisfies

$$\mathbb{E}\sum_{x\in\mathbf{X}} g(x, \mathbf{X} \setminus x) = \int_{\mathbb{R}^d} \mathbb{E}\left[\beta^*(u, \mathbf{X})g(u, \mathbf{X})\right]\, \mathrm{d}u \tag{43}$$

for any integrable g.

Now suppose $\mathbf{X}$ has probability density $f(\mathbf{x})$ (with respect to the uniform Poisson process Π with intensity 1 on W). Then the expectation of any integrable function $h(\mathbf{X})$ may be written in the form (39) or (41). Applying this to both sides of (43) above, we get

$$\mathbb{E}\left[f(\Pi)\sum_{x\in\Pi} g(x,\Pi\setminus x)\right] = \int_{\mathbb{R}^d} \mathbb{E}\left[\beta^*(u,\Pi)f(\Pi)g(u,\Pi)\right]\,\mathrm{d}u. \tag{44}$$

If we write

$$h(x,\mathbf{X}) = f(\mathbf{X}\cup\{x\})g(x,\mathbf{X}),$$

then the left side of (44) can be rewritten

$$\begin{aligned}\mathbb{E}\left[f(\Pi)\sum_{x\in\Pi} g(x,\Pi\setminus x)\right] &= \mathbb{E}\left[\sum_{x\in\Pi} h(x,\Pi\setminus x)\right]\\ &= \int_W \mathbb{E}[h(u,\mathbf{X})]\,\mathrm{d}u,\end{aligned}$$

where the last line is obtained by applying equation (43) to the process Π, since the conditional intensity of Π is identically equal to 1 on W. Thus we get

$$\int_{\mathbb{R}^d} \mathbb{E}\left[\beta^*(u,\Pi)f(\Pi)g(u,\Pi)\right]\,\mathrm{d}u = \int_W \mathbb{E}\left[f(\Pi\cup\{u\})g(u,\Pi)\right]\,\mathrm{d}u$$

for all integrable functions g. It follows that

$$\beta^*(u,\Pi)f(\Pi) = f(\Pi\cup u)$$

almost surely, for almost all $u\in W$. Thus we have obtained the following result.

Theorem 4.1. *Let f be the probability density of a finite point process $\mathbf{X}$ in a bounded region W of $\mathbb{R}^d$. Assume that*

$$f(\mathbf{x})>0 \quad\Rightarrow\quad f(\mathbf{y})>0 \text{ for all } \mathbf{y}\subset\mathbf{x}.$$

Then the conditional intensity of $\mathbf{X}$ exists and equals

$$\beta^*(u,\mathbf{x}) = \frac{f(\mathbf{x}\cup u)}{f(\mathbf{x})} \tag{45}$$

almost everywhere.

Example 4.8 (Uniform Poisson). The uniform Poisson process on W with intensity β has density

$$f(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})}$$

where α is a certain normalising constant. Applying (45) we get

$$\beta^*(u, \mathbf{x}) = \beta$$

for $u \in W$.

Example 4.9 (Hard core process). The probability density of the hard core process (Example 4.7)

$$f(\mathbf{x}) = \alpha \mathbf{1}\{\mathbf{x} \in H\}$$

yields

$$\lambda(u, \mathbf{x}) = \mathbf{1}\{\mathbf{x} \cup u \in H\}.$$

Lemma 4.2. *Let* $\mathbf{X}$ *be a finite point process in a bounded region* W *in* $\mathbb{R}^d$. *Suppose that* $\mathbf{X}$ *has a probability density* f *and a conditional intensity* β^*. *Then* f *is completely determined by* β^*.

Proof. We may invert the relationship (45) by starting with the empty configuration $\emptyset$ and adding one point at a time:

$$\begin{aligned} f(\{x_1, \ldots, x_n\}) &= f(\emptyset) \frac{f(\{x_1\})}{f(\emptyset)} \frac{f(\{x_1, x_2\})}{f(\{x_1\})} \cdots \frac{f(\{x_1, \ldots, x_n\})}{f(\{x_1, \ldots, x_{n-1}\})} \\ &= f(\emptyset) \beta^*(x_1, \emptyset) \beta^*(x_2, \{x_1\}) \ldots \beta^*(x_n, \{x_1, \ldots, x_{n-1}\}). \end{aligned}$$

If the values of β^* are known, then this determines f up to a constant $f(\emptyset)$, which is then determined by the normalisation of f. □

It is often convenient to formulate a point process model in terms of its conditional intensity $\beta^*(u, \mathbf{x})$, rather than its probability density $f(\mathbf{x})$. The conditional intensity has a natural interpretation (in terms of conditional probability) which may be easier to understand than the density. Using the conditional intensity also eliminates the normalising constant needed for the probability density.

However, we are not free to choose the functional form of $\beta^*(u, \mathbf{x})$ at will. It must satisfy certain consistency relations. The next section describes a large class of models which turns out to characterise the most general functional form of $\beta^*(u, \mathbf{x})$.

4.6 Finite Gibbs Models

Definition 4.2. *A* **finite Gibbs process** *is a finite point process* $\mathbf{X}$ *with probability density* $f(\mathbf{x})$ *of the form*

$$f(\mathbf{x}) = \exp(V_0 + \sum_{x \in \mathbf{x}} V_1(x) + \sum_{\{x,y\} \subset \mathbf{x}} V_2(x, y) + \ldots) \tag{46}$$

where $V_k : \mathsf{N}_k \to \mathbb{R} \cup \{-\infty\}$ *is called the* **potential** *of order* k.

Gibbs models arise in statistical physics, where $\log f(\mathbf{x})$ may be interpreted as the **potential energy** of the configuration $\mathbf{x}$. The term $-V_1(u)$ can be interpreted as the energy required to create a single point at a location u. The term $-V_2(u,v)$ can be interpreted as the energy required to overcome a force between the points u and v.

Example 4.10 (Hard core process). Give parameters $\beta, r > 0$, define $V_1(u) = \log \beta$ and

$$V_2(u,v) = \begin{cases} 0 & \text{if } ||u-v|| > r \\ -\infty & \text{if } ||u-v|| \le r \end{cases}$$

and $V_k \equiv 0$ for all $k \ge 3$. Then $\sum_{\{x,y\}\subset\mathbf{x}} V_2(x,y)$ is equal to zero if all pairs of points in $\mathbf{x}$ are at least r units apart, and otherwise this sum is equal to $-\infty$. Taking $\exp(-\infty) = 0$, we find that (46) is

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \mathbf{1}\{\mathbf{x} \in H\}$$

where H is the hard core constraint set defined in Example 4.7, and $\alpha = \exp(V_0)$ is a normalising constant. This is the probability density of the hard core process.

Theorem 4.2. *Let f be the probability density of a finite point process $\mathbf{X}$ in a bounded region W in $\mathbb{R}^d$. Suppose that*

$$f(\mathbf{x}) > 0 \quad \Rightarrow \quad f(\mathbf{y}) > 0 \text{ for all } \mathbf{y} \subset \mathbf{x}. \tag{47}$$

Then f can be expressed in the Gibbs form (46).

Proof. This is a consequence of the Möbius inversion formula (the 'inclusion-exclusion principle'). The functions V_k can be obtained explicitly as

$$\begin{aligned} V_0 &= \log f(\emptyset) \\ V_1(u) &= \log f(\{u\}) - \log f(\emptyset) \\ V_2(u,v) &= \log f(\{u,v\}) - \log f(\{u\}) - \log f(\{v\}) + \log f(\emptyset) \end{aligned}$$

and in general

$$V_k(\mathbf{x}) = \sum_{\mathbf{y}\subseteq\mathbf{x}} (-1)^{n(\mathbf{x})-n(\mathbf{y})} \log f(\mathbf{y}).$$

Then equation (46) can be verified by induction on $n(\mathbf{x})$. □

Any process satisfying (47) also has a conditional intensity, by Theorem 4.1. The corresponding conditional intensity is

$$\beta^*(u,\mathbf{x}) = \exp\left[V_1(u) + \sum_{x\in\mathbf{X}} V_2(u,x) + \sum_{\{x,y\}\subset\mathbf{X}} V_3(u,x,y) + \dots \right] \tag{48}$$

Hence, this is the most general form of a conditional intensity:

Theorem 4.3. *Let* $\mathbf{X}$ *be a finite point process in a bounded region* W *of* $\mathbb{R}^d$. *Assume that* $\mathbf{X}$ *has a probability density* f *satisfying (47).*

A function $\beta^*(u, \mathbf{x})$ *is the conditional intensity of* $\mathbf{X}$ **if and only if** *it can be expressed in the form (48).*

Example 4.11 (Strauss process). For parameters $\beta > 0$, $0 \le \gamma \le 1$ and $r > 0$, suppose

$$\begin{aligned} V_1(u) &= \log \beta \\ V_2(u, v) &= (\log \gamma)\, \mathbf{1}\{||u - v|| \le r\}. \end{aligned}$$

Then we have

$$\beta^*(u, \mathbf{x}) = \beta \gamma^{t(u,\mathbf{x})}$$

where

$$t(u, \mathbf{x}) = \sum_{x \in \mathbf{x}} \mathbf{1}\{||u - x|| \le r\}$$

is the number of points of $\mathbf{x}$ which are close to u. Also

$$f(\mathbf{x}) = \alpha\, \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})}$$

where

$$s(\mathbf{x}) = \sum_{x,y \in \mathbf{x}} \mathbf{1}\{||x - y|| \le r\}$$

is the number of pairs of close points in $\mathbf{x}$. The normalising constant α is not available in closed form.

When $\gamma = 1$, this reduces to the Poisson process with intensity β. When $\gamma = 0$, we have

$$\beta^*(u, \mathbf{x}) = \mathbf{1}\{||u - x|| > r \text{ for all } x \in \mathbf{x}\}$$

and

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \mathbf{1}\{\mathbf{x} \in H\}$$

so we get the hard core process.

For $0 < \gamma < 1$ the Strauss process has 'soft inhibition' between neighbouring pairs of points.

Figure 35 shows simulated realisations of the Strauss process with two values of the interaction parameter γ. The case $\gamma = 0$ was already illustrated in Figure 34.

The intensity function of the Strauss process is, applying equation (34),

$$\begin{aligned} \beta(u) &= \mathbb{E}[\beta^*(u, \mathbf{X})] \\ &= \mathbb{E}[\beta \gamma^{t(u,\mathbf{X})}] \le \beta \end{aligned}$$

It is not easy to evaluate $\beta(u)$ explicitly as a function of β, γ, r.

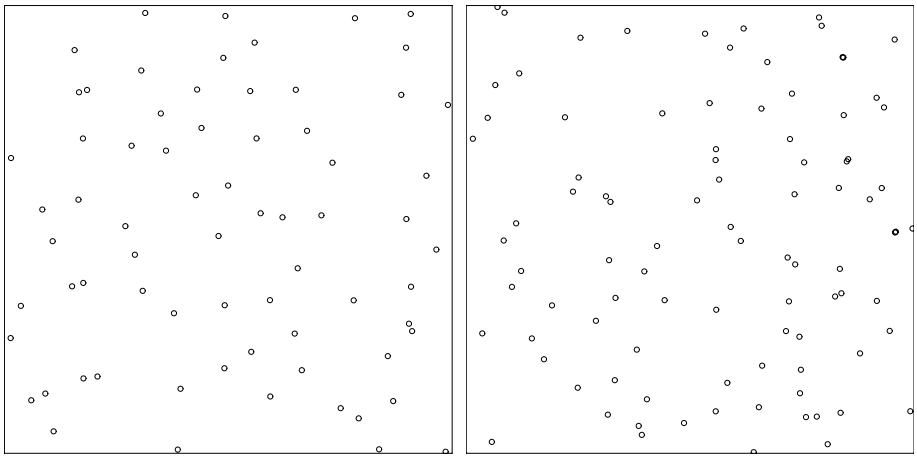

Fig. 35. Realisations of the Strauss process with interaction parameter $\gamma = 0.2$ (Left) and $\gamma = 0.5$ (Right) in the unit square, both having activity $\beta = 200$ and interaction range $r = 0.07$.

4.7 Parameter Estimation

Finally we return to the question of fitting a parametric model to point pattern data. Suppose we have observed a point pattern $\mathbf{x}$ in a bounded window $W \subset \mathbb{R}^2$, and wish to model it as a realisation of a finite point process $\mathbf{X}$ with probability density $f(\mathbf{x}; \theta)$ where θ is the parameter. The likelihood is

$$L(\theta) = f(\mathbf{x}; \theta).$$

Define the maximum likelihood estimator

$$\widehat{\theta} = \mathsf{argmax}_\theta L(\theta).$$

Example 4.12 (Uniform Poisson process). Suppose the model X is a uniform Poisson process in W with intensity β. The probability density of this model was found in Example 4.6 to be

$$f(\mathbf{x}; \beta) = \beta^{n(\mathbf{x})} \exp((1 - \beta)\lambda_2(W)).$$

Thus the log likelihood is

$$L(\beta) = n(\mathbf{x}) \log \beta + (1 - \beta)\lambda_2(W)$$

so the score is

$$U(\beta) = \frac{\mathrm{d}}{\mathrm{d}\beta} \log L(\beta) = \frac{n(\mathbf{x})}{\beta} - \lambda_2(W)$$

so the maximum likelihood estimate of β is

$$\widehat{\beta} = \frac{n(\mathbf{x})}{\lambda_2(W)},$$

the same as the method-of-moments estimator.

Example 4.13 (Hard core process). The probability density for a hard core process was found in Example 4.7 to be

$$f(\mathbf{x}; r) = \frac{\mathbf{1}\{\mathbf{x} \in H_r\}}{\mathbb{P}(\Pi \in H_r)}$$

where H_r is the set of all point patterns $\mathbf{x}$ with the property that every pair of distinct points in $\mathbf{x}$ is at least r units apart. Again Π denotes the uniform Poisson process with intensity 1.

For any configuration $\mathbf{x}$ let $m(\mathbf{x}) = \min\{||x_i - x_j|| : x_i, x_j \in \mathbf{x}\}$ be the minimum distance between any pair of distinct points in $\mathbf{x}$. Then $H_r = \{\mathbf{x} : m(\mathbf{x}) \geq r\}$. Thus the likelihood can be written

$$L(r) = f(\mathbf{x}; r) = \frac{\mathbf{1}\{m(\mathbf{x}) \geq r\}}{\mathbb{P}(m(\Pi) \geq r)}.$$

The numerator is equal to 1 for all $r \leq m(\mathbf{x})$ and to 0 for $r > m(\mathbf{x})$. The denominator is clearly a decreasing (or at least non-increasing) function of r. Thus, $L(r)$ is a non-decreasing function of r on the range $0 \leq r \leq m(\mathbf{x})$, and is zero for $r > m(\mathbf{x})$. It follows that the maximum likelihood estimate of r is

$$\widehat{r} = m(\mathbf{x}).$$

For more complicated models, we rapidly run into problems in applying the method of maximum likelihood. The likelihood of a cluster process is known analytically [9], but is difficult to maximise. The likelihood of many Gibbs models contains a normalising constant, which is usually an intractable function of θ. Hence, analytic maximisation of L is often difficult [33].

Instead, some form of approximation is often employed as a substitute for exact maximum likelihood estimation. One strategy is to approximate the likelihood itself, for example using a series expansion (such as **virial expansion**), or using simulation (**Monte Carlo maximum likelihood** [21, 33]). Monte Carlo methods are probably the most popular approach, although they have many technical difficulties and are highly computationally-intensive.

Some experts also claim to be able to 'fit by eye' [40].

4.8 Estimating Equations

Another strategy for parameter estimation is to replace the maximum likelihood estimating equations

$$U(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} \log L(\theta) = 0 \tag{49}$$

by another system of equations.

Let $\Psi(\theta, \mathbf{X})$ be a function such that, for any θ,

$$\mathbb{E}_\theta[\Psi(\theta, \mathbf{X})] = 0 \tag{50}$$

where $\mathbb{E}_\theta$ denotes the expectation over the distribution of X when θ is the true parameter value. When data $\mathbf{x}$ are observed, suppose we estimate θ by the solution $\widehat{\theta}$ of

$$\Psi(\theta, \mathbf{x}) = 0. \tag{51}$$

Then (51) is called an **unbiased estimating equation** [22].

This concept embraces both maximum likelihood and the method of moments. A key result in the classical theory of maximum likelihood is that (49) is an unbiased estimating equation. The method of moments is characterised by the estimating equation (35), which can be rewritten in the form (51) where $\Psi(\theta, \mathbf{x}) = T(\mathbf{x}) - \mathbb{E}_\theta[T(\mathbf{X})]$. This is clearly an unbiased estimating equation.

The term 'unbiased' should not be misinterpreted as suggesting that $\widehat{\theta}$ is an unbiased estimator of θ. However under reasonable limiting conditions $\widehat{\theta}$ is consistent and asymptotically unbiased.

Another family of estimating equations for point process models was suggested by Tákacs and Fiksel [20, 44, 45]. Consider again the identity (31). In our context the conditional intensity of the point process $\mathbf{X}$ depends on the parameter θ, so we denote the conditional intensity by $\beta^*_\theta(u, \mathbf{x})$, and equation (31) becomes

$$\mathbb{E} \sum_{x \in \mathbf{X}} g(x, \mathbf{X} \setminus x) = \int_{\mathbb{R}^d} \mathbb{E}\left[\beta^*_\theta(x, \mathbf{X}) g(x, \mathbf{X})\right] \, \mathrm{d}x$$

for arbitrary integrable functions g. It follows that if we define

$$\Psi(\theta, \mathbf{x}) = \sum_{x_i \in \mathbf{x} \cap A} g(x_i, \mathbf{x} \setminus \{x_i\}) - \int_A \beta^*_\theta(u, \mathbf{x}) g(u, \mathbf{x}) \, \mathrm{d}u \tag{52}$$

for any chosen function g and any $A \subseteq W$, then (50) holds, and we have an unbiased estimating equation.

Example 4.14. To fit the Strauss process (Example 4.11) to a point pattern dataset $\mathbf{x}$, recall that the conditional intensity is $\beta^*_\theta(u, \mathbf{x}) = \beta\gamma^{t(u,\mathbf{x})}$. Assume r is fixed and known. Taking $g \equiv 1$ and $A = W$ in (52) gives

$$\Psi(\theta, \mathbf{x}) = n(\mathbf{x}) - \beta \int_W \gamma^{t(u,\mathbf{x})} \, \mathrm{d}u.$$

Taking $g(u, \mathbf{x}) = t(u, \mathbf{x})$ gives

$$\begin{aligned} \Psi(\theta, \mathbf{x}) &= \sum_{i=1}^{n(\mathbf{x})} t(x_i, \mathbf{x} \setminus \{x_i\}) - \beta \int_W t(u, \mathbf{x}) \gamma^{t(u,\mathbf{x})} \, \mathrm{d}u \\ &= 2s(\mathbf{x}) - \beta \int_W t(u, \mathbf{x}) \gamma^{t(u,\mathbf{x})} \, \mathrm{d}u \end{aligned}$$

where $s(\mathbf{x})$ is the number of unordered pairs of points in $\mathbf{x}$ which are closer than a distance r apart. Equating these two functions Ψ to zero gives us two equations in the two unknowns β and γ which can be solved numerically to give estimates $\widehat{\beta}, \widehat{\gamma}$. To simplify these expressions, notice that $t(u, \mathbf{x})$ has integer values, and let

$$m_k = \int_W \mathbf{1}\{t(u, \mathbf{x}) = k\} \, \mathrm{d}u$$

for each $k = 0, 1, 2, \ldots$. Then the equations to be solved are

$$n(\mathbf{x}) = \beta \sum_{k=0}^{\infty} \gamma^k m_k$$
$$2s(\mathbf{x}) = \beta \sum_{k=0}^{\infty} k\gamma^k m_k.$$

4.9 Likelihood Devices

Another approach is to replace the likelihood function by another function altogether. An example is the pseudolikelihood proposed by Besag [10, 11].

Definition 4.3. *For a point process* $\mathbf{X}$ *with conditional intensity* $\beta^*_\theta(u, \mathbf{x})$*, where* θ *is the unknown parameter, define the* **pseudolikelihood**

$$\mathsf{PL}(\theta, \mathbf{x}) = \left(\prod_{x_i \in \mathbf{x}} \beta^*_\theta(x_i, \mathbf{x}) \right) \exp\left[- \int_W \beta^*_\theta(u, \mathbf{x}) \, \mathrm{d}u \right]. \tag{53}$$

The **maximum pseudolikelihood estimator (MPLE)** *of* θ *is the value maximising* $\mathsf{PL}(\theta)$*.*

A rationale for using this function is offered in [10, 12]. If $\mathbf{x}$ is treated as fixed, and the pseudolikelihood is considered as a function of θ, then the pseudolikelihood has the same functional form as the likelihood of an inhomogeneous Poisson process. The pseudolikelihood for a Poisson process is identical to its likelihood, up to a constant factor. Thus, the pseudolikelihood may be regarded as an approximation to the likelihood which ignores the dependence between points. It may be expected to be a good approximation to the likelihood when interpoint dependence is weak.

Example 4.15. For the Strauss process the log pseudolikelihood is

$$\begin{aligned} \log \mathsf{PL}(\beta, \gamma, r) &= \sum_{i=1}^{n(\mathbf{x})} (\log \beta + t(x_i, \mathbf{x} \setminus \{x_i\}) \log \gamma) - \int_W \beta \gamma^{t(u,\mathbf{x})} \, \mathrm{d}u \\ &= n(\mathbf{x}) \log \beta + 2s(\mathbf{x}) \log \gamma - \int_W \beta \gamma^{t(u,\mathbf{x})} \, \mathrm{d}u. \end{aligned}$$

The score components for β, γ are

$$\frac{\partial}{\partial\beta}\log \mathsf{PL} = \frac{n(\mathbf{x})}{\beta} - \int_W \gamma^{t(u,\mathbf{x})}\,\mathrm{d}u$$
$$\frac{\partial}{\partial\gamma}\log \mathsf{PL} = \frac{2s(\mathbf{x})}{\gamma} - \beta\int_W t(u,\mathbf{x})\gamma^{t(u,\mathbf{x})-1}\,\mathrm{d}u.$$

The maximum pseudolikelihood estimators of β, γ are the roots of these two functions.

Notice that the maximum pseudolikelihood estimating equations derived in Example 4.15 are equivalent to those derived using the Takacs-Fiksel method in Example 4.14. This is true in greater generality, as shown in [3, 17].

Powerful advantages of the pseudolikelihood are that it avoids the normalising constant present in the likelihood, that it is usually easy to compute and maximise, and that it retains good properties such as consistency and asymptotic normality in a large-sample limit [25, 24]. A disadvantage of the pseudolikelihood is that it is known to be biased and statistically inefficient in small samples.

An algorithm for fitting very general point process models by maximum pseudolikelihood was developed in [6]. It is now possible to perform parametric modelling and inference for spatial point processes using tools similar to those available for other types of data. For further details, see [8, 7].

References

1. Babu, G.J., Feigelson, E.D.: Astrostatistics. Chapman & Hall, London (1996)
2. Baddeley, A.J.: Spatial sampling and censoring. In: Barndorff-Nielsen, O.E., Kendall, W.S., van Lieshout, M.N.M. (eds) Stochastic Geometry. Likelihood and Computation. Monographs on Statistics and Applied Probability, **80**, Chapman & Hall/CRC, Boca Raton, FL (1999)
3. Baddeley, A.: Time-invariance estimating equations. Bernoulli, **6**, 783–808 (2000)
4. Baddeley, A., Jensen, E.B. Vedel: Stereology for Statisticians. Chapman & Hall/CRC, Boca Raton, FL (2005)
5. Baddeley, A.J., Silverman, B.W.: A cautionary example on the use of second-order methods for analyzing point patterns. Biometrics, **40**, 1089–1094 (1984)
6. Baddeley, A., Turner, R.: Practical maximum pseudolikelihood for spatial point patterns (with discussion). Austral. N. Z. J. Stat., **42**(3), 283–322 (2000)
7. Baddeley, A., Turner, R.: Modelling spatial point patterns in R. In: Baddeley, A., Mateu, J., Stoyan, D. (eds) Case Studies in Spatial Point Pattern Modelling. Springer, Berlin Heidelberg New York (to appear)
8. Baddeley, A., Turner, R.: Spatstat: an R package for analyzing spatial point patterns. J. Statist. Software, **12**(6), 1–42 (2005). URL: `www.jstatsoft.org`, ISSN: 1548-7660
9. Baddeley, A.J., van Lieshout, M.N.M, Møller, J.: Markov properties of cluster processes. Adv. in Appl. Probab., **28**, 346–355 (1996)

10. Besag, J.: Statistical analysis of non-lattice data. The Statistician, **24**, 179–195 (1975)
11. Besag, J.: Some methods of statistical analysis for spatial data. Bull. Internat. Statist. Inst., **44**, 77–92 (1978)
12. Besag, J., Milne, R., Zachary, S.: Point process limits of lattice processes. J. Appl. Probab., **19**, 210–216 (1982)
13. The Comprehensive R Archive Network. URL `http://www.cran.r-project.org`
14. Cuzick, J., Edwards, R.: Spatial clustering for inhomogeneous populations (with discussion). J. Royal Statist. Soc. Ser. B, **52**, 73–104 (1990)
15. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes. Springer, Berlin Heidelberg New York (1988)
16. Diggle, P.J.: Statistical Analysis of Spatial Point Patterns. Second edition, Arnold, London (2003)
17. Diggle, P.J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D., Tanemura, M.: On parameter estimation for pairwise interaction processes. Internat. Statist. Rev., **62**, 99–117 (1994)
18. Diggle, P.J., Gratton, R.J.: Monte Carlo methods of inference for implicit statistical models (with discussion). J. Royal Statist. Soc. Ser. B, **46**, 193–227 (1984)
19. Elliott, P., Wakefield, J., Best, N., Briggs, D. (eds): Spatial Epidemiology: Methods and Applications. Oxford University Press, Oxford (2000)
20. Fiksel, T.: Estimation of interaction potentials of Gibbsian point processes. Statistics, **19**, 77–86 (1988)
21. Geyer, C.J.: Likelihood inference for spatial point processes. In: Barndorff-Nielsen, O.E., Kendall, W.S., van Lieshout, M.N.M. (eds) Stochastic Geometry. Likelihood and Computation. Monographs on Statistics and Applied Probability, **80**, Chapman & Hall/CRC, Boca Raton, FL (1999)
22. Godambe, V.P. (ed): Estimating Functions. Oxford Statistical Science Series, **7**, Oxford Science Publications, Oxford (1991)
23. Hornik, K.: The R FAQ: Frequently asked questions on R. URL `http://www.ci.tuwien.ac.at/~hornik/R/`. ISBN 3-901167-51-X
24. Jensen, J.L., Künsch, H.R.: On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. Ann. Inst. Statist. Math., **46**, 475–486 (1994)
25. Jensen, J.L., Møller, J.: Pseudolikelihood for exponential family models of spatial point processes. Ann. Appl. Probab., **1**, 445–461 (1991)
26. Kallenberg, O.: Random Measures. Third edition, Akademie Verlag/Academic Press, Berlin/New York (1983)
27. Kendall, D.G.: Foundations of a theory of random sets. In: Harding, E.F., Kendall, D.G. (eds) Stochastic Geometry. John Wiley & Sons, Chichester (1974)
28. Kingman, J.F.C.: Poisson Processes. Oxford University Press, Oxford (1993)
29. Ludwig, J.A., Reynolds, J.F.: Statistical Ecology: a Primer on Methods and Computing. John Wiley & Sons, New York (1988)
30. Matérn, B.: Spatial Variation. Meddelanden från Statens Skogsforskningsinstitut, **49**(5), 1–114 (1960)
31. Matheron, G.: Ensembles fermés aléatoires, ensembles semi-markoviens et polyèdres poissoniens. Adv. in Appl. Probab., **4**, 508–543 (1972)
32. Matheron, G.: Random Sets and Integral Geometry. John Wiley & Sons, New York (1975)

33. Møller, J., Waagepetersen, R.P.: Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall/CRC, Boca Raton, FL (2003)
34. Ohser, J., Mücklich, F.: Statistical Analysis of Microstructures in Materials Science. John Wiley & Sons, Chichester (2000)
35. Ornstein, L.S., Zernike, F.: Accidental deviations of density and opalesence at the critical point of a single substance. Proceedings, Section of Sciences, Royal Academy of Sciences/ Koninklijke Akademie van Wetenschappen, Amsterdam, **17**, 793–806 (1914)
36. Palm, C.: Intensitätsschwankungen im Fernsprechverkehr. Ericsson Technics, **44**, 1–189 (1943)
37. Parzen, E.: Stochastic Processes. Holden-Day, San Francisco (1962)
38. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2004). ISBN 3-900051-00-3
39. Ripley, B.D.: The second-order analysis of stationary point processes. J. Appl. Probab., **13**, 255–266 (1976)
40. Ripley, B.D.: Spatial Statistics. John Wiley & Sons, New York (1981)
41. Serra, J.: Image Analysis and Mathematical Morphology. Academic Press, London (1982)
42. Stoyan, D., Kendall, W.S., Mecke, J.: Stochastic Geometry and its Applications. Second edition, John Wiley & Sons, Chichester (1995)
43. Stoyan, D., Stoyan, H.: Fractals, Random Shapes and Point Fields. John Wiley & Sons, Chichester (1995)
44. Takacs, R.: Estimator for the pair potential of a Gibbsian point process. Statistics, **17**, 429–433 (1986)
45. Takacs, R., Fiksel, T.: Interaction pair-potentials for a system of ants' nests. Biom. J., **28**, 1007–1013 (1986)
46. van Lieshout, M.N.M.: Markov Point Processes and their Applications. Imperial College Press, London (2000)
47. van Lieshout, M.N.M, Baddeley, A.J.: A nonparametric measure of spatial interaction in point patterns. Statist. Neerlandica, **50**, 344–361 (1996)