5

Point processes III – Models

5.1 Operations on point processes

Model construction is a fundamental step in practical applications of stochastic geometry. It is explained here for the case of point processes; however, many of the modelling and simulation ideas discussed here can be used for other random structures.

This section describes three fundamental operations frequently used in modelling, which produce new point processes from old. Several important models can be derived from simpler ones (such as the Poisson process) by means of these operations. The resulting models, as well as some others, are described in the remainder of this chapter.

The three fundamental operations to be described in this section are:

- thinning,
- clustering,
- superposition.

Matthes *et al.* (1978), Kallenberg (2002) and Daley and Vere-Jones (2008) discuss limit theorems for the results of repeated use of these operations both separately and in combination. In this section only the operations themselves are described.

Thinning

A thinning operation uses some definite rule to delete points of a basic point process Φ_b, thus yielding the *thinned point process* Φ. Considered as a random closed set, Φ is a subset of Φ_b:

$$\Phi \subset \Phi_b. \tag{5.1}$$

A simple form of thinning is *p-thinning*: each point of Φ_b has probability $1 - p$ of suffering deletion, and its deletion is independent of locations and possible deletions of any other points of Φ_b. Thus p is the probability that a point is retained.

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

A natural generalisation allows the retention probability p to depend on the location x of the point. A deterministic function $p(x)$ is given on $\mathbb{R}^d$, with $0 \le p(x) \le 1$. If the point x belongs to Φ_b, it is deleted with probability $1 - p(x)$ and again its deletion is independent of locations and possible deletions of any other points. To emphasise the spatial dependence here, the generalised operation is called $p(x)$-*thinning*.

In a further generalisation the function $p(x)$ is itself random. Formally, a random field $\boldsymbol{\pi} = \{\pi(x) : x \in \mathbb{R}^d\}$ is given, which is independent of Φ_b, where $0 \le \pi(x) \le 1$ for all $x \in \mathbb{R}^d$. A realisation φ of the thinned process Φ is constructed by taking a realisation φ_b of Φ_b and applying $p(x)$-thinning to φ_b, where $\{p(x) : x \in \mathbb{R}^d\}$ is a realisation of the random field $\boldsymbol{\pi}$. Given $\pi(x) = p(x)$ and given $\Phi_b = \varphi_b$, the probability of x in Φ_b also belonging to Φ is $p(x)$. This further generalisation is known as $\pi(x)$-*thinning*.

All these thinnings are *independent thinnings*. This means there is no interaction between the points, so that the thinning functions (which are independent of Φ_b) determine the operation completely.

Yet another generalisation allows dependence on the configuration of Φ_b, giving the class of *dependent thinnings*. A practical context for a dependent thinning would be a point pattern representing plant locations. Suppose there is mutual inhibition between plants leading to the death of a fraction of them. This suggests a point process constructed by a thinning with deletion probability which is higher in regions of higher point density. This could be obtained, for example, by making deletion probability depend on the distance to the nearest point in the initial pattern.

Figure 5.1 demonstrates the result of independent and dependent thinnings applied to the same initial point pattern. An important dependent thinning leads to the Matérn hard-core processes; an example is considered in Section 5.4.

If the summary characteristics of the basic process Φ_b are known, then it is straightforward to calculate the characteristics of point processes produced by independent thinning. Indeed, if Φ is the result of $p(x)$-thinning of Φ_b, then its intensity measure Λ is given by

$$\Lambda(B) = \int_B p(x)\Lambda_b(\mathrm{d}x), \tag{5.2}$$

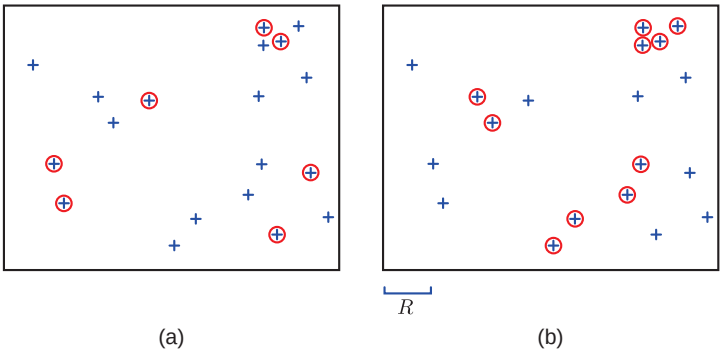

Figure 5.1 Thinnings of a point pattern. Points of the origin pattern are marked by $+$, deleted points by $\oplus$. (a) Independent thinning carried out by a p-thinning; (b) dependent thinning deleting all points with nearest neighbour within distance R.

where Λ_b is the intensity measure of Φ_b. If G_b is the generating functional of Φ_b, then that of Φ is

$$G(v) = G_b(v_p) \qquad \text{for } v \in \boldsymbol{V}, \tag{5.3}$$

where

$$v_p(x) = v(x)p(x) + 1 - p(x); \tag{5.4}$$

$\boldsymbol{V}$ is the set of functions defined in Section 4.3.6.

Analogous formulae for $\pi(x)$-thinnings follow by averaging with respect to the distribution of the random field $\boldsymbol{\pi}$.

If Φ_b is stationary, then the p-thinned process Φ is also stationary; in general this is false for $p(x)$-thinning. From Formula (5.2) the intensity is given by

$$\lambda = p\lambda_b. \tag{5.5}$$

A $\pi(x)$-thinned point process can be stationary: if Φ is a stationary point process and $\boldsymbol{\pi} = \{\pi(x)\}$ a stationary random field. Formula (5.5) still applies with $p = \mathbf{E}\big(\pi(x)\big)$.

When both Φ_b and $\boldsymbol{\pi}$ are motion-invariant so is Φ. The existence of a second-order product density $\varrho_b^{(2)}(r)$ for Φ_b implies the same for Φ. The corresponding second-order characteristics are as follows: in the case of a p-thinning the product density is given simply by

$$\varrho^{(2)}(r) = p^2\varrho_b^{(2)}(r), \tag{5.6}$$

and the reduced second moment functions as well as the pair correlation functions of Φ and Φ_b coincide.

In the case of a $\pi(x)$-thinning let $k(r)$ denote the covariance function of the random field $\boldsymbol{\pi}$ by

$$k(r) = \mathbf{E}\big(\pi(o)\pi(\mathbf{r})\big) - p^2 \qquad \text{where } \|\mathbf{r}\| = r. \tag{5.7}$$

Then

$$\varrho^{(2)}(r) = \big(k(r) + p^2\big)\,\varrho_b^{(2)}(r), \tag{5.8}$$

and the reduced second moment function $K(r)$ of Φ is given by

$$K(r) = \int_0^r \Big(\big(k(x) + p^2\big)\,\mathrm{d}K_b(x)\Big) \Big/ p^2. \tag{5.9}$$

If $\varrho_b^{(2)}(r)$ is interpreted intuitively as in Section 4.3.3, then Formulae (5.6) and (5.8) are plausible: if a pair of points are both to contribute to the product density, then both of them must survive.

It is easy to describe Φ if Φ_b is a *Poisson process*. Let Φ_b be Poisson with intensity measure Λ_b. Under a $p(x)$-thinning Φ is again Poisson with intensity measure given by (5.2). This result (sometimes called *Prekopa's Theorem*) follows easily from (5.3) and the formula

for the generating functional of a Poisson process. Alternatively, one can compute the void-probabilities of a p-thinned Poisson process:

$$\begin{aligned}\mathbf{P}(\Phi(K)=0) &= \sum_{k=0}^{\infty} \mathbf{P}(\Phi_{\mathrm{b}}(K)=k)\cdot \mathbf{P}(\text{all these } k \text{ points are deleted})\\ &= \sum_{k=0}^{\infty} \mathrm{e}^{-\Lambda_{\mathrm{b}}(K)}\frac{\Lambda_{\mathrm{b}}(K)^k}{k!}(1-p)^k = \exp\bigl(-p\Lambda_{\mathrm{b}}(K)\bigr).\end{aligned}$$

In particular if Φ_{b} is a homogeneous Poisson process of intensity λ_{b}, then its p-thinning is a homogeneous Poisson process of intensity $p\lambda_{\mathrm{b}}$. Note that the points which are *removed* by this thinning also form a Poisson process, and the two processes are independent.

The $\pi(x)$-thinnings of Poisson processes are doubly stochastic Poisson processes (i.e. Cox processes) as described in Section 5.2. Multiple thinnings combined with rescalings lead to Cox processes; see Daley and Vere-Jones (2008, Section 11.3).

Example 5.1. *An interrupted point process: Lightning gap locations (Kautz et al., 2011)*
Figure 5.2 shows the locations of 78 gaps in the canopy of a Vietnamese mangrove forest resulting from lighting strikes, together with the distribution of the forest matrix, interrupted

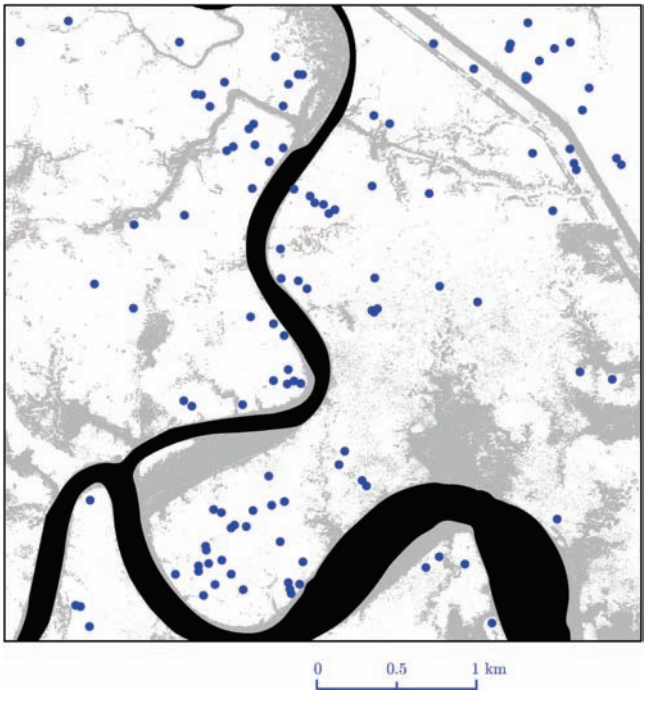

Figure 5.2 Positions of 78 lightning gaps (marked by •) in a 4 km × 4 km sampling window of a Vietnamese mangrove forest (white regions) with rivers (black) and open (grey) areas such as water channels and mud flats, generated in the years 2003–2007. Reprinted from Kautz *et al.* (©2011) with permission of Elsevier.

by water channels and mud flats. Lightning strikes leave traces only in woodland. Thus here a quite natural process of thinning is acting: the observed gap locations are taken to be result of thinning a motion-invariant point process Φ_b (representing 'all lightning strikes') by using the indicator of a motion-invariant random closed set Ξ (representing the woodland) as the thinning random field $\boldsymbol{\pi}$. In the notation of Stoyan (1979a) the resulting point process is an *interrupted point process*; it is a $\pi(x)$-thinning with $\pi(x) = \mathbf{1}_\Xi(x)$. It can also be written as

$$\Phi = \Phi_b \cap \Xi.$$

Let Φ_b have intensity λ_b and pair correlation function $g_b(r)$ and let Ξ have area fraction p and covariance $C(r)$. Note that the covariance function $k(r)$ of the random field $\{\mathbf{1}_\Xi\}$ is given by

$$k(r) = C(r) - p^2.$$

From the formulae above the intensity of Φ is

$$\lambda = p\lambda_b, \tag{5.10}$$

and its pair correlation function is

$$g(r) = \frac{C(r)}{p^2}\, g_b(r) \qquad \text{for } r \geq 0. \tag{5.11}$$

Naturally this model is only an approximation, both by assuming motion-invariance and independence of lightning and woodland. Kautz *et al.* (2011) confirmed the latter assumption by a test and analysed statistically the data further. Figure 5.3 displays estimates of the pair

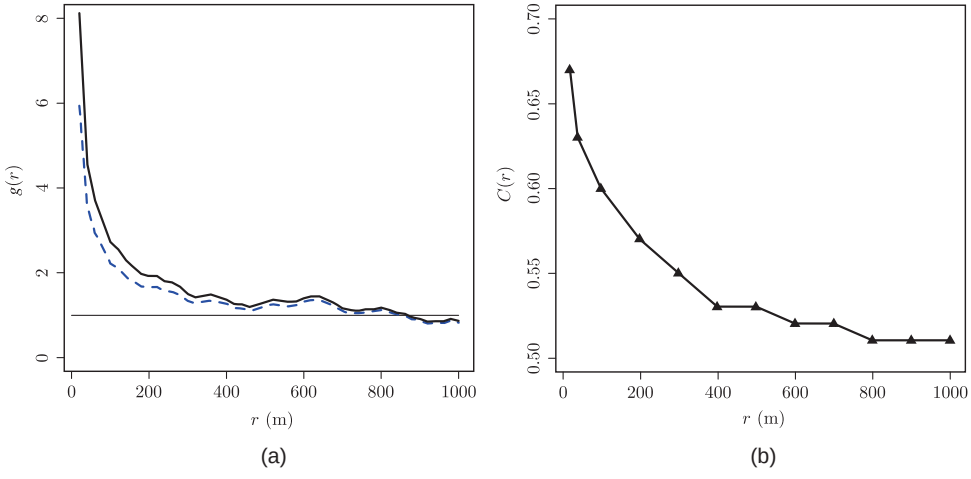

Figure 5.3 (a) The empirical pair correlation function $\hat{g}(r)$ (solid line) for the point pattern of Figure 5.2 and the pair correlation function (dashed line) of the full lightning strike process Φ_b obtained by Formula (5.11). (b) The estimated covariance $\hat{C}(r)$ of the forest matrix of Figure 5.2. Reprinted from Kautz *et al.* (©2011) with permission of Elsevier.

correlation function $g(r)$ and covariance $C(r)$. The estimated first-order statistics are:

$$\begin{aligned}\hat{p} &= 0.7,\\ \hat{\lambda} &= 4.9\,\mathrm{km}^{-2},\\ \hat{\lambda}_\mathrm{b} &= 7.0\,\mathrm{km}^{-2}.\end{aligned}$$

Formula (5.11) yields the graph of $\hat{g}_\mathrm{b}(r)$ shown in Figure 5.3, which is similar to pair correlation functions of cluster processes. Thus, it is reasonable to assume that the lightning strikes are spatially clustered, and Kautz *et al.* (2011) proposed a suitable process with clustering both in time and space to model the given pattern.

Clustering

In a clustering operation each point x of a given point process Φ_p (the subscript 'p' stands for 'parent') is replaced by a cluster N^x of daughter points. The replacement clusters N^x are themselves point processes, which have only a finite number of points each. The union of all these clusters is the *cluster point process* Φ,

$$\Phi = \bigcup_{x\in\Phi_\mathrm{p}} N^x. \tag{5.12}$$

It is tacitly assumed that almost certainly points of different clusters do not coincide, so that $N^x \cap N^y$ is empty whenever $x \neq y$. Furthermore, it is assumed that Φ is locally finite. Depending on the particular model, x may or may not itself belong to N^x. Figure 5.4 shows a sample of a cluster point process.

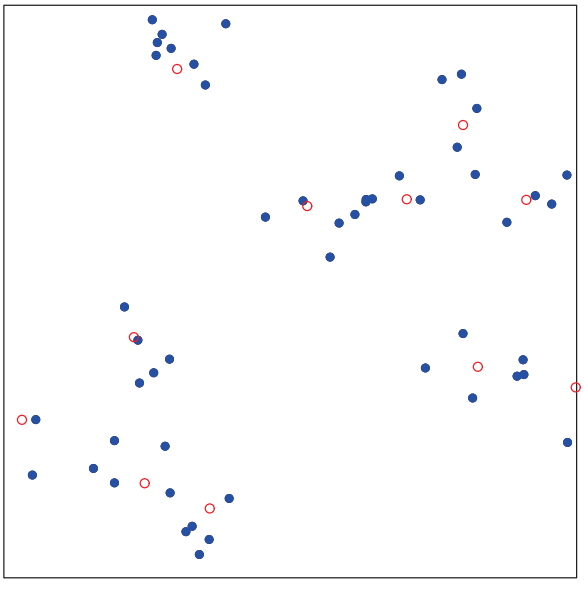

Figure 5.4 A sample of a cluster point process. Parent points are marked by ○, daughter points by ●.

Note in passing that thinning operations can formally be interpreted as special cases of cluster operations: a thinning can be obtained by supposing N^x to be either the singleton set $\{x\}$ or the empty set, depending on the thinning criterion evaluated at x. Random shifts can be interpreted similarly: i.i.d. shifted points are daughter points of a given pattern of parent points; in this case all N^x are i.i.d. singletons.

Cluster processes have been used as models for many natural phenomena, as documented by the references in Section 5.3. For example, parents are locations of plants and daughters are the seedfall locations; the cluster pattern is generated as the locations where new plants grow. Often cluster processes serve as point process models of a higher degree of variability than Poisson processes.

An important special case is the Neyman–Scott process, which is described in detail in Section 5.3. The formulae in the rest of this section are given for a somewhat more general model. It is supposed that the parent point process $\Phi_{\mathrm{p}} = \{x_1, x_2, \ldots\}$ is stationary of intensity λ_{p} and that the clusters N^x are of the form

$$N^{x_i} = N_i + x_i \tag{5.13}$$

for each x_i in Φ_{p}. The N_i are i.i.d. finite point sets with the common distribution $\boldsymbol{c}$, independent of the parent point process Φ_{p}, and the $+ x_i$ terms shift the cluster centres to x_i. This is called *homogeneous independent clustering*. If Φ_{p} is a Poisson process, then the resultant process Φ is called a *Poisson cluster process*.

Whatever the form of Φ_{p} in this model is, the intensity λ of Φ is given by

$$\lambda = \lambda_{\mathrm{p}} \bar{c} \tag{5.14}$$

where $\bar{c}$ is the mean number of points in the cluster N_0, which has the same distribution as the N_i and is called the *representative cluster* of the cluster process. Clearly, $\bar{c}$ and λ_{p} of Φ_{p} are assumed to be finite.

A general formula for the second-order product density $\varrho^{(2)}(x_1, x_2)$ can be found in Illian *et al.* (2008, p. 370).

If G, G_{c} and G_{p}, denote the generating functionals of Φ, N_0 and Φ_{p} respectively, then

$$G(v) = G_{\mathrm{p}}\left(G_{(\cdot)}(v)\right) = \mathbf{E}\left(\prod_{x \in \Phi_{\mathrm{p}}} G_{(x)}(v)\right) \qquad \text{for } v \in \mathbf{V}, \tag{5.15}$$

where

$$G_{(x)}(v) = G_{\mathrm{c}}\big(v(x + \cdot)\big) = \mathbf{E}\left(\prod_{y \in N_0} v(x + y)\right). \tag{5.16}$$

Formula (5.15) expresses a close similarity to the theory of branching processes, and indeed the clustering operation is precisely what is used to create new generations of a *branching random walk* or *spatial branching process*; see Biggins (1977), Mollison (1977), Matthes *et al.* (1978) and Liemant *et al.* (1988). Biological applications can be found in Shimatani (2002, 2010).

Superposition

Let Φ_1 and Φ_2 be two point processes with distributions P_1 and P_2, intensities λ_1 and λ_2, etc. Consider the union

$$\Phi = \Phi_1 \cup \Phi_2. \tag{5.17}$$

Suppose that with probability one the point sets Φ_1 and Φ_2 do not overlap (in the case of independence this is equivalent to the corresponding intensity measures having no common atoms; see Daley and Vere-Jones, 2008.) The set-theoretic union then coincides with the superposition operation of general point process theory. In the general theory superpositions may have multiple points, a complication not considered here.

It is clear that the intensity measure Λ and intensity λ (in case of stationarity) of Φ are given by

$$\Lambda = \Lambda_1 + \Lambda_2, \tag{5.18}$$

$$\lambda = \lambda_1 + \lambda_2. \tag{5.19}$$

If Φ_1 and Φ_2 are independent, then the distribution P of Φ is written by means of the convolution symbol $*$ as

$$P = P_1 * P_2. \tag{5.20}$$

The generating functional G of Φ satisfies

$$G(v) = G_1(v) \cdot G_2(v) \qquad \text{for } v \in \mathbf{V}, \tag{5.21}$$

and the second-order product density $\varrho^{(2)}$ is

$$\varrho^{(2)}(x_1, x_2) = \varrho_1^{(2)}(x_1, x_2) + \varrho_2^{(2)}(x_1, x_2) + 2\lambda_1\lambda_2 \qquad \text{for } x_1 \text{ and } x_2 \in \mathbb{R}^d, \tag{5.22}$$

which leads easily to formulae for the pair correlation function $g(r)$ and the reduced second moment function $K(r)$; see Illian *et al.* (2008, p. 371).

The nearest-neighbour distance distribution function of Φ is given by

$$1 - D(r) = \frac{\lambda_1}{\lambda}\big(1 - D_1(r)\big)\big(1 - H_{s,2}(r)\big) + \frac{\lambda_2}{\lambda}\big(1 - D_2(r)\big)\big(1 - H_{s,1}(r)\big) \qquad \text{for } r \geq 0, \tag{5.23}$$

see van Lieshout and Baddeley (1996), where $H_{s,i}(r)$ is the spherical contact distribution function of Φ_i. The spherical contact distribution function of Φ is

$$H_s(r) = 1 - \big(1 - H_{s,1}(r)\big)\big(1 - H_{s,2}(r)\big) \qquad \text{for } r \geq 0. \tag{5.24}$$

If Φ_1 and Φ_2 are independent Poisson processes, then so is Φ. This follows directly from the fact that the sum of two independent Poisson variables is itself Poisson.

Multiple superpositions combined with rescalings lead to Poisson processes; see Daley and Vere-Jones (2008, Section 11.2).

5.2 Doubly stochastic Poisson processes (Cox processes)

5.2.1 Introduction

An obvious generalisation of a Poisson process is made by supposing that the intensity measure is itself random, with the point process being Poisson *conditional* on the realisation of the intensity measure. Such a process is called a *doubly stochastic Poisson process* or *Cox process*.

Formally a Cox process is defined by choosing a distribution Q on $[\mathbb{M}, \mathcal{M}]$, the space of all nonnegative locally finite measures on $\mathbb{R}^d$. Suppose P_Λ is the distribution of the Poisson process of intensity measure Λ, and Ψ is a random measure with distribution Q. (See Chapter 7 for elements of the theory of random measures.) Then the *Cox process* Φ *with driving random measure* Ψ has distribution

$$P_\Phi(Y) = \int P_\Lambda(Y) Q(\mathrm{d}\Lambda) \qquad \text{for } Y \in \mathcal{N}. \tag{5.25}$$

To ensure that Φ is simple, that is, that Φ has no multiple points, it is assumed that Q is concentrated on the subset of diffuse measures.

A Cox process can be thought of as arising from a two-step random mechanism, hence the term 'doubly stochastic'. The first step generates a measure Λ on $\mathbb{R}^d$ according to the driving random measure distribution Q and the second step generates a Poisson process of intensity measure Λ.

Cox processes are both general and amenable to calculation; therefore they find important applications as stochastic models. They occur frequently in physics, for example, in optics. Systematic study began with Cox (1955), after earlier examples. Mecke (1968) and Krickeberg (1972) made important early contributions, Grandell (1976) is an excellent early monograph on the subject; further material is to be found in all books on point processes, for example Daley and Vere-Jones (2003, 2008), Møller and Waagepetersen (2004) and Illian *et al.* (2008).

Early applications of Cox processes in spatial statistics are discussed by Matérn (1971), Bartlett (1975) and Grandell (1981). Diggle (1983), Diggle and Milne (1983) and Lotwick (1984) considered a marked Cox process as a model for a pattern of points of two types, such as locations of two different species of trees. Positive and negative dependence are combined in this model. Kingman (1977) showed that Cox processes are more natural candidates than Poisson processes for the distribution of reproducing populations in space.

An interesting classical example is given by Stapper *et al.* (1980), who used a mixed Poisson process (see the next section) to model the pattern of point defects on the surfaces of silicon wafers arising in semiconductor manufacture. (The randomised intensity is given by a gamma distribution.) Modern developments are presented in Kuo *et al.* (2006).

Statistical problems for Cox processes have been studied in many papers. Krickeberg (1982) and Karr (1984, 1986) made early major contributions. The books by Møller and Waagepetersen (2004) and Illian *et al.* (2008) describe the modern approach.

5.2.2 Examples of Cox processes

The mixed Poisson process

The *mixed Poisson process* is a simple instance of a Cox process. It can be thought of as a homogeneous Poisson process with randomised intensity parameter. The driving random measure Ψ is the random measure

$$\Psi = X\nu_d, \tag{5.26}$$

where the randomised intensity X is a nonnegative random variable. Every sample of such a process looks like a sample of some homogeneous Poisson process. This is a typical example of a non-ergodic Cox process.

$\pi(x)$-thinning of a Poisson process

In this case the driving random measure Ψ is given by

$$\Psi(B) = \int_B \pi(x)\Lambda_{\mathrm{b}}(\mathrm{d}x) \qquad \text{for Borel } B, \tag{5.27}$$

where Λ_{b} is the intensity measure of the original Poisson process Φ_{b}.

The Log-Gaussian Cox process

A random field $\{Z(x)\}$ is Gaussian if, for any finite collection of locations $x_1, \ldots, x_k$, any linear combination $b_1 Z(x_1) + \cdots + b_k Z(x_k)$ with real $b_1, \ldots, b_k$ has a one-dimensional normal distribution. If $\{Z(x)\}$ is stationary and isotropic, its distribution is determined completely by the mean μ_Z and the covariance function $k_Z(r)$. However, such a field cannot be used as the intensity field of a Cox process since it can take negative values. Thus, a suitable transformation has to be applied to the field $\{Z(x)\}$ to yield a Cox process. An elegant transformation, resulting in a mathematically tractable model, is

$$\Lambda(x) = \exp\bigl(Z(x)\bigr) \qquad \text{for } x \in \mathbb{R}^d. \tag{5.28}$$

The corresponding process is termed a *log-Gaussian Cox process.* It was independently introduced by Coles and Jones (1991), Rathbun (1996) and Møller *et al.* (1998) and is now a very popular model, which has a simple formula for all product densities; see Example 5.3. It can be seen as a link between point process statistics and geostatistics.

Random-set-generated Cox process

A regular-closed stationary random closed set Ξ divides $\mathbb{R}^d$ into two parts or phases: the set Ξ and its complement Ξ^{c}. In both phases, a Poisson process is generated with intensities λ_1 and λ_2, respectively.

In so-called 'ideal cluster materials' there exist deterministic or random ellipsoidal domains ('particle clouds') with high concentration of nuclei/reinforcements within and low without; see Buryachenko (2007, pp. 3 and 151). In another application the random-set-generated Cox process was used for a data set of a pattern of seedling locations in a commercial tree plantation, where the soil had been treated in two different ways. The random set reflects these two different soil treatments (Penttinen and Niemi, 2007). The interrupted process of Example 5.1 is a special case with $\lambda_1 = \lambda$ and $\lambda_2 = 0$. The random intensity of this Cox process is simply

$$\Lambda(x) = \lambda_1 \mathbf{1}_\Xi(x) + \lambda_2\bigl(1 - \mathbf{1}_\Xi(x)\bigr), \tag{5.29}$$

where $\mathbf{1}_\Xi(x)$ denotes the indicator function of the set Ξ.

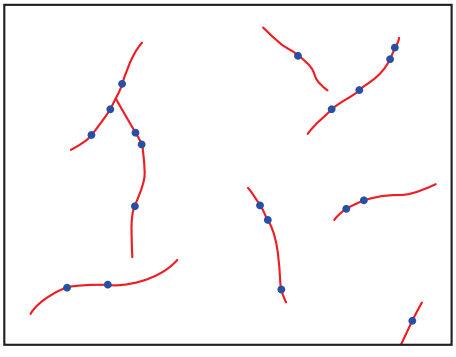

Figure 5.5 A sample of a particular Cox process. The points are randomly scattered on a system of random fibres. Compare this with Hodder and Orton (1976, Figure 7.1), which shows the distribution of finds of late Iron Age inscribed coins and Roman roads in central and southern England. Many coin finds lie on or very close to these roads. In another interpretation the points here can be considered as centres of nucleation.

A closely related class of Cox processes are *shot-noise Cox processes* as discussed in Møller and Waagepetersen (2004). Neyman–Scott processes belong to this class if the number of points per cluster follows a Poisson distribution; an example is the Matérn cluster process on p. 173.

Points scattered on random fibres and networks

Let χ be a random fibre process (see Section 8.3). The random measure Ψ defined by

$$\Psi(B) = N_L h_1(\chi \cap B) \qquad \text{for Borel } B \tag{5.30}$$

yields a driving measure for a Cox process. Here N_L is a positive constant and $h_1(\chi \cap B)$ is the total length of all fibre pieces of χ in B. The resultant Cox process is formed of points scattered uniformly with intensity N_L on the fibres of χ, as illustrated in Figure 5.5. The parameter N_L is the mean number of points per unit length of fibres.

Such Cox processes appear in materials science in the context of nucleation on fibres (or, similarly, on surfaces), see Section 6.6.4, and in telecommunication networks, see Voss *et al.* (2010). There χ is the edge system of a random tessellation, representing the infrastructure (roads, railways), while the points are for example mobile users in wireless networks. In another approach Ang *et al.* (2012) consider random point patterns on deterministic/fixed networks, for example crime locations on the streets of Chicago. They make second-order statistics based on travel distances in the network. This falls outside of the usual domain of Cox point process theory.

5.2.3 Formulae for characteristics of Cox processes

The generating functional G of the Cox process Φ_C is given by

$$G(v) = L_Q(1 - v) \qquad \text{for } v \in \mathbf{V}, \tag{5.31}$$

where L_Q is the Laplace functional of the driving random measure Ψ of Φ. This follows from

$$\begin{aligned} G(v) &= \int_{\mathbb{M}} G_\Lambda(v) Q(\mathrm{d}\Lambda) \\ &= \int_{\mathbb{M}} \exp\left(- \int_{\mathbb{R}^d} \big(1 - v(x)\big) \Lambda(\mathrm{d}x) \right) Q(\mathrm{d}\Lambda) \\ &= L_Q(1 - v), \end{aligned}$$

where G_Λ is the generating functional of the Poisson process of intensity measure Λ.

The void-probabilities of a Cox process are given by

$$\begin{aligned} v_K &= \int_{\mathbb{M}} \exp\big(-\Lambda(K)\big) Q(\mathrm{d}\Lambda) \\ &= L_Q(\mathbf{1}_K) \qquad \text{for compact } K, \end{aligned} \tag{5.32}$$

since

$$\begin{aligned} v_K &= \mathbf{P}(\Phi \cap K = \emptyset) = \mathbf{P}\big(\Phi(K) = 0\big) \\ &= \mathbf{E}\big(\exp(-\Psi(K))\big) = \int_{\mathbb{M}} \exp\big(-\Lambda(K)\big) Q(\mathrm{d}\Lambda). \end{aligned}$$

The factorial moment measures $\alpha^{(n)}$ of Φ coincide with the moment measures $\mu_\Psi^{(n)}$ of Ψ,

$$\alpha^{(n)} = \mu_\Psi^{(n)} \qquad \text{for } n = 1, 2, \ldots, \tag{5.33}$$

and, in particular, the intensity measures are equal,

$$\Lambda = \Lambda_\Psi. \tag{5.34}$$

If Ψ is stationary, then so is Φ, and both have the same intensity parameter. In particular, every mixed Poisson process is stationary.

The Palm distributions P_x for Φ are given by

$$P_x = \delta_{\delta_x} * P_{Q_x}, \tag{5.35}$$

where P_{Q_x} is the distribution of a Cox process with driving random measure Q_x, and Q_x is the Palm distribution of Ψ at x. Note that Palm distributions for random measures are defined in Section 7.1.3.

Formula (5.35) is a good starting point for calculating the reduced second moment function $K(r)$ and the nearest-neighbour distance distribution function $D(r)$.

For all Cox processes the number of counts in bounded sets is *super-Poissonian*. That is to say, their variances exceed those for homogeneous Poisson processes of the same intensity. This is a consequence of the mixed Poisson distribution of the number of counts in a bounded set.

Example 5.2. *Mixed Poisson processes using two intensities λ_1 and λ_2*

Consider a mixed Poisson process Φ as above, with the randomised intensity X having a two-point distribution

$$X = \begin{cases} \lambda_1, & \text{with probability } 1-p, \\ \lambda_2, & \text{with probability } p, \end{cases}$$

where $0 < p < 1$. Its intensity is given by

$$\lambda = (1-p)\lambda_1 + p\lambda_2.$$

Its Palm distribution P_o is derived from the Palm distributions $P_o^{(1)}$ and $P_o^{(2)}$ of Poisson processes with intensities λ_1 and λ_2 respectively:

$$P_o(Y) = \frac{1}{\lambda}\left(\lambda_1(1-p)P_o^{(1)}(Y) + \lambda_2 p P_o^{(2)}(Y)\right) \qquad \text{for } Y \in \mathcal{N}.$$

Consequently the nearest-neighbour distance distribution function is given by

$$D(r) = \frac{1}{\lambda}\Big(\lambda_1(1-p)\left(1-\exp(-\lambda_1 b_d r^d)\right) + \lambda_2 p\left(1-\exp(-\lambda_2 b_d r^d)\right)\Big).$$

The reduced second moment function K of Φ is

$$K(r) = \frac{\lambda_1^2(1-p)+\lambda_2^2 p}{\lambda^2}\, b_d r^d \qquad \text{for } r \geq 0.$$

Example 5.3. *Log-Gaussian Cox processes*

Recall that for a log-Gaussian Cox processes the random intensity is $\Lambda(x) = \exp\big(Z(x)\big)$, where $\{Z(x)\}$ is a stationary and isotropic Gaussian random field with mean μ_Z, variance σ_Z^2 and covariance function $k_Z(r)$. Then the first- and second-order point process characteristics are

$$\lambda = \exp\left(\mu_Z + \frac{\sigma_Z^2}{2}\right) \tag{5.36}$$

and

$$g(r) = \exp\big(k_Z(r)\big) \qquad \text{for } r \geq 0. \tag{5.37}$$

Moreover, the n^{th}-order product density can be expressed in terms of λ and $g(r)$ as

$$\varrho^{(n)}(x_1, \ldots, x_n) = \lambda^n \prod_{1 \leq i < j \leq n} g(r_{ij}), \tag{5.38}$$

where $r_{ij} = \|x_i - x_j\|$; see Møller *et al.* (1998).

Example 5.4. *Random-set-generated Cox processes*

The intensity and pair correlation function of a random-set-generated Cox process are obtained from the first- and second-order characteristics of a stationary and isotropic random set Ξ, namely, the area fraction (volume fraction) p and covariance $C(r)$; see Section 6.3. These

two characteristics, together with the Poisson process intensities λ_1 and λ_2 of the two phases, lead to

$$\lambda = p\lambda_1 + (1-p)\lambda_2 \tag{5.39}$$

and

$$g(r) = \frac{1}{\lambda^2}\left(C(r)(\lambda_1-\lambda_2)^2 - 2p(\lambda_1-\lambda_2)\lambda_2 + \lambda_2^2\right) \qquad \text{for } r \geq 0. \tag{5.40}$$

5.3 Neyman–Scott processes

Neyman–Scott processes (also called *centre–satellite processes*) are examples of Poisson cluster processes, which are often used in spatial statistics. They result from homogeneous independent clustering applied to a homogeneous Poisson process. The parent points form a homogeneous Poisson process of intensity λ_{p} while the daughter points of a representative cluster N_0 are random in number and are scattered independently with identical spatial probability density $f(y)$ around the origin. The parent points do not occur in the observed point pattern; only daughter points are included.

Under these assumptions the resultant cluster process Φ is stationary. If the scattering probability density function $f(y)$ is radially symmetric, then Φ is isotropic.

The Neyman–Scott process can be interpreted as a particular kind of Boolean model, where the clusters play the rôle of the 'grains'. This is useful for the derivation of formulae for void-probabilities.

The general Formula (5.14) for the intensity λ of cluster processes applies here:

$$\lambda = \lambda_{\mathrm{p}}\bar{c}, \tag{5.41}$$

where $\bar{c}$ is the mean number of daughter points per parent. The generating functional in Formula (5.15) specialises to

$$G(v) = \exp\left(-\lambda_{\mathrm{p}}\int_{\mathbb{R}^d}\left(1 - G_n\left(\int_{\mathbb{R}^d} v(x+y)f(y)\,\mathrm{d}y\right)\right)\mathrm{d}x\right), \tag{5.42}$$

where G_n is the generating function for the random number of points of N_0.

Further progress can be made by considering a fundamental formula for the Palm distribution P_o of a Poisson cluster process. Let P be the distribution of Φ and $\boldsymbol{c}_o$ be the 'Palm' distribution of the cluster N_0. Then

$$P_o = P * \boldsymbol{c}_o \tag{5.43}$$

as given by Ambartzumian (1966) and Mecke (1967). The Palm distribution $\boldsymbol{c}_o$ is given by

$$\boldsymbol{c}_o(Y) = \frac{1}{\bar{c}}\,\mathbf{E}\left(\sum_{x\in N_0}\mathbf{1}_Y(N_0 - x)\right) \qquad \text{for } Y \in \mathcal{N}, \tag{5.44}$$

which is the formula for the 'Palm' distribution of a point process with finite mean total number of points.

Formula (5.43) can be seen as plausible if it is considered as a generalisation of the Slivnyak–Mecke theorem for the Palm distribution of a homogeneous Poisson process. The resulting Palm distribution P_o for Φ may be interpreted as the superposition of a finite point process following the Palm distribution $\boldsymbol{c}_o$ for N_0 and an independent point process with the distribution P of the complete Φ.

The distribution $\boldsymbol{c}_o$ can be understood as follows. Let $\{p_n\}$ be the distribution of the number of points in the representative cluster N_0 so that

$$p_n = \mathbf{P}\big(N_0(\mathbb{R}^d) = n\big).$$

Then

$$\overline{c} = \sum_{n=0}^{\infty} n p_n.$$

Simple calculations with conditional probabilities yield $\boldsymbol{c}_o(Y_k)$ for $Y_k = \{\varphi \in \mathbb{N} : \varphi(\mathbb{R}^d) = k\}$, that is, the 'Palm' probability of the event that there are precisely k points in the cluster N_0. It is

$$\pi_0 = \boldsymbol{c}_o(Y_0) = 0, \tag{5.45}$$

$$\pi_k = \boldsymbol{c}_o(Y_k) = \frac{1}{\overline{c}}\,\mathbf{E}\left(\sum_{x\in N_0} \mathbf{1}_{Y_k}(N_0 - x)\right) = \frac{1}{\overline{c}}\,k p_k \qquad \text{for } k = 1, 2, \ldots. \tag{5.46}$$

The reason that $\pi_0 = 0$ is clear: a cluster which contains the typical point cannot be empty. Thus the distribution $\boldsymbol{c}_o$ is *number-weighted* and in general different from the distribution of the representative cluster N_0.

Formula (5.43) allows other characteristics to be expressed in terms of quantities that are related to Palm distributions of Neyman–Scott processes. In the isotropic case the reduced second moment function $K(r)$ satisfies

$$K(r) = b_d r^d + \frac{1}{\lambda \overline{c}} \sum_{n=2}^{\infty} p_n n(n-1)\boldsymbol{F}(r) \qquad \text{for } r \geq 0, \tag{5.47}$$

where $\boldsymbol{F}(r)$ is the distribution function of the distance between two independent random points of the same cluster.

From Formula (5.47) the pair correlation function $g(r)$ is given by

$$g(r) = 1 + \frac{1}{\lambda \overline{c}} \sum_{n=2}^{\infty} p_n n(n-1) \frac{\boldsymbol{f}(r)}{d b_d r^{d-1}} \qquad \text{for } r \geq 0, \tag{5.48}$$

where $\boldsymbol{f}(r)$ denotes the probability density function of $\boldsymbol{F}(r)$. It is obtained from the spatial probability density $f(y)$ of the cluster points as follows: the probability density function of the vector difference of two independent points of the same cluster is

$$\boldsymbol{f}(\mathbf{r}) = \int f(y) f(y + \mathbf{r}) \mathrm{d}y,$$

and changing to polar coordinates ($\mathbf{r} \to r$) yields $\boldsymbol{f}(r)$.

Formula (5.43) also leads to the nearest-neighbour distance distribution function $D(r)$:

$$1 - D(r) = \mathbf{P}\big(\Phi(B(o,r)) = 0\big) \cdot \boldsymbol{c}_o\big(\{\varphi \in \mathbb{N} : \varphi\big(B(o,r)\big) = 1\}\big) \qquad \text{for } r \geq 0, \tag{5.49}$$

where the first factor on the right-hand side is equal to $1 - H_s(r)$, that is, given by the spherical contact distribution function; see Stoyan and Stoyan (1994, p. 312). The spherical contact distribution function $H_s(r)$ of the cluster process can be obtained by means of Formula (3.30) using the fact that Φ can be interpreted as a Boolean model with typical grain N_0. Unfortunately, the use of these formulae for concrete models is not easy.

Clearly the number of points in a compact convex set is super-Poissonian for a Neyman–Scott process; see Stoyan (1983). That is, its variance is greater than that for a Poisson process of the same intensity. A similar inequality holds for the corresponding K-functions.

Example 5.5. *Matérn cluster process (Matérn, 1960, 1986)*

In this example, the representative cluster N_0 has a distribution as follows. The number of points in N_0 has a Poisson distribution with the positive parameter $\mu = \bar{c}$. The sum in (5.47) and (5.48) is then $\bar{c}^2$. The points of N_0 are independently uniformly scattered in the ball $B(o, R)$, where R is a further model parameter. Hence

$$\lambda = \lambda_{\mathrm{p}}\mu, \tag{5.50}$$

and the probability density function $\boldsymbol{f}(r)$ occurring in Formula (5.48) for $g(r)$ is of the form

$$\boldsymbol{f}(r) = \begin{cases} \dfrac{1}{R}\left(1 - \dfrac{r}{2R}\right), & \text{if } d = 1, \\[2ex] \dfrac{4r}{\pi R^2}\left(\cos^{-1}\dfrac{r}{2R} - \dfrac{r}{2R}\sqrt{1 - \dfrac{r^2}{4R^2}}\right), & \text{if } d = 2, \\[2ex] \dfrac{3}{2}\dfrac{r^2}{R^6}\left(R - \dfrac{r}{2}\right)^2\left(2R + \dfrac{r}{2}\right), & \text{if } d = 3, \end{cases} \tag{5.51}$$

for $0 < r < 2R$; otherwise $\boldsymbol{f}(r) = 0$.

These results can be found already in Santaló (1976, p. 212, Note 6).

A sample of a simulated Matérn cluster process is shown in Figure 5.6; Figure 4.7 shows the pair correlation function for particular parameters.

Variations on the scatter distribution produce other models. If in the construction above the uniform distribution in the disc is replaced by the isotropic d-dimensional normal distribution with variance parameter σ, then the result is the *(modified) Thomas process*. For this model the pair correlation function is

$$g(r) = 1 + \frac{\bar{c}}{\lambda}(4\pi\sigma^2)^{-d/2}\exp\left(-\frac{r^2}{4\sigma^2}\right) \qquad \text{for } r \geq 0 \tag{5.52}$$

(see Møller and Waagepetersen, 2004). Here $\mu = \bar{c}$ is the mean of the Poisson distribution of cluster size and σ^2 is the variance of the symmetric normal distribution. Both the Matérn and Thomas process can be generalised by replacing the Poisson distribution of the random number of cluster points by any other discrete distribution; the formulae for $\boldsymbol{f}(r)$ can be

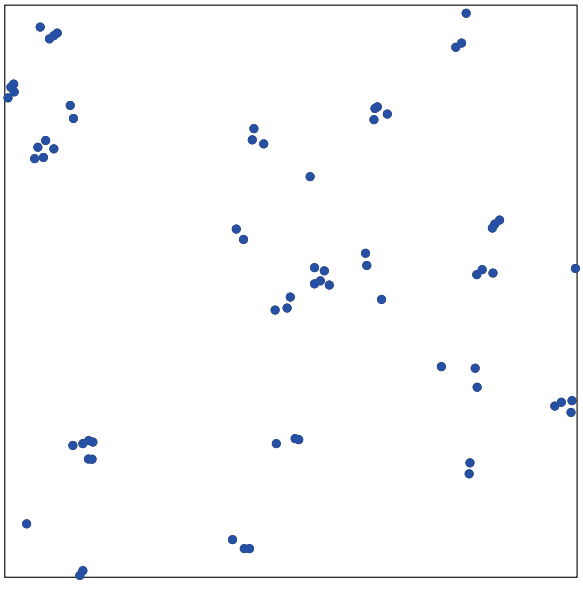

Figure 5.6 A simulated realisation of a Matérn cluster process in a unit square with $\lambda = 25$, $R = 0.025$ and $\mu = 2$.

plugged into the general formula (5.48). Tanaka *et al.* (2008) generalise the Thomas process model by introducing two types of clusters with σ_1^2 and σ_2^2.

Buryachenko (2007, p. 151) discusses the application of cluster processes in materials science and writes that processes with convex clusters are useful in the context of recrystallisation and residual stresses. Illian *et al.* (2008) give examples for the application of cluster processes in forestry, where sometimes it is necessary to use more general models than the Thomas process; see also Example 2.1.

A simple subclass of Poisson cluster processes is formed by the Gauss-Poisson processes. Milne and Westcott (1972) discuss this subclass, in which each cluster consists of zero, one, or two points only.

Example 5.6. *Calculation of $D(r)$ for a planar Gauss-Poisson process*

In this example the typical cluster N_0 has an isotropic distribution, being composed of zero, one or two points with probability p_0, p_1 or p_2 respectively. If N_0 consists of only one point, then that point is the origin o. If N_0 is composed of two points, then these are separated by a unit distance and have midpoint o.

Formula (5.49) yields $D(r)$ by means of

$$\mathbf{E}\big(\nu_2(N_0 \oplus B(o,r))\big) = p_1\pi r^2 + p_2\big(2\pi r^2 - \gamma_r(1)\big),$$

where

$$\gamma_r(1) = \begin{cases} 2r^2 \cos^{-1}\dfrac{1}{2r} - \dfrac{\sqrt{4r^2-1}}{2} & \text{for } 2r \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

This quantity $\gamma_r(1)$ is the area of intersection of two circles of radius r and centres separated by unit distance. Then (3.30) gives $H_s(r)$. Moreover,

$$\overline{c} = p_1 + 2p_2$$

and

$$\boldsymbol{c}_o\big(\{\varphi \in \mathbb{N} : \varphi\big(B(o,r)\big) = 1\}\big) = \frac{1}{\overline{c}} \cdot \begin{cases} (p_1 + 2p_2) & \text{for } 0 \le r < 1, \\ p_1 & \text{for } r \ge 1. \end{cases}$$

Therefore

$$1 - D(r) = \frac{1}{\overline{c}} \exp\Big(-\lambda_{\mathrm{p}} \left(p_1 \pi r^2 + p_2\big(2\pi r^2 - \gamma_r(1)\big)\right)\Big) \cdot \begin{cases} p_1 + 2p_2 & \text{for } 0 \le r < 1, \\ p_1 & \text{for } r \ge 1. \end{cases} \tag{5.53}$$

Some processes are simultaneously Neyman–Scott *and* Cox processes. A Neyman–Scott process is Cox if the number of points per cluster has a Poisson distribution. This category includes the Matérn cluster process and the modified Thomas process. For example, the driving measure Ψ of the Matérn cluster process is

$$\Psi(B) = \mu \sum_{x \in \Phi_{\mathrm{p}}} \nu_d\big(B \cap B(x, R)\big) \qquad \text{for Borel } B,$$

where Φ_{p} denotes the Poisson process of parent points. Thall (1983) shows that for a Poisson cluster process that is also a Cox process, the number of points per cluster either takes on each positive integer value with positive probability or is identically equal to one. Thus, a Gauss-Poisson process cannot be Cox.

Van Lieshout (1995) and Baddeley *et al.* (1996) show that certain nonstationary Neyman–Scott processes can also be interpreted as Gibbs processes, at least if the Baddeley–Møller 'nearest neighbour' generalisation is used.

The Neyman–Scott process has found many applications as a model for spatial phenomena. Neyman and Scott (1958) introduce it as a model for the locations of galaxies in space; in this context the cluster centres are unobservable. (This model is now superseded; see Martínez and Saar, 2002.) The survey paper by Neyman and Scott (1972) contains further examples: the distribution of larvae in fields and the geometry of bombing. In the case of larvae in fields the parent points are positions of egg masses and the daughter points are the positions of larvae. In the case of the geometry of bombing, the parent points are the aiming points of bomb release and the daughter points are the points of impact of individual bombs. There are many applications of cluster processes in ecology, beginning with Warren (1971); see Stoyan and Penttinen (2000) and Illian *et al.* (2008).

The simulation of Neyman–Scott processes in a window W is easy: first the parent Poisson process is simulated in $W_{\oplus R}$, where R is a radius such that $\mathbf{P}\big(N_0 \subset B(o, R)\big)$ is very small or zero. Then around each parent point x a cluster distributed as $N_0 + x$ is generated. The sample is then the union of all daughter points in W. Brix and Kendall (2002) propose a procedure that simulates only those parents giving rise to daughter points in W. See Illian *et al.* (2008) for more details.

Statistics for Neyman–Scott processes is surprisingly complicated. Various methods have been tried, in particular the minimum contrast method as in Møller and Waagepetersen (2004) with $D(r)$, $H_s(r)$, $K(r)$ and $g(r)$. Experience gained by Stoyan and Stoyan (1996) and Brix (1999) indicates that it is preferable to work with $g(r)$. Tanaka *et al.* (2008) suggest an approximative maximum likelihood method.

5.4 Hard-core point processes

A hard-core point process is a point process in which the constituent points are forbidden to lie closer together than a certain positive minimum distance. In this section the minimum distance is denoted by D. These hard-core models describe patterns produced by the locations of centres of non-overlapping discs or balls of diameter D. There are many models of this type, which are often rather complicated. In this book some of them are presented in Section 6.5.3 the context of random sets. The reason for this is that these point processes belong to systems of hard balls, which are considered as random sets. The models considered in Section 6.5.3 are:

- the Stienen model,
- the lilypond model,
- packings of hard balls, and
- the random sequential adsorption (RSA) model.

Furthermore, Section 5.5 discusses the Gibbs hard-core process, which assigns an infinite *energy* to point pairs having a distance less than D so that the likelihood that a realisation containing such point pairs is zero.

This section considers the popular *(second) Matérn hard-core process*, for which exact mathematical calculations are possible and the corresponding dependent thinning procedure can be easily described. In his famous booklet, Matérn (1960, 1986) suggests two such models, but here only the model yielding a higher eventual intensity of points is considered. A similar hard-core process has also been proposed by Porod (1952). Its intensity is, however, not higher than that yielded by the model described here.

The model results from a dependent thinning applied to a homogeneous Poisson process Φ_b of intensity λ_b. The points of Φ_b are marked independently by random numbers uniformly distributed over (0, 1). The dependent thinning retains the point x of Φ_b with mark $m(x)$ if the ball $B(x, D)$ contains no points of Φ_b with marks smaller than $m(x)$. Formally the thinned process Φ is given by

$$\Phi = \big\{x \in \Phi_b : m(x) < m(y) \text{ for all } y \in \Phi_b \cap B(x, D) \setminus \{x\}\big\}. \tag{5.54}$$

It is stationary and isotropic.

The process Φ may be an appropriate model in certain ecological contexts. The points of Φ_b are for example the locations of seedlings while the marks are the instants at which the seedlings commence growth. So Φ consists of the positions x of those plants which are the eldest seedlings in their required growing area $B(x, D)$. A simulation is shown in Figure 5.7.

The intensity λ of Φ is given by

$$\lambda = p\lambda_b, \tag{5.55}$$

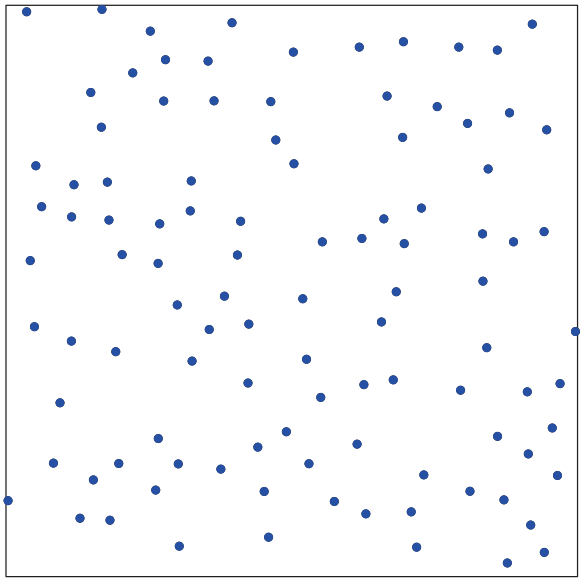

Figure 5.7 A simulated realisation of a planar second Matérn hard-core process in a unit square. The parameters are $\lambda_b = 200$ and $D = 0.05$.

where p is the *Palm retaining probability* of the typical point of Φ_b. It is given by

$$p = \int_0^1 r(t)\,\mathrm{d}t = \frac{1 - \exp(-\lambda_b c)}{\lambda_b c} \qquad \text{with } c = b_d D^d, \tag{5.56}$$

where $r(t) = \exp(-\lambda_b c t)$ is the retaining probability of a point of mark t. This formula follows from the observation that the point process

$$\{x \in \Phi_b : m(x) < t\}$$

is simply a t-thinning (the usual p is here t) of a Poisson process, hence itself a Poisson process of intensity $\lambda_b t$. So $r(t)$ is the probability that a ball of radius D contains no points of the t-thinned process. Consequently,

$$\lambda = \frac{1 - \exp(-\lambda_b c)}{c}. \tag{5.57}$$

If $D = 1$, then the maximum intensity λ is approximately 0.32 if $d = 2$, and 0.24 if $d = 3$. These maxima result from letting λ_b tend to infinity.

The second-order product density $\varrho^{(2)}(r)$ is given by

$$\varrho^{(2)}(r) = \lambda^2 \begin{cases} 0, & \text{if } r \le D, \\ \dfrac{2\Gamma_D(r)\big(1 - \exp(-\lambda_b c)\big) - 2c\big(1 - \exp(-\lambda_b \Gamma_D(r))\big)}{c\Gamma_D(r)\big(\Gamma_D(r) - c\big)}, & \text{if } D < r \le 2D, \end{cases} \tag{5.58}$$

with

$$\Gamma_D(r) = \nu_d\big(B(o, D) \cup B(\mathbf{r}, D)\big) = 2c - \gamma_D(r) \qquad \text{if } r > D, \tag{5.59}$$

where $r = \|\mathbf{r}\|$. In the practically important cases for $r \le 2D$

$$\gamma_D(r) = \begin{cases} 2D^2 \cos^{-1}\left(\dfrac{r}{2D}\right) - \dfrac{r}{2}\sqrt{4D^2 - r^2}, & \text{if } d = 2, \\ \dfrac{4\pi}{3} D^3 \left(1 - \dfrac{3r}{4D} + \dfrac{r^3}{16D^3}\right), & \text{if } d = 3. \end{cases} \tag{5.60}$$

Note that $\gamma_D(r)$ vanishes for $r \ge 2D$.

In the planar case the approximation $\gamma_D(r) = \pi D^2 - 2Dr$ may be useful. The pair correlation function of this process (with particular values chosen for the parameters) is graphed in Figure 4.7.

These formulae can be obtained by using $\varrho^{(2)}(r) = \lambda_b^2 k(r)$, where $k(r)$ is the two-point Palm probability that two points of Φ_b separated by distance r are both retained. Details of calculation of the above using the second-order factorial moment measure are to be found in Stoyan and Stoyan (1985). In this paper, as well as in Stoyan (1987, 1988), the Matérn model is generalised to have a variable hard-core, leading then to a soft-core model. So it is possible to obtain more realistic models for point patterns for which the original Matérn model is too simple. Statistically analysed examples include systems of locations of trees and villages.

Further generalisations that lead to germ–grain models with variable balls as grains are considered in Månsson and Rudemo (2002) and Andersson *et al.* (2006). Teichmann *et al.* (2013) study an elegant and general model where in a point pair of distance r in Φ_b the point with the larger mark is eliminated with probability $p(r)$. The classical second Matérn process is obtained with $p(r) = \mathbf{1}_{[0,D]}(r)$.

Closely related to this class of processes is Matheron's dead leaves model; see Stoyan and Schlather (2000). All these models, if considered as germ–grain models, have only a very small volume fraction $p = \lambda b_d D^d$, usually smaller than $1/2^d$.

The simulation of a Matérn hard-core process is not difficult. First a Poisson process of intensity λ_b is simulated in $W_{\oplus D}$. These points are marked independently with uniform marks in $(0, 1)$. Then the specific thinning rule explained on p. 176 is applied. Note that points deemed to be thinned out still have their job to do in the thinning rule and hence should be removed only after all points are classified as retained or thinned out; the retained points in W then form the required sample.

The thinning leading to the Matérn hard-core process can be refined such that higher intensities λ are possible, by the price of more complicated algorithms; see Møller *et al.* (2010) and Hörig and Redenbach (2012).

5.5 Gibbs point processes

5.5.1 Introduction

Gibbs point processes (*Gibbs processes* for short) first arose in the theories of statistical physics. They are related to the so-called Boltzmann–Gibbs distributions, which describe equilibrium states of subsystems of very large closed physical systems. In this book the

physical background will be largely ignored; the treatment will concentrate on explaining the mathematical reasoning and statistical aspects. Gibbs processes will be treated as a class of point processes and described in terms of the characteristics used to describe the classes of point processes discussed above. Detailed discussion of the mathematical–physical aspects may be found in Minlos (1968), Ruelle (1969, 1970) and Preston (1974, 1976). Historical aspects of the mathematical theory are covered briefly in Kallenberg (1983a, p. 177).

Note that Gibbs processes are *not* a universal class of models which can be used in all situations. They are good models for patterns with some degree of regularity (more regular than Matérn hard-core processes) or for moderate clustering, but can be deficient in cases of strong clustering (but see also the remark by Baddeley *et al.*, 1996, p. 347).

To commence the discussion in the spirit of this chapter, note that new point processes can be produced from old not only by thinning, clustering and superposition, but also by transforming their distributions by means of probability densities. The theory of Gibbs processes is a development of this idea.

Example 5.7. *Altering the distribution of a Poisson process of finite total intensity measure* Suppose that Ψ is a Poisson process on $\mathbb{R}^d$ with intensity measure Λ of total mass $\Lambda(\mathbb{R}^d) = 1$. Let the distribution of Ψ be Q. A new point process distribution P_λ is defined for positive λ by

$$P_\lambda(Y) = \int_Y f_\lambda(\varphi) Q(\mathrm{d}\varphi) \qquad \text{for } Y \in \mathcal{N}, \tag{5.61}$$

where

$$f_\lambda(\varphi) = \lambda^{\varphi(\mathbb{R}^d)} \exp(1-\lambda). \tag{5.62}$$

The point process under the new probability distribution P_λ is a Poisson process on $\mathbb{R}^d$ of intensity measure $\lambda\Lambda$.

Proof. Let Φ be a Poisson process of intensity measure $\lambda\Lambda$ and with distribution P_Φ. Then calculation shows that P_λ and P_Φ give the same measure to the set

$$Y_{K,n} = \{\varphi \in \mathbb{N} : \varphi(\mathbb{R}^d) = n, \varphi(K) = 0\}$$

for each nonnegative integer n and each compact set K. The void-probabilities of both processes agree since

$$P_\lambda(Y_{K,n}) = \lambda^n \exp(1-\lambda) \exp\bigl(-\Lambda(K)\bigr) \exp\bigl(-(1-\Lambda(K))\bigr) \frac{\bigl(1-\Lambda(K)\bigr)^n}{n!} = Q(Y_{K,n}),$$

and therefore their distributions are equal. □

Fundamental to the theory of Gibbs processes is the idea of a general *basic* or *weight process* Ψ of distribution Q. A new probability distribution P can be defined on $[\mathbb{N}, \mathcal{N}]$ by means of a probability density $f(\varphi)$ as

$$P(Y) = \int_Y f(\varphi) Q(\mathrm{d}\varphi) \qquad \text{for } Y \in \mathcal{N}. \tag{5.63}$$

If Φ is a point process of distribution P, then $f(\varphi)$ is the *likelihood* that Φ takes the realisation φ compared with Ψ taking on the same realisation. Obviously this construction is very general. In particular whole classes of realisations of Φ can be prohibited by setting f to be zero on the corresponding subset of $\mathbb{N}$.

The form of the density $f(\varphi)$ is often suggested by the field of application. It can be chosen conveniently to model interactions between the points of the process. The Gibbs process approach is straightforward if the process to be modelled contains only a finite number of points confined to a bounded region B. Problems arise when considering more general point processes. However, the construction via (5.63) may fail for point processes in the whole $\mathbb{R}^d$, as the following example shows.

Example 5.8. *Counter-example*

The homogeneous Poisson process $\Phi^{(1)}$ of intensity λ_1 cannot be obtained in the simple manner described above from another homogeneous Poisson process $\Phi^{(2)}$ of intensity λ_2 if $\lambda_1 \neq \lambda_2$.

Proof. Consider a family $\{W_n\}$ of compact sets with $W_n \uparrow \mathbb{R}^d$. Set, for $\lambda > 0$,

$$Y_\lambda = \{\varphi \in \mathbb{N} : \varphi(W_n)/v_d(W_n) \to \lambda\}.$$

Then clearly $Y_{\lambda_1} \cap Y_{\lambda_2}$ is empty but $\mathbf{P}(\Phi^{(i)} \in Y_{\lambda_i}) = 1$, $i = 1, 2$. So the distributions of the two processes are mutually singular, and thus a representation of the form (5.63) cannot be obtained. □

To obtain Gibbs processes in the situation of point processes on the whole $\mathbb{R}^d$, the density idea described above must be applied to conditional distributions confined to bounded regions. Section 5.5.3 describes the case of stationary point processes on $\mathbb{R}^d$. Before this, Section 5.5.2 deals with the simpler case of finite point processes.

5.5.2 Gibbs point processes in bounded regions

To facilitate understanding, the following considers a Gibbs point process Φ of distribution P and with *exactly n points in a specific bounded set B* and no points outside B ('absence of an outer field'). Physicists term this a *canonical ensemble*. This case of fixed n is of great practical importance for statistical purposes, where one often works under the condition that the number of points observed in the window of observation is fixed. It is assumed that the distribution of the point process is given by a probability density function $f : \mathbb{R}^{nd} \to [0, \infty)$ so that

$$\mathbf{P}(\Phi \in Y) = \int\limits_{\{x_1,\ldots,x_n\}\in Y} \ldots \int f(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n \qquad \text{for } Y \in \mathcal{N}_{B,n}, \tag{5.64}$$

where $\mathcal{N}_{B,n}$ denotes the trace of $\mathcal{N}$ on the set of all point sequences with n points in B. Because point processes are unordered sets of points, $f(x_1, \ldots, x_n)$ is taken not to depend on the order of the arguments. The form used to describe f is

$$f(x_1, \ldots, x_n) = \exp\bigl(-E(x_1, \ldots, x_n)\bigr)/Z. \tag{5.65}$$

Here Z is a normalising constant called the (*configurational*) *partition function*, and the function $E : \mathbb{R}^{nd} \to \mathbb{R} \cup \{\infty\}$ is called the *energy function* or *multiparticle functional*; also it does not depend on the order of the arguments. These terms derive from statistical mechanics. The convention is adopted that

$$\exp(-\infty) = 0.$$

Frequently E is chosen to be of a specialised form, as a sum of *pair potentials*

$$E(x_1, \ldots, x_n) = \sum_{1 \le i < j \le n} \sum \theta(\|x_i - x_j\|). \tag{5.66}$$

The name 'pair potential' for the function $\theta : [0, \infty) \to (-\infty, \infty]$ also comes from statistical physics.

This leads to the formula

$$f(x_1, \ldots, x_n) = \exp\left(- \sum_{1 \le i < j \le n} \sum \theta(\|x_i - x_j\|) \right) \Big/ Z. \tag{5.67}$$

Often the function θ is rewritten so that (5.67) takes the form

$$f(x_1, \ldots, x_n) = \exp\left(-\beta \sum_{1 \le i < j \le n} \sum \theta(\|x_i - x_j\|) \right) \Big/ Z, \tag{5.68}$$

where β is called the *inverse temperature*.

A typical example of a pair potential is graphed in Figure 5.8. Its features correspond to features of the point process of density f constructed by using Formulae (5.63), (5.65) and (5.66):

(a) Because $\theta(r)$ is infinite for $r < D$ the inter-point distance can never be less than D. So the point process is in fact a hard-core model.

(b) Because $\theta(r)$ is large when r is only a little larger than D such distances r are possible but are unlikely to occur as inter-point distances.

(c) In contrast, inter-point distances close to r_1, where θ takes its minimum, should occur relatively frequently.

From the theory of statistical mechanics, distributions of the form given in (5.65) arise when the point process Φ is constrained to have fixed mean energy

$$\int_{B^n} E(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

and maximum entropy

$$H = -\int_{B^n} f(x_1, \ldots, x_n) \ln\big(f(x_1, \ldots, x_n)\big) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

among such fixed-energy systems.

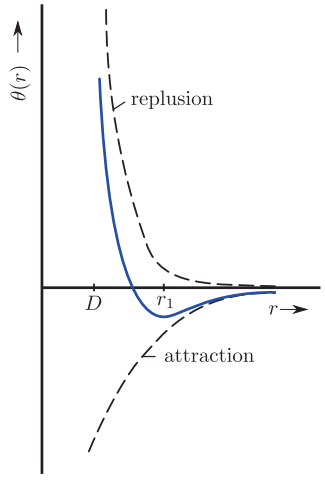

Figure 5.8 A typical pair potential, the result of superposition of attractive and repulsive forces.

If the distribution of Φ is given by Equation (5.64), as an integral of a probability density, then the underlying weight process Ψ has a density proportional to the uniform density on B^n. Clearly the form of B is crucial to the distribution of the resulting Gibbs process Φ.

Often, Formula (5.65) with E as in (5.66) is re-cast in terms of an *interaction function* $e(t) = \exp\bigl(-\theta(t)\bigr)$. This motivates the frequent description of such point processes Φ as *pairwise interaction processes* or (special cases of) *Markov point processes*; see van Lieshout (2000) and Møller and Waagepetersen (2004). Then the exponential of the sum (5.66) is replaced by a product of individual exponential terms.

The normalising constant Z in Formula (5.65) is given by

$$Z = \int_{B^n} \exp\bigl(-E(x_1, \ldots, x_n)\bigr)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n. \tag{5.69}$$

Computation of the partition function Z is notoriously difficult. Ogata and Tanemura (1981) propose an approximation method for the case of pair potentials, the second cluster approximation:

$$Z \approx \nu_d(B)^n \left(1 - \frac{a}{\nu_d(B)}\right)^{n(n-1)/2}, \tag{5.70}$$

where

$$a = \left(\int_0^\infty \bigl(1 - \exp(-\theta(r))\bigr) r^{d-1}\, \mathrm{d}r\right) db_d.$$

(Recall that db_d is the surface area of the unit ball in $\mathbb{R}^d$.) Ogata and Tanemura (1984) and Gates and Westcott (1986) discuss the quality of this approximation. It is poor if the pair potential $\theta(r)$ is negative for small r or has large deviations from zero. Ripley (1988) and Diggle *et al.* (1994) describe other more precise approximations. The partition function is

often determined by MCMC simulation; see Ogata and Tanemura (1984), Geyer (1992), Huang and Ogata (2001) and Møller and Waagepetersen (2004).

Example 5.9. *The square-well pair potential*
For $0 < D < \rho$ and $-\infty < b < \infty$ let

$$\theta_{D,\,b,\,\rho}(r) = \begin{cases} \infty, & \text{if } r \le D, \\ -b, & \text{if } D < r \le \rho, \\ 0, & \text{if } r > \rho. \end{cases}$$

Then the density $f(\varphi)$ has the form

$$f(\varphi) = \begin{cases} \exp\bigl(bs_\varphi(\rho)\bigr) \big/ Z, & \text{if } \varphi \subset B \text{ and } s_\varphi(D) = 0, \\ 0, & \text{otherwise,} \end{cases}$$

where for $\varphi = \{x_1, \ldots, x_n\}$,

$$s_\varphi(\rho) = \sum_{1 \le i < j \le n} \alpha_\rho(x_i, x_j)$$

with

$$\alpha_\rho(x_i, x_j) = \begin{cases} 1, & \text{if } \|x_i - x_j\| \le \rho, \\ 0, & \text{otherwise.} \end{cases}$$

So $s_\varphi(\rho)$ is simply the number of inter-point distances in φ that are less than or equal to ρ.

In Example 5.9 the parameter D is a hard-core distance while ρ is a radius of interaction. The cases $b > 0$ and $b < 0$ correspond respectively to clustering and regularity. Particular cases are:

- the uniform distribution of n points in B (when $b = 0$, $D = 0$);
- the *Strauss model* with fixed number of points (when $D = 0$).

The Strauss model is discussed in many papers on Gibbs processes; see the books by van Lieshout (2000), Møller and Waagepetersen (2004), Daley and Vere-Jones (2008) and Illian *et al.* (2008), often for random n. The model need not then be well-defined for positive b; see Møller (1999). Even for fixed n, problems may appear in simulations for large positive b: the algorithm may get stuck in the same point configuration for many iterations.

Ogata and Tanemura (1981) study other parametric families of pair potentials θ, namely

$$\begin{aligned} \theta_1(r) &= -\ln\bigl(1 + (\alpha r - 1)\mathrm{e}^{-\beta \mathrm{r}^2}\bigr), \\ \theta_2(r) &= -\ln\bigl(1 + (\alpha - 1)\mathrm{e}^{-\beta \mathrm{r}^2}\bigr), \\ \theta_3(r) &= \beta \left(\frac{\sigma}{r}\right)^n - \alpha \left(\frac{\sigma}{r}\right)^m, \end{aligned}$$

for $\alpha \ge 0$, $\beta, \sigma > 0$ and $n > m$.

The last potential $\theta_3(r)$ is a common model in statistical mechanics, called the *Lennard-Jones potential*. The case $n = 12$, $m = 6$ is of particular interest. It provides both types of interaction, attraction and repulsion, at different scales.

Ogata and Tanemura also describe *maximum likelihood estimation* of parameters for these models in case of sparseness, using approximations to the likelihood via (5.65) and (5.70). A general approximation method by using MCMC is proposed in Gu and Zhu (2001). Asymptotic optimality properties of maximum likelihood estimators are studied by Mase (1992).

Penttinen (1984) proposed an alternative approach. In the planar case, his method gives for the parameter b of $\theta_{D,b,\rho}$ (if D and ρ are known) the estimator

$$\hat{b} = -\ln \frac{2\nu_2(B)s_\Phi(\rho)}{\pi n^2(\rho^2 - D^2)}.$$

For the Strauss model this simplifies to

$$\hat{b} = -\ln \frac{2\nu_2(B)s_\Phi(\rho)}{\pi n^2 \rho^2}.$$

Here and below B and W, the window of observation, coincide.

In contrast, the maximum likelihood method leads for the Strauss process to the estimator

$$\hat{b} = -\ln \frac{2\nu_2(B)s_\Phi(\rho)}{\pi n(n-1)\rho^2}.$$

More general statistical methods for Gibbs processes with finite point number are described in the books by Møller and Waagepetersen (2004), Illian *et al.* (2008) and Gelfand *et al.* (2010). These books also describe a number of varied examples.

Gibbs processes with a *random* total number of points are also studied; these go in physics under the name *grand canonical ensemble*. The reader is referred to the literature, for example van Lieshout (2000), Møller and Waagepetersen (2004), Illian *et al.* (2008) and Gelfand *et al.* (2010).

Simulation of Gibbs processes with fixed point number

For the simulation of Gibbs processes usually an iterative procedure called *Markov chain Monte Carlo* (MCMC) method is applied. The idea is to construct a Markov chain, the states of which are point configurations in B and the stationary distribution of which is that of the Gibbs process. The algorithm simulates the Markov chain for a long time. Once the chain is in its stationary regime, every state of it is a sample from the stationary distribution of the chain; of course, states close together in time are spatially correlated. The following sketches two particular algorithms. Van Lieshout (2000), Møller and Waagepetersen (2004), Gelfand *et al.* (2010) and Huber (2011) provide a detailed overview of MCMC algorithms for point pattern simulation and give the theoretical justifications.

The first method is the classical Metropolis algorithm as in Metropolis *et al.* (1953); see also Torquato (2002) and Billera and Diaconis (2001), who explain the 'magic trick' of this famous algorithm, why it works.

The steps of the *Metropolis algorithm* in the Gibbs process case are as follows.

1. Generate an initial configuration φ of n points in B and compute its energy $E(\varphi)$.

2. Move a point by displacing it along each coordinate axis by amounts randomly and uniformly in the interval $[-\delta, \delta]$, where δ is a positive parameter. This leads to a new configuration φ'. Compute the new energy $E(\varphi')$ and the energy difference $\Delta E = E(\varphi') - E(\varphi)$.

3. Accept the new configuration φ' with probability $\min\{1, \exp(-\Delta E)\}$. This means: if the energy of the new configuration is smaller than that of the old, the move $\varphi \to \varphi'$ is accepted. Otherwise it is accepted with probability $\exp(-\Delta E)$; non-acceptance of φ' means that φ is retained.

4. Continue by going to step 2 or, after sufficiently many moves, goto step 5.

5. Accept the current configuration as a sample of the Gibbs process.

The order in which the points are moved is fixed. Moves that lead to positions outside of B are corrected in suitable form. Torquato (2002) recommended choosing the parameter δ so that about 50% of the proposed moves would be accepted.

Metropolis *et al.* (1953) considered the planar case with B a square under periodic boundary conditions, which means that their intention is to simulate a stationary Gibbs process. As an illustration, they simulated the hard disc case with $n = 224$.

Instead of the random shifts, other 'moves' may also be used, for example jumps; see the literature on MCMC, for example Frenkel and Smit (2002) and Brooks *et al.* (2011).

The second method uses a *birth-and-death algorithm*. As in Illian *et al.* (2008), in the following a simple but nontrivial example is discussed followed by a description of the general algorithm.

Example 5.10. *Simulation of the Gibbs hard-core process with n points*

The starting configuration can be any n-point pattern in B in which the smallest inter-point distance is not less than D, for example a lattice of n points.

Assume that the current state of the Markov chain after l steps is $\{x_1, \ldots, x_n\}$. The $(l+1)^{\text{st}}$ step is:

1. A point in the set $\{x_1, \ldots, x_n\}$ is deleted at random, where each point has the same probability $1/n$ to be deleted. The deleted point is, say, x_k.

2. A new point x is drawn in B with uniform distribution. If

$$\min\{\|x - x_i\| : i = 1, 2, \ldots, n,\ i \neq k\} \geq D,$$

then x replaces x_k to form a new set of n points $\{x_1, \ldots, x_{k-1}, x, x_{k+1}, \ldots, x_n\}$. Otherwise, if

$$\min\{\|x - x_i\| : i = 1, 2, \ldots, n,\ i \neq k\} < D,$$

the point x is rejected and a new proposal point is drawn until the proposed x can be accepted.

When n and D are large, it often takes a long time to have a new point accepted.

Inspection of patterns simulated with this algorithm shows that there is some boundary effect called 'drift towards the boundary': the point density along the edge of B is slightly higher than in its interior. This is a correct result, which is in full agreement with the Gibbs distribution in a bounded set B with no influence from outside. If a pattern looking like a 'stationary sample' is required, periodic boundary conditions can be used.

The structure of the general birth-and-death algorithm closely resembles that of the algorithm for the Gibbs hard-core process above. A randomly chosen point in the pattern is deleted and a new point is generated following the conditional probability density function as above. The simulation starts with an initial point configuration of positive density, which does not 'contradict' the pair potential.

Suppose that after the first l steps the point pattern is $\{x_1, \ldots, x_n\}$. The $(l+1)^{\text{st}}$ step is:

1. A point in the set $\{x_1, \ldots, x_n\}$ is deleted at random, where each point in $\{x_1, \ldots, x_n\}$ has the same probability $1/n$ to be deleted. The deleted point is, say, x_k.

2. A new point is simulated based on the conditional probability density function

$$f(x|x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n) = \frac{f(x_1, \ldots, x_{k-1}, x, x_{k+1}, \ldots, x_n)}{\displaystyle\int_B f(x_1, \ldots, x_{k-1}, y, x_{k+1}, \ldots, x_n)\,\mathrm{d}y}. \quad (5.71)$$

The new configuration $\{x_1, \ldots, x_{k-1}, x, x_{k+1}, \ldots, x_n\}$ is the result after the $(l+1)^{\text{st}}$ step. As above for the Gibbs hard-core process, the x in substep 2 may be simulated by the rejection method: let M be an upper bound of the nonnormalised conditional density function

$$h(x) = \exp\left(-\sum_{j=1, j\neq k}^{n} \theta(\|x - x_j\|)\right).$$

A random point x is generated uniformly in B along with an independent uniform random number u on $[0, 1]$. The point x is accepted if $h(x) \geq Mu$. Otherwise a new point x is proposed.

The algorithm for the Gibbs hard-core process follows this approach with $h(x) = 1$ if $\|x - x_j\| > D$ for all $j = 1, \ldots, n$, $j \neq k$, and zero otherwise. Illian *et al.* (2008, Section 3.6.2) presented details for the application of the birth-and-death algorithm.

5.5.3 Stationary Gibbs point processes

The methods of Section 5.5.2 cannot be used directly in the case when B is replaced by $\mathbb{R}^d$ and the number of points is infinite. A more sophisticated approach is required, which considers the restrictions of point processes to bounded sets B, and these must be conditioned on the behaviour of the point process outside the bounded region B. Formally one considers probabilities of the form $\pi_B(Y \mid \psi)$. This is the probability that the pattern in B belongs to the configuration set Y, given that outside B the pattern takes on the configuration ψ. So

$$\pi_B(Y \mid \psi) = \mathbf{P}(\Phi \cap B \in Y \mid \Phi \cap B^{\mathrm{c}} = \psi), \quad (5.72)$$

where B is a bounded Borel set, Y is in $\mathcal{N}$, and ψ belongs to the set $\mathbb{N}_{B^{\mathrm{c}}}$, where

$$\mathbb{N}_{B^{\mathrm{c}}} = \{\varphi \in \mathbb{N} : \varphi(B) = 0\}.$$

These probabilities must satisfy certain measurability and compatibility conditions, which are omitted here. The objective of this section is to show how in many cases the probabilities π_B can be expressed elegantly using a weight process and a pair potential.

The mathematical problems of the above formulation were intensively studied in the 1970s by, for example, Georgii (1976), Preston (1976), Matthes *et al.* (1978, 1979), Nguyen and Zessin (1979b), Glötzl (1980) and Kallenberg (1983a).

Rather than deriving the π_B from the point process Φ, one can attempt to move in the reverse direction. Indeed, this is the more useful procedure for model specification.

Suppose the *local specification* $\{\pi_B\}$ is given for all bounded Borel B (using the terminology of Preston, 1976). A point process Φ with distribution P is said to be a *Gibbs point process with respect to the local specification* $\{\pi_B\}$, if for all bounded Borel B

$$P(Y) = \int \pi_B(Y \mid \varphi \cap B^{\mathrm{c}}) P(\mathrm{d}\varphi). \tag{5.73}$$

The equation (5.73) is called the *DLR-equation* after Dobrushin, Lanford and Ruelle. Note that its solution P may not be unique.

The concept of a Gibbs point process is closely connected to the older notion of a *Boltzmann–Gibbs distribution* in statistical physics. The corresponding physical question concerns the existence and uniqueness of equilibrium states for dynamic particle systems given local dynamic specifications analogous to (5.73). See Ruelle (1969) for a discussion of the general physical theory and Georgii (1988) for a general study of uniqueness and phase transition in Gibbs processes.

Formula (5.65) gives an exponential form for the density of P in the bounded case. By analogy the probabilities π_B are considered in exponential form

$$\pi_B(Y \mid \psi) = \frac{1}{Z(B, \psi)} \int \mathbf{1}_Y\big((\varphi \cap B) \cup \psi\big) \exp\big(-\overline{E}(\varphi \cap B, \psi)\big)\, Q(\mathrm{d}\varphi), \tag{5.74}$$

where Q is the distribution of a basic or weight process and $Z(B, \psi)$ is a normalising constant. The function $\overline{E} : \mathbb{N}_{\mathrm{finite}} \times \mathbb{N} \to (-\infty, \infty]$ is the *conditional energy* (here $\mathbb{N}_{\mathrm{finite}} = \{\varphi \in \mathbb{N} : \varphi(\mathbb{R}^d) < \infty\}$). The conditional energy $\overline{E}(\chi, \psi)$ can be interpreted as the energy required to add the point set χ to the configuration ψ.

To complete the analogy with (5.65), Formula (5.74) can be derived as the solution to a maximum entropy problem when the energy is held constant.

Two simplifying conditions are imposed in the subsequent discussion.

First, $\overline{E}$ is assumed to have an additive form expressed in terms of the *local energy* E

$$\overline{E}(\chi, \psi) = E(x_1, \psi) + E(x_2, \psi \cup \{x_1\}) + \cdots + E(x_n, \psi \cup \{x_1, \ldots, x_{n-1}\}), \tag{5.75}$$

where $\chi = \{x_1, \ldots, x_n\}$. Naturally a consistency condition must be imposed on E to ensure that the sum does not depend on the order of the x_i. The term $E(x, \varphi)$ can be interpreted as the energy required to add the point x to the configuration φ and is called *local energy*.

Second, the weight process is supposed to be Poisson with intensity measure Λ_Q. This is the natural generalisation of the 'uniform' weight process used to treat Gibbs processes with nonrandom finite total number of points as in Section 5.5.2.

Clearly it is important to decide when a distribution P leads to a local specification $\{\pi_B\}$ in the form (5.74) with a local energy. Some condition resembling absolute continuity is

required. The answer has been given by Georgii (1976), Nguyen and Zessin (1979b) and Glötzl (1980).

Theorem 5.1. *A necessary and sufficient condition for the point process distribution P to have local specification $\{\pi_B\}$ in the form of (5.74) with $\overline{E}$ given by (5.75) is as follows: the reduced Campbell measure $\mathscr{C}^!$ corresponding to P satisfies the absolute continuity condition*

$$\mathscr{C}^! \ll \Lambda_Q \times P, \tag{5.76}$$

where Λ_Q is the intensity measure of the Poisson weight process. Furthermore, if (5.76) holds, then for almost all x and φ (with respect to P)

$$\frac{\mathrm{d}\mathscr{C}^!}{\mathrm{d}(\Lambda_Q \times P)}(x, \varphi) = \exp\bigl(-E(x, \varphi)\bigr). \tag{5.77}$$

Equation (5.77) suggests a plausible argument for the interpretation of $E(x, \varphi)$ as the energy required to add x to φ. For simplicity consider the case when P and Q are both stationary with intensities λ and λ_Q. Then the Radon–Nikodym derivative on the left-hand side of (5.77) can be written informally as

$$\frac{\mathrm{d}\mathscr{C}^!}{\mathrm{d}(\Lambda_Q \times P)}(x, \varphi) = \frac{\mathscr{C}^!(\mathrm{d}x \times \mathrm{d}\varphi)}{\lambda_Q \nu_d(\mathrm{d}x) P(\mathrm{d}\varphi)}.$$

By Formula (4.65), which expresses the reduced Palm measure in terms of the reduced Campbell measure $\mathscr{C}^!$, the infinitesimal quotient can be written as

$$\frac{\mathscr{C}^!(\mathrm{d}x \times \mathrm{d}\varphi)}{\lambda_Q \nu_d(\mathrm{d}x) P(\mathrm{d}\varphi)} = \frac{P_o^!(\mathrm{d}\varphi)\lambda \nu_d(\mathrm{d}x)}{\lambda_Q \nu_d(\mathrm{d}x) P(\mathrm{d}\varphi)},$$

which is proportional to

$$\frac{P_o^!(\mathrm{d}\varphi)}{P(\mathrm{d}\varphi)}.$$

This last quotient is the ratio of the infinitesimal probabilities for Φ under P (in the denominator) and P conditioned to have a point at o (in the numerator). The standard physical interpretation regards the negative logarithm of the quotient as the energy required to make the transition from φ to $\varphi \cup \{o\}$.

The question of uniqueness of the process given by the local specification is of great interest and importance in probability theory, but appears not to be of great relevance in spatial statistics. This is because all practical applications involve limited quantities of data observed through bounded windows. The non-uniqueness phenomenon may manifest itself in increased sensitivity to boundary conditions. However, it seems that non-uniqueness phenomena usually begin to play an important part when the numbers involved approach the thermodynamic ranges of the order of 10^{23} points. On the other hand, when the relatively low numbers of points in statistical problems are being considered, the problem of edge-effects is likely to be of far greater significance (B. D. Ripley, 1984, personal communication).

Together with Formula (4.65), Equation (5.77) implies that

$$\lambda P_o^!(Y) = \lambda_Q \int_Y \exp\bigl(-E(o,\varphi)\bigr)\, P(\mathrm{d}\varphi), \tag{5.78}$$

which in turn implies that the reduced Palm distribution $P_o^!$ is absolutely continuous with respect to P. This enables a quick determination that a point process is *not* a Gibbs process (with respect to a Poisson process, using local energy). For example, the lattice of points of integral coordinates, translated in both coordinates by random shifts uniformly over $(0, 1)$, is a stationary point process which cannot be so expressed. Indeed its reduced Palm probability measure yields simply the lattice with the origin deleted and this is clearly not absolutely continuous with respect to the original law.

In the following it is assumed that $\lambda_Q = 1$, which can be compensated by the freedom to choose the local energy.

An equivalent formulation of Equation (5.78) is

$$\lambda \int g(\varphi) P_o^!(\mathrm{d}\varphi) = \int g(\varphi) \exp\bigl(-E(o,\varphi)\bigr)\, P(\mathrm{d}\varphi) \tag{5.79}$$

for every nonnegative measurable $g : \mathbb{N} \to \mathbb{R}$, which has several applications, for example in the Takacs–Fiksel method.

A more general equation is the *GNZ (Georgii–Nguyen–Zessin) equation*

$$\int \int h(x, \varphi \setminus \{x\})\varphi(\mathrm{d}x) P(\mathrm{d}\varphi) = \int \int h(x,\varphi) \exp\bigl(-E(x,\varphi)\bigr)\, P(\mathrm{d}\varphi)\, \mathrm{d}x \tag{5.80}$$

or

$$\mathbf{E}\left(\sum_{x\in\Phi} h(x, \Phi \setminus \{x\})\right) = \int \mathbf{E}\bigl(h(x,\Phi)\exp(-E(x,\Phi))\bigr)\mathrm{d}x,$$

where h is a function as in (4.64). The Campbell–Mecke theorem yields (5.79).

An important class of general Gibbs processes corresponding to the class of models specified by the Formulae (5.74) and (5.76) is obtained by requiring that the local energy $E(x, \varphi)$ can be expressed in terms of a *pair potential* θ as in Section 5.5.2:

$$E(x,\varphi) = \alpha + \sum_{y\in\varphi} \theta(\|x - y\|). \tag{5.81}$$

The constant α is called the *chemical activity*. If the pair potential $\theta(r)$ is fixed, then there is the tendency that the Gibbs process intensity λ increases with decreasing α.

Note that the assumption (5.75) implies that the Gibbs process is *hereditary*, that is, for every 'forbidden' point configuration φ and every $x \in \mathbb{R}^d$, the configuration $\varphi \cup \{x\}$ is also forbidden. Dereudre and Lavancier (2009) and Dereudre *et al.* (2012) consider Gibbs processes that are nonhereditary and give a generalised form of Equation (5.80). Such processes appear, for example, in the context of Delaunay and Voronoi tessellations with cells of limited variability (neither too large, too small, nor too flat); see Dereudre and Lavancier (2011).

It is difficult to obtain formulae for the point process characteristics introduced in Chapter 4. However, there are various approximations in the pair potential case. For the intensity λ,

Baddeley and Nair (2012) suggest the following approach: Substituting $g(\varphi) \equiv 1$ in (5.79) gives

$$\lambda = \mathrm{e}^{-\alpha} \int \exp\left(- \sum_{y \in \varphi} \theta(y) \right) P(\mathrm{d}\varphi).$$

The integral on the right-hand side is a mean value with respect to the Gibbs process. The idea is to approximate it by an analogous mean with respect to the Poisson process of the same intensity. The latter mean is equal to $\exp(-\lambda G)$ with

$$G = \int_{\mathbb{R}^d} \big(1 - \exp(-\theta(x)\big) \mathrm{d}x. \tag{5.82}$$

Thus the intensity approximation is the solution of

$$\lambda = \mathrm{e}^{-\alpha} \exp(-\lambda G). \tag{5.83}$$

In the particular case of a *Gibbs hard-core process*, which is the Gibbs process with pair potential given by

$$\theta(r) = \begin{cases} \infty, & \text{if } r \le D, \\ 0, & \text{otherwise,} \end{cases} \tag{5.84}$$

the constant G in (5.82) is given by

$$G = b_d D^d \tag{5.85}$$

and the approximation of λ is then the solution of

$$\lambda = \mathrm{e}^{-\alpha} \exp(-\lambda b_d D^d). \tag{5.86}$$

Clearly, for high intensities this approximation may fail; then in the approximation a lattice process may be used instead of a Poisson process.

Mase (1990) suggests an approximation method for calculating means of the form

$$\mathbf{E}\left(\prod_{x \in \Phi} c(x) \right)$$

for nonnegative measurable functions c on $\mathbb{R}^d$. Quantities such as the spherical contact distribution function $H_s(r)$ and the nearest-neighbour distance distribution function $D(r)$ belong to this class. For example, consider

$$c(x) = 1 - \mathbf{1}_B(x)$$

for a Borel set B. Then

$$\mathbf{E}\left(\prod_{x \in \Phi} c(x) \right) = \mathbf{E}\mathbf{1}_{\{\Phi(B)=0\}} = \mathbf{P}\big(\Phi(B) = 0\big).$$

If $\theta(r)$ has a finite *interaction radius* ρ, that is to say, if $\theta(r)$ is zero for $r \geq \rho$, then the following relationship holds:

$$\lambda P_o^!(Y_r) = P(Y_r)\,\mathrm{e}^{-\alpha} \qquad \text{for } r \geq \rho, \tag{5.87}$$

where $Y_r \in \mathcal{N}_{\mathbb{N}_{B(o,r)^c}}$, the trace of $\mathcal{N}$ on the set $\mathbb{N}_{B(o,r)^c}$ of point sequences φ that $\varphi \cap B(o, r) = \emptyset$. In terms of the nearest-neighbour distance distribution function $D(r)$ and the spherical contact distribution function $H_s(r)$, the relationship given in (5.87) becomes

$$\lambda\big(1 - D(r)\big) = \big(1 - H_s(r)\big)\mathrm{e}^{-\alpha} \qquad \text{for } r \geq \rho. \tag{5.88}$$

For large windows, when $D(r)$ and $H_s(r)$ can be estimated quite accurately, Formula (5.88) suggests a way to estimate α.

In the case of a Gibbs hard-core process, substituting $r = D$ in Formula (5.88) yields a close relationship between λ, D and α:

$$\lambda = \big(1 - H_s(D)\big)\mathrm{e}^{-\alpha}. \tag{5.89}$$

Formula (6.103) in Section 6.5.3 for the spherical contact distribution function of a random dense hard ball packing can be used to calculate $H_s(r)$ for the Gibbs hard-core process for $d = 3$ and $r > \frac{D}{2}$.

In the physical literature there are ways to derive approximations for the pair correlation function $g(r)$ of stationary Gibbs processes, given λ and $\theta(r)$; see the classical book by Hansen and McDonald (2006). These are acceptable in the case of small intensity λ, as 'low-density' approximations. The simplest approximation is

$$g(r) \simeq \exp\big(-\theta(r)\big) \qquad \text{for } r \geq 0, \tag{5.90}$$

(see Hansen and McDonald, 2006, p. 35). Better results are obtained by means of the so-called Percus–Yevick approximation, which yields approximations for $g(r)$ in terms of λ and $\theta(r)$:

$$g(r) \approx \frac{c(r)}{1 - \exp\big(\theta(r)\big)} \qquad \text{for } r \geq 0, \tag{5.91}$$

(see Diggle *et al.*, 1987, and Hansen and McDonald, 2006, p. 91), where $c(r)$ is a further function called *direct correlation function*, which is defined implicitly by the solution of the so-called Ornstein–Zernicke equation

$$c(r) = h(r) - \lambda(c * h)(r) \qquad \text{for } r \geq 0, \tag{5.92}$$

with $h(r) = g(r) - 1$ and

$$(c * h)(r) = \begin{cases} \displaystyle\int_0^\infty \int_0^{2\pi} p(s)q\big(\sqrt{r^2 + s^2 - 2rs\cos\alpha}\big)s\,\mathrm{d}\alpha\,\mathrm{d}s, & \text{for } d = 2, \\ \displaystyle\int_0^\infty \int_0^{2\pi} \int_0^{\pi} p(s)q\big(\sqrt{r^2 + s^2 - 2rs\cos\alpha\sin\beta}\big)s^2\sin\beta\,\mathrm{d}\beta\,\mathrm{d}\alpha\,\mathrm{d}s & \text{for } d = 3, \end{cases} \tag{5.93}$$

following Ornstein and Zernike (1914). Combining (5.91) and (5.92) yields an integral equation for $g(r)$. Torquato (2002, p. 72ff) and Buryachenko (2007, p. 156) discuss its solution by

means of Fourier methods. In the case of hard balls for volume fractions up to 0.5 excellent approximations of $g(r)$ are obtained; see for example Hansen and McDonald (2006, p. 94).

The theory of stationary Gibbs processes was extended to the *marked case*; see Coeurjolly *et al.* (2012). This opens the way to important random-set models, for example the quermass models. These form a class of germ–grain models with morphological interaction between the grains, which leads to structures that cannot come from Boolean models.

An approximation for $g(r)$ for three-dimensional Gibbs hard-core processes with low intensity is given in Kanaun and Levin (1994), based on Willis (1978):

$$g(r) = \big(1 - \mathbf{1}_{[0,D)}(r)\big)\left(1 + \Big(\frac{2+p}{2(1-p)^2} - 1\Big)\cos\Big(\frac{2\pi r}{D}\Big)\exp\Big(4(1-\frac{r}{D})\Big)\right) \qquad \text{for } r \geq 0, \tag{5.94}$$

where p denotes the volume fraction V_V of the system of hard balls of diameter D centred at the points of the Gibbs hard-core process. Comparison of the graph corresponding to the $g(r)$ in this formula with Figure 4.7 shows that many details of $g(r)$ appearing in the case of high volume fraction p are not reflected by the low-density approximation (5.94).

The planar case is also discussed in Buryachenko (2007).

Methods of statistics for stationary Gibbs processes are described in Illian *et al.* (2008) and Baddeley and Dereudre (2013). A particular method is the Takacs–Fiksel method, which is based on the GNZ equation (5.80); see Coeurjolly *et al.* (2012), where the statistical properties of this method are thoroughly investigated. An alternative may be the approach in Diggle *et al.* (1987), which is based on (5.91) and uses estimates of the pair correlation function. Another popular approach is the 'method' to ignore that the given sample belongs to a stationary process and to apply methods for bounded regions B, where B is identified with the window of observation and it is believed that the estimated pair potential is also valid for the stationary process. This approach opens the way to application of the powerful methods in van Lieshout (2000) and Møller and Waagepetersen (2004).

5.5.4 Spatial birth-and-death processes

A spatial birth-and-death process is a continuous-time Markov process with state space $\mathbb{N}$ the family of point patterns; see Preston (1977), Glötzl (1981), van Lieshout (2000) and Daley and Vere-Jones (2008). These processes provide models of spatio–temporal (or space–time) point processes, which are studied in the books by Daley and Vere-Jones (2008), Illian *et al.* (2008) and Cressie and Wikle (2011). There the term 'birth-and-death' process is sometimes used in a more general sense than in this section, for systems where points randomly appear, exist for a random time and then disappear. The relevance of the birth-and-death processes discussed here lies in the close relationship to Gibbs processes and especially in the way in which they provide a means to simulate Gibbs processes as suggested by Ripley (1977).

A brief description of a spatial birth-and-death process is that it is a point process in space which changes at isolated instants of time either by the *birth* of a new point, which is added to the pattern, or by the *death* of an old point, which is deleted. Moreover, it has a *Markov* property in time: the probability of a change depends only on the current configuration of the process. Births are controlled by a *birth-rate* $\boldsymbol{b}$, a positive measurable function

$$\boldsymbol{b} : \mathbb{R}^d \times \mathbb{N} \to \mathbb{R}$$

with

$$\int_B \boldsymbol{b}(x, \varphi)\,\mathrm{d}x < \infty$$

for all bounded Borel sets B and all φ in $\mathbb{N}$. The probability that a birth occurs within the set B in the short time interval $[t, t+s)$, given that the process has configuration φ at time t, is

$$s\int_B \boldsymbol{b}(x, \varphi)\,\mathrm{d}x + o(s).$$

The occurrence of a death is controlled by a *death-rate* $\boldsymbol{d}$, another positive measurable function

$$\boldsymbol{d} : \mathbb{R}^d \times \mathbb{N} \to \mathbb{R}.$$

The probability that the point x is deleted from the pattern $\varphi \cup \{x\}$ in the time interval $[t, t+s)$, given that $\varphi \cup \{x\}$ is the state of the birth-and-death process at time t, is

$$s\boldsymbol{d}(x, \varphi) + o(s).$$

The two functions $\boldsymbol{b}$ and $\boldsymbol{d}$ characterise the birth-and-death process. It is of course necessary to impose conditions on $\boldsymbol{b}$ and $\boldsymbol{d}$ to eliminate various pathological properties. Suitable conditions as well as a formal construction can be found in Preston (1976).

The *equilibrium* or time-stationary distributions of the process are of particular interest. A point process distribution P is an *equilibrium distribution* of the birth-and-death process if when the birth-and-death process is started with distribution P at time 0 its distribution continues as P for all time. The actual configuration may change but the statistics of the configuration have a time-invariant distribution. An equilibrium state is a *time-reversible* state of the birth-and-death process if in addition a realisation of the ensuing evolution cannot be statistically distinguished from its reversal in time. Technically for P to be time-reversible it is required that

$$\int g(\varphi)\Omega h(\varphi)P(\mathrm{d}\varphi) = \int h(\varphi)\Omega g(\varphi)P(\mathrm{d}\varphi), \tag{5.95}$$

where Ω is the closure of the infinitesimal generator of the birth-and-death process, defined for nonnegative measurable locally bounded g (i.e. there is a bounded Borel set A_g with $g(\varphi) = g(\varphi \cap A_g)$) by

$$\begin{aligned}\Omega g(\varphi) = &\int \boldsymbol{b}(x, \varphi)\big(g(\varphi \cup \{x\}) - g(\varphi)\big)\varphi(\mathrm{d}x)\\ &+ \int \boldsymbol{d}(x, \varphi\backslash\{x\})\big(g(\varphi\backslash\{x\}) - g(\varphi)\big)\varphi(\mathrm{d}x).\end{aligned}$$

For these definitions to be sensible, it is necessary to assume $\boldsymbol{b}$ and $\boldsymbol{d}$ are well behaved, and in particular that they generate a uniquely defined Markov process.

Glötzl (1981) presents a simple example: Assume $\boldsymbol{b}(x, \varphi) \equiv 1$ and $\boldsymbol{d} \equiv 1$. Then the generated process in any bounded set B is independent of what happens outside. After a random waiting time, which follows an exponential distribution with parameter $\nu_d(B)$ a new point is born in B at an uniformly distributed location in B, independently of other points. Each

point dies after a random lifetime, which is exponentially distributed with parameter 1 and independent of the behaviour of other points. In analogy to queueing theory, this model can be called *spatial M/M/∞ queue*. The shapshots of this process (i.e. the system of all points existing at the same instant), when in equilibrium, are Poisson point processes of intensity 1.

A similar model, but with lifetimes of the points following a general distribution is called *spatial M/G/∞ queue*; see Daley and Vere-Jones (2008). This latter model is not a birth-and-death process in the sense considered here, if the lifetimes are not exponentially distributed.

Under suitable conditions the spatial birth-and-death process will converge to a statistical equilibrium which is the time-reversible distribution P; see Preston (1976, Theorem 7.1). Lotwick and Silverman (1981) prove this by using coupling arguments; see also Møller (1999). If the birth-and-death process has a time-reversible equilibrium distribution P, then, whatever initial distribution is given, it converges to P. This fact forms the basis for a practical simulation procedure of stationary Gibbs processes as described in the next section.

Glötzl (1981) demonstrates the relationship between time-reversible birth-and-death process and Gibbs processes.

Theorem 5.2. *Let the rates* $\boldsymbol{b}$ *and* $\boldsymbol{d}$ *satisfy*

$$\frac{\boldsymbol{b}(x,\varphi)\boldsymbol{b}(y,\varphi\cup\{x\})}{\boldsymbol{d}(x,\varphi)\boldsymbol{d}(y,\varphi\cup\{x\})}=\frac{\boldsymbol{b}(y,\varphi)\boldsymbol{b}(x,\varphi\cup\{y\})}{\boldsymbol{d}(y,\varphi)\boldsymbol{d}(x,\varphi\cup\{y\})}$$

for all x *and* y *in* $\mathbb{R}^d$ *and* φ *in* $\mathbb{N}$.

Then the distribution P *is a time-reversible distribution of the spatial birth-and-death process with rates* $\boldsymbol{b}$ *and* $\boldsymbol{d}$ *if and only if* P *is the distribution of a Gibbs process based on a homogeneous Poisson weight process with intensity* 1 *and local energy*

$$E(x,\varphi)=-\ln\left(\frac{\boldsymbol{b}(x,\varphi)}{\boldsymbol{d}(x,\varphi)}\right)$$

for x *in* $\mathbb{R}^d$ *and* φ *in* $\mathbb{N}$.

5.5.5 Simulation of stationary Gibbs processes

Stationary Gibbs processes are also simulated by means of MCMC methods, and there are two widely used methods.

Application of birth-and-death processes

A first method uses spatial birth-and-death processes as the first 'MC' component. The simulation is carried out in a bounded region, usually a rectangle or cuboid W, with periodic boundary conditions. In this way one hopes to obtain point patterns that behave like samples of stationary point processes. The idea is to repeat the point patterns periodically outside of W. Figure 4.9 in Chapter 4 illustrates the periodic continuation for a rectangular window.

The simulation method is as follows:

(1) choose a starting configuration;

(2) choose the rates $\boldsymbol{b}$ and $\boldsymbol{d}$;

(3) simulate the birth-and-death process with these rates until it is considered sufficiently close to equilibrium.

Then Theorem 5.2 is applied to assert that the final configuration behaves approximately as a sample of the desired Gibbs process.

The starting configuration used in (1) can be of n points generated by the uniform distribution on W, where n is some estimate of the expected mean point number in W, or a lattice satisfying a hard-core condition or the empirical pattern which is under investigation.

The rates chosen in (2) can take many forms. The form which is most easily programmed is to take constant birth rate $\boldsymbol{b}(x, \varphi) = 1$ with new points distributed uniformly over W, and death rate $\boldsymbol{d}(x, \varphi) = \exp\big(E(x, \varphi)\big)$. However, this leads to slow convergence in many cases, as observed already by Ripley (1977). The reason is that new-born points are frequently removed almost immediately, in particular in the hard-core case. It is usually preferable to take constant death rate $\boldsymbol{d}(x, \varphi) = 1$ and birth rate proportional to the configuration $\boldsymbol{b}(x, \varphi) = \exp\big(-E(x, \varphi)\big)$. This means new points have to be generated by rejection sampling, as described below, but in practice the convergence to equilibrium is then faster, compensating for the more complicated way of simulating new points.

The simulation step (3) then takes place in continuous time running in variable time-steps from one incident of birth or death to the next, as follows.

Suppose that at time t a birth or death has just happened, and the configuration φ in W consists of n points $x_1, \ldots, x_n$. In the computation of the birth rate $\boldsymbol{b}(x, \varphi) = \exp\big(-E(x, \varphi)\big)$ the continuation of the point pattern outside of W must be taken into account. It must be determined whether the next incident is actually a birth or a death, when it is to happen, and (if it is a birth) where it is to happen or (if it is a death) which point is to be removed.

First, simulate the time T to the next incident as an exponential random variable with mean

$$\left(n + \int_W \exp\big(-E(x, \varphi)\big)\,\mathrm{d}x\right)^{-1},$$

evaluating the integral $I(\varphi) = \int_W \exp\big(-E(x, \varphi)\big)\,\mathrm{d}x$ numerically. Then use simulation to arrange for the next incident to be the death of point x_i, for $i = 1, \ldots,$ or n, each with probability

$$\frac{1}{n + I(\varphi)},$$

and a birth otherwise.

If it is to be a birth, then use rejection sampling to add a new point at x with probability density

$$\frac{\exp\big(-E(x, \varphi)\big)}{I(\varphi)}.$$

Thus it has been determined that the next birth or death has occurred at time $t + T$, and whether the next incident was a birth or a death, and which point has been affected. Now the whole procedure is repeated with, of course, recalculating $I(\varphi)$ whenever the point pattern φ is altered, and using independent random variables for each simulation step.

Application of the Metropolis–Hastings algorithm

An alternative to the birth-and-death-process simulation is the Metropolis–Hastings algorithm. Usually it is also carried out in a rectangular or cuboidal window W with periodic boundary conditions. Then methods for the simulation of Gibbs processes in bounded windows and random point number are applied. (Use of periodic boundary conditions ensures that the patterns look like realisations of a stationary point process, the random-number models lead to the density fluctuations which appear in stationary processes in bounded windows.)

The general idea of a Metropolis–Hastings simulation algorithm is that in each step a randomly chosen *proposal* for a change of the current configuration is made, and by some *acceptance rule* it is decided to accept the proposal or not. This rule makes that the resulting Markov chain has the target stationary distribution. Whether a proposal is accepted or not depends on chance.

For Gibbs processes it is based on the nonnormalised density

$$f_n(x_1, \ldots, x_n) = \exp\left(-\left(\alpha_f n + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \theta(\|x_i - x_j\|)\right)\right) \tag{5.96}$$

with $x_1, \ldots, x_n \in W$ for $n = 0, 1, \ldots$, where α_f is a parameter which controls the intensity in a way similar to the chemical activity α in (5.81); see p. 198.

The proposals are controlled by the following three simulation parameters:

- $b(\{x_1, \ldots, x_n\})$: the probability that the 'birth' of a new point will be proposed, while $1 - b(\{x_1, \ldots, x_n\})$ is the probability that the 'death' of a point will be proposed;
- $q_{\text{birth}}(\{x_1, \ldots, x_n\}; x)$: the proposal density function for the location of the new point x;
- $q_{\text{death}}(\{x_1, \ldots, x_n\}; x_k)$: the probability for the proposal to delete the point x_k from the set $\{x_1, \ldots, x_n\}$.

Simple concrete choices of the proposal parameters are as follows:

$$\begin{aligned} b(\{x_1, \ldots, x_n\}) &\equiv \frac{1}{2}\,, \\ q_{\text{birth}}(\{x_1, \ldots, x_n\}; x) &= \frac{1}{\nu(W)} \qquad \text{for } x \in W, \\ q_{\text{death}}(\{x_1, \ldots, x_n\}; x_k) &= \frac{1}{n} \qquad \text{for } x_k \in \{x_1, \ldots, x_n\}. \end{aligned}$$

The simulation starts with a point pattern $\{x_{01}, \ldots, x_{0m}\}$ for which

$$f_m(\{x_{01}, \ldots, x_{0m}\}) > 0\,,$$

that is, which does not contradict the basic properties of the process. (For example, it is a hard-core pattern if a hard-core process is to be simulated.)

Suppose that after l iteration steps the configuration is $\{x_1, \ldots, x_n\}$. In the $(l+1)^{\text{st}}$ step, initially a decision is made as to whether a new point may be added to the pattern (birth) or removed (death), with probability $b(\{x_1, \ldots, x_n\})$ or $1 - b(\{x_1, \ldots, x_n\})$, respectively.

If the $(l+1)^{\text{st}}$ step is a birth, the position of the new point $x \in W$ is proposed from the density function $q_{\text{birth}}(\{x_1, \ldots, x_n\}; x)$ and the proposal is accepted with probability

$$p_{\text{birth}} = \min\{1, \rho_{\text{birth}}\}, \tag{5.97}$$

where

$$\rho_{\text{birth}} = \frac{f_n(\{x_1, \ldots, x_n, x\})\big(1 - b(\{x_1, \ldots, x_n, x\})\big) q_{\text{death}}(\{x_1, \ldots, x_n, x\}; x)}{f_n(\{x_1, \ldots, x_n\}) b(\{x_1, \ldots, x_n\}) q_{\text{birth}}(\{x_1, \ldots, x_n\}; x)} \tag{5.98}$$

is the so-called Metropolis–Hastings birth ratio. With this ratio and the corresponding death ratio the Markov chain is controlled such that its stationary distribution is the desired point process distribution. If the proposal is accepted, the new point is added to the point configuration. If the proposal is not accepted, the configuration does not change in this step.

If the $(l+1)^{\text{st}}$ step is a death and if $\{x_1, \ldots, x_n\}$ is not empty, a point x_k in $\{x_1, \ldots, x_n\}$ is proposed to be deleted with probability $q_{\text{death}}(\{x_1, \ldots, x_n\}; x_k)$. This deletion is accepted with probability

$$p_{\text{death}} = \min\{1, \rho_{\text{death}}\}, \tag{5.99}$$

where

$$\rho_{\text{death}} = \frac{f_n(\{x_1, \ldots, x_n\} \setminus \{x_k\})\, b(\{x_1, \ldots, x_n\} \setminus \{x_k\})\, q_{\text{birth}}(\{x_1, \ldots, x_n\} \setminus \{x_k\}; x_k)}{f_n(\{x_1, \ldots, x_n\})\big(1 - b(\{x_1, \ldots, x_n\})\big)\, q_{\text{death}}(\{x_1, \ldots, x_n\}; x_k)} \tag{5.100}$$

is now the Metropolis–Hastings death ratio. Here $\{x_1, \ldots, x_n\} \setminus \{x_k\}$ denotes the point pattern $\{x_1, \ldots, x_n\}$ without the point x_k, $k \in \{1, \ldots, n\}$. If the proposal is accepted, the point x_k is removed from the configuration. If the proposal is not accepted, the configuration does not change in this step. (If the configuration was empty, the empty configuration does not change in a death step.)

Note that the Metropolis–Hastings algorithm works also in the nonhereditary case; see Dereudre and Lavancier (2011).

Example 5.11. *Simulation of a planar Strauss process with random number of points (Illian et al., 2008, p. 152)*

The Strauss process with parameters $\alpha_f = -8.0$, $b = -\exp(0.3) = -1.35$ and $\rho = 0.08$ is simulated in W, the unit square. The parameters of the Metropolis–Hastings algorithm are chosen as above.

Suppose the configuration at step l is $\{x_1, \ldots, x_n\}$. At first, a decision has to be made as whether to add or delete a point, either of which happens with probability $\frac{1}{2}$. If a point is to be added to the configuration, the candidate x is chosen uniformly within W. Then the value of the Metropolis–Hastings ratio for a birth is calculated,

$$\rho_{\text{birth}} = \frac{1}{n+1} \exp\left(-\alpha_f + b \sum_{i=1}^{n} \mathbf{1}(0 < \|x - x_i\| \le \rho)\right).$$

The proposal x is accepted with probability

$$p_{\text{birth}} = \min\{1, \rho_{\text{birth}}\}.$$

This means, a random number u is drawn from $[0, 1]$ and if $u \leq p_{\text{birth}}$, the proposal x is accepted and the new configuration is $\{x_1, \ldots, x_n, x\}$. If $u > p_{\text{birth}}$, the old configuration $\{x_1, \ldots, x_n\}$ does not change. (Note that a similar mechanism is used above to decide whether the next step is a potential birth or death.)

If a point is to be deleted, a number k is randomly chosen from $\{1, \ldots, n\}$ and x_k is proposed to be removed. The Metropolis–Hastings ratio for a death

$$\rho_{\text{death}} = n \exp\left(\alpha_f - b \sum_{i=1, i \neq k}^{n} \mathbf{1}(0 < \|x_k - x_i\| \leq \rho)\right)$$

and $p_{\text{death}} = \min\{1, \rho_{\text{death}}\}$ are evaluated. The proposal is accepted, that is, x_k is removed, with probability α_{death} as above. If the proposal is accepted, the new configuration is $\{x_1, \ldots, x_{k-1}, x_{k+1}, x_n\}$, otherwise the configuration does not change.

In this simulation example, the initial configuration is a realisation from a binomial process with 120 points. Figure 5.9 on p. 199 shows the simulation results after 100, 5000 and 10 000 iterations, where the numbers of points were by chance 120, 166 and 166, points respectively. (It is only by coincidence that the last two numbers are the same.) Figure 5.10 on p. 200 shows a trace plot of the evolution of the number of points in the 10 000 iteration steps.

In this particular case, the burn-in took around 2000 iterations after which the generated point patterns were considered to be realisations from the Strauss process.

Technical remarks on Gibbs process simulation

(1) In all kinds of simulation, the Gibbs process samples arise as equilibrium limits. In practice it suffices to continue the simulation until the approximation to statistical equilibrium is judged as acceptable. That means, a burn-in period must be accepted, the length of which is not known *a priori*. This implies that the user has to observe the simulation process in trace plots (Møller and Waagepetersen, 2004) and to decide if the algorithm is likely to have converged. In favourable cases the so-called 'perfect simulation' approach can be employed to avoid the burn-in problem: see for example Kendall and Møller (2000) and Huber (2011). Possible indicators are the current point number n or the numbers of births, deaths or moves in fixed numbers of simulation steps. The computing times can be long, in particular for Gibbs processes with repulsion and high point density.

A typical example of such a simulation might lead to the following: after perhaps $10n$ steps the influence of the starting configuration will have largely disappeared, where n is the average number of points in W in the stationary state. Patterns which are separated by $2n$, or perhaps $4n$ steps can then be considered as 'independent'. Practical experience shows that excessively long simulation times do not solve all problems: numerical effects can then make difficulties.

For example, in the case of a hard-core potential and high intensity (very negative α) the burn-in period may be very long and perhaps the packing algorithms as in Section 6.5.3 provide a better approach than MCMC methods.

(2) A difficult problem in the case of stationary Gibbs processes is related to chemical activity α and intensity λ. If the birth-and-death process simulation is used, α can be explicitly built in the algorithm; the intensity may be first unknown and is obtained (approximately) by

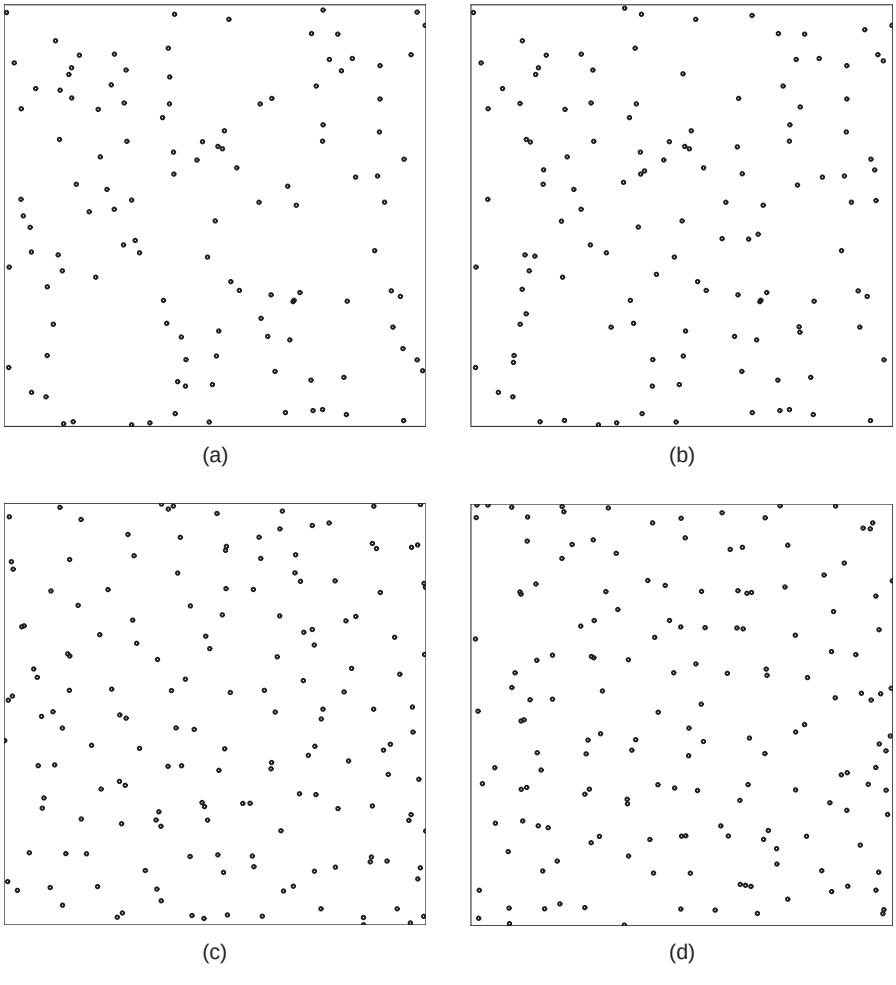

Figure 5.9 Simulated samples from a Strauss process with random number of points generated by the Metropolis–Hastings algorithm on the unit square. The parameters of the process and the specific implementation of the algorithm are described in Example 5.11. (a) Initial configuration ($n = 120$), configuration after (b) 100 steps ($n = 120$), (c) 5000 steps ($n = 166$) and (d) 10 000 steps ($n = 166$). Some tendency towards regularity can be detected but close pairs are also allowed. Reproduced from Illian *et al.* (2008, Figure 3.21).

statistical analysis of simulated samples. If the intensity is prescribed, α has to be determined by experimentation.

In the Metropolis–Hastings approach one has to determine that value of α_f which yields the desired intensity. Samples that behave (approximately) as realisations from a stationary Gibbs process can be obtained by simulating with fixed point number (defined by the desired intensity) and periodic boundary conditions but in an extended window W_{ext} containing the window W in which samples are needed. Then, as it must be, the point number in W shows random fluctuations, while the number in W_{ext} is fixed.

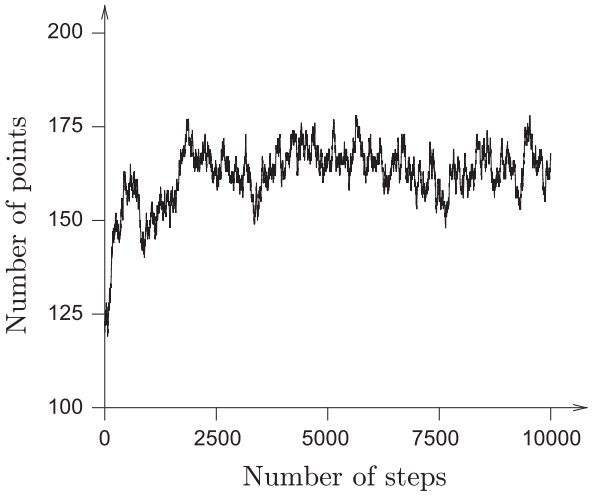

Figure 5.10 A trace plot for the number of points of a run of the Metropolis–Hastings algorithm described in Example 5.11. Reproduced from Illian *et al.* (2008, Figure 3.22(a)).

(3) Some kind of verification of the various approximations is necessary at the end of such simulations. Clearly, one can re-estimate the process parameters and see if the estimated parameters are close to those used in the simulation. Stoyan and Grabarnik (1991) suggested the marking of points of a stationary Gibbs process with 'energy marks': the point x_n has the mark

$$m_n = E(x_n, \Phi\backslash\{x_n\}) = \exp\left(\alpha + \sum_{y\in\Phi\backslash\{x_n\}} \theta(\|x_n - y\|)\right). \tag{5.101}$$

By means of Formula (5.79) it is easy to show that the mean of the marks is $1/\lambda$, where λ is the intensity. Thus a necessary condition for a 'good sample' is that the sum of all energy marks is approximately equal to the window area.

5.6 Shot-noise fields

5.6.1 Definition and examples

A random field $\{Z(x)\}$ is a family of random variables as explained in books on geostatistics. It has values $Z(x)$ in *all* $x \in \mathbb{R}^d$. An important class of random fields is formed by shot-noise fields $\{S_\Psi(x)\}$. Such a random field describes a structure which results from the superposition of (random) *impulses*, also called *responses* or *effects*, which are related to the points of a (marked) point process Ψ. The name origins from the one-dimensional case, where the points are instants of time and the impulses acoustic signals.

The impulses are usually assumed to be homogeneous, that is, they only depend on the mark of the point and the difference $x - x_n$ of the location x of interest and the point x_n. Then

$s(x - x_n, m_n)$ is the contribution of point x_n with mark m_n to $S_\Psi(x)$. Formally, $s(x, m)$ is the contribution of a point at origin o with mark m to the shot-noise field.

In many cases a suitable name for the impulse function $s(x, m)$ is *attenuation function,* as its value decreases with the distance of x from o and really describes an attenuation process.

Superposition of the impulses of all points of Ψ yields the *shot-noise field*:

$$S_\Psi(x) = \sum_{[x_n;m_n]\in\Psi} s(x - x_n, m_n) \qquad \text{for } x \in \mathbb{R}^d. \tag{5.102}$$

An important special case is

$$s(x - x_n, m_n) = m_n f(x - x_n), \tag{5.103}$$

where f is a probability density function. In this approach the 'point mass' m_n is continuously distributed around x_n.

However, there are also situations where the superposition follows the max-rule, that is, the value at x is the maximum of the $s(x - x_n, m_n)$,

$$M(x) = \max_{[x_n;m_n]\in\Psi} s(x - x_n, m_n) \qquad \text{for } x \in \mathbb{R}^d; \tag{5.104}$$

multiplicative superposition may also be considered, for example in the context of resource interference in ecology; see Wu *et al.* (1985).

Usually, the main interest is to determine characteristics of the field $\{S_\Psi(x)\}$. Some formulae are given in the next section for the stationary case and independent marks.

Example 5.12. *Competition load (Adler, 1996)*

The points of Ψ represent plants in a (planar) plant community, where each plant competes with its neighbours, and the marks are size parameters such as tree stem diameters. Clearly, the strength of the competition decreases with distance from the reference plant and a given point is influenced by the competition load from all its neighbours. The result is a random competition field $\{S_\Psi(x)\}$, where $S_\Psi(x)$ is the sum of the competition contributions from all plants in the vicinity of location x. In this context, the attenuation function $s(x, m)$ is called 'local competition function'. Here the value of the random field at a deterministic location x is probably less important than at the typical plant.

In Adler (1996) the following attenuation function

$$s(x, m) = m^\alpha \exp\left(-\delta \frac{\|x\|}{m^\beta}\right) \tag{5.105}$$

is considered, where α, β and δ are model parameters.

Example 5.13. *Interference field in wireless communication (Baccelli and Błaszczyszyn, 2009b)*

The points of Ψ are transmitters and the 'impulses' correspond to the signals emitted by antennas. The total power received from all transmitters at a given location x is of interest. The strength of the signals depends on the distances from the antennas and on noise.

The impulse function $s(x, m)$ can have the form

$$s(x, m) = \frac{m}{l(\|x\|)}, \tag{5.106}$$

which corresponds to the case of omnidirectional antennas, where l is the so-called *path-loss function*. An example is

$$l(r) = (Ar)^{\beta}, \tag{5.107}$$

called *simplified path-loss function*, where A is positive and β greater than 2. (In fact, this $l(r)$ is only a mathematical approximation; for small r the values are too large, and the mean of the corresponding random field $\{S_\Psi(x)\}$ is infinite.) An exponential mark distribution is also a popular choice.

Shot-noise fields are also used to construct point process models, the so-called *shot-noise Cox processes*; see Møller and Waagepetersen (2004), Illian *et al.* (2008) and Jalilian *et al.* (2013). In optics and meteorology, as well as in other fields of physics, shot-noise fields have also been successfully used; see Frieden (1983), Rodrigez-Itube *et al.* (1987, 1988) and Cox and Isham (1988).

5.6.2 Moment formulae for stationary shot-noise fields

In the stationary case, mean, variance and variogram (defined in Formula (6.89)) can be derived analytically for the shot-noise random fields, but the relevant equations are simple only for the mean. Despite their complexity the reader may get an impression of the structure of these equations, by considering the volume integrals for the second-order characteristics in this section, taken from Cox and Isham (1980) and Schmidt (1985).

In order to avoid overly complex formulae, assume that the marks are quantitative (real-valued) and independent, that the ground process Φ (Ψ without the marks) is motion-invariant and that the impulse function is rotation-invariant with respect to o, which is used to replace $s(x, m)$ by $s(r, m)$, where r is the distance of x from o.

With these assumptions the field $\{S_\Psi(x)\}$ is stationary and isotropic,

$$S_\Psi(x) = \sum_{[x_n;m_n]\in\Psi} s(x - x_n, m_n) \qquad \text{for } x \in \mathbb{R}^d. \tag{5.108}$$

This means that the value of the random field at location x is a sum of impulses $s(x - x_n, m_n)$ in the points x_n, which depend on their marks m_n.

Mean value:

$$\mathbf{E}\big(S_\Psi(o)\big) = \lambda \int_{\mathbb{R}^d} e(x)\mathrm{d}x \tag{5.109}$$

with

$$e(x) = \int_0^\infty s(x, m) f_{\mathcal{M}}(m)\mathrm{d}m, \tag{5.110}$$

where $f_{\mathcal{M}}(m)$ is the mark probability density function. Due to stationarity, the mean of $S_\Psi(x)$ is the same for all $x \in \mathbb{R}^d$, therefore only the value for $x = o$ has to be considered. Equation (5.109) is a simple consequence of the Campbell theorem for marked point processes.

For the Adler competition field the mean is

$$\mathbf{E}\big(S_\Psi(o)\big) = 2\pi\lambda \int_0^\infty \int_0^\infty m^\alpha r \exp\left(-\delta \frac{r}{m^\beta}\right) f_{\mathcal{M}}(m)\mathrm{d}r\mathrm{d}m = \frac{2\pi\lambda}{\delta^2}\,\mathbf{E}\left(\mathcal{M}^{\alpha+2\beta}\right), \quad (5.111)$$

where $\mathbf{E}\left(\mathcal{M}^{\alpha+2\beta}\right)$ denotes the $(\alpha + 2\beta)^{\text{th}}$ moment of the marks.

Variance:

$$\mathbf{var}\big(S_\Psi(o)\big) = \lambda \int_{\mathbb{R}^d} e^{(2)}(x)\mathrm{d}x + \lambda^2 \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e(x)e(x+h)\mathrm{d}x\,\mathcal{K}(dh) - \big(\mathbf{E}(S(o))\big)^2 \quad (5.112)$$

with $e(x)$ defined in (5.110) and

$$e^{(2)}(x) = \int_0^\infty \big(s(x,m)\big)^2 f_{\mathcal{M}}(m)\,\mathrm{d}m,$$

where $\mathcal{K}$ is the reduced second moment measure of the ground process Φ, as introduced in Section 4.5. For a homogeneous Poisson process Formula (5.112) simplifies to

$$\mathbf{var}\big(S_\Psi(o)\big) = \lambda \int_{\mathbb{R}^d} e^{(2)}(x)\mathrm{d}x, \quad (5.113)$$

since in this case $\mathcal{K} = \nu_d$. For the Adler field

$$\mathbf{var}\big(S(o)\big) = \frac{2\pi\lambda}{4\delta^2}\,\mathbf{E}\left(\mathcal{M}^{2(\alpha+\beta)}\right). \quad (5.114)$$

For the interference field similar formulas can be derived; see Baccelli and Błaszczyszyn (2009b). The Laplace transform of $S_\Psi(o)$ is also known in the case of a Poisson process.

Variogram:

$$\gamma(r) = \mathbf{var}\big(S_\Psi(o)\big) + \big(\mathbf{E}(S_\Psi(o))\big)^2 - \lambda \int_{\mathbb{R}^d} e_r(x)\mathrm{d}x - \lambda^2 \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e(x)e(x+h-\mathbf{r})\mathrm{d}x\mathcal{K}(\mathrm{d}h) \quad (5.115)$$

with $e(x)$ defined in (5.110) and

$$e_r(x) = \int_0^\infty s(x,m)s(x-\mathbf{r},m)f_{\mathcal{M}}(m)\mathrm{d}m, \quad (5.116)$$

where $\mathbf{r}$ is any point with distance r from the origin o.

Mean value of the shot-noise field at the typical point

In ecological modelling, the mean competition load at the typical point is of substantial interest, since it reflects the strength of the competition on a typical individual in the community, while

$\mathbf{E}\big(S_\Psi(o)\big)$ is a spatial mean reflecting the strength of the competition on potential individuals. In the planar case, again based on polar coordinates,

$$\mathbf{E}_o\big(S_\Psi(o)\big) = 2\pi\lambda \int_0^\infty rg(r) \int_0^\infty s(r, m) f_{\mathcal{M}}(m)\, \mathrm{d}m\mathrm{d}r. \tag{5.117}$$

For the Adler field

$$\mathbf{E}_o\big(S_\Psi(o)\big) = 2\pi\lambda \int_0^\infty \int_0^\infty m^\alpha rg(r) \exp\left(-\delta \frac{r}{m^\beta}\right) f_{\mathcal{M}}(m)\mathrm{d}r\,. \tag{5.118}$$

For a homogeneous Poisson process the pair correlation function $g(r)$ is equal to one and consequently

$$\mathbf{E}\big(S_\Psi(o)\big) = \mathbf{E}_o\big(S_\Psi(o)\big)\,, \tag{5.119}$$

which is again a special case of the Slivnyak–Mecke theorem.