CHAPTER 6

Models Constructed via Conditioning: Cox, Cluster, and Marked Point Processes

In this chapter, we bring together a number of the most widely used classes of point process models. Their common theme is the generation of the final model by a two-stage construction: first, the generation of an indexed family of processes, and then an operation applied to members of the family to produce the final process. The first two classes (Cox and cluster processes) extend the simple Poisson process in much the same way that the mixed and compound Poisson distributions extend the basic Poisson distribution. Independence plays a central role and leads to elegant results for moment and generating functional relationships. Both processes are used typically in contexts where the realizations are stationary and therefore define infinite collections of points. To deal with these issues, we anticipate the transition from finite to general point processes to be carried out in Chapter 9 and present in Section 6.1 a short review of some key results for more general point processes and random measures.

The third class of processes considered in this chapter represents a generalization in a different direction. In many situations, events are characterized by both a location and a weight or other distinguishing attribute. Such processes are already covered formally by the general theory, as they can be represented as a special type of point process on a product space. However, marked point processes are deserving of study in their own right because of their wide range of applications, such as in queueing theory, and their conceptual importance in contexts such as Palm theory (see [MKM] especially).

6.1. Infinite Point Families and Random Measures

Although the framework developed for finite point processes in Chapter 5 needs to be extended, it nevertheless contains the essential ingredients of the

more general theory. We retain the assumption that the points are located within a complete, separable metric space (c.s.m.s.) $\mathcal{X}$, and will generally interpret $\mathcal{X}$ as either $\mathbb{R}^1$ or $\mathbb{R}^2$.

The space $\mathcal{X}^{\cup}$ as in (5.3.10) is no longer the appropriate space for defining the realizations; instead we move to a description of the realizations in terms of *counting measures*, meaning measures whose values on Borel sets are nonnegative integers. The interpretation is that the value of the measure on such a set counts the number of points falling inside that set.

A basic assumption, which really defines the extent of current point process theory, is that the measures are *boundedly finite*: only a finite number of points fall inside any bounded set (i.e. there are no finite accumulation points). In the martingale language of Chapters 7 and 14, this is equivalent to requiring the realizations to be 'nonexplosive'. The space $\mathcal{X}^{\cup}$ is then replaced by the space[1] $\mathcal{N}_{\mathcal{X}}^{\#}$ of all *boundedly finite counting measures on* $\mathcal{X}$. A remarkable feature is that a relatively simple and natural distance between counting measures can be defined and allows $\mathcal{N}_{\mathcal{X}}^{\#}$ to be interpreted as a metric space in its own right. It then acquires a natural topology and a natural family of Borel sets $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#})$ that can be used to define measures on $\mathcal{N}_{\mathcal{X}}^{\#}$. We shall not give details here but refer to Chapter 9 and Appendix A2.6.

Thus, the way is open to formally introducing a point process on $\mathcal{X}$ as a *random counting measure on* $\mathcal{X}$, meaning technically a measurable mapping from a probability space $(\Omega, \mathcal{E}, \mathcal{P})$ into the space $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#}))$. Often, the latter space itself is taken as the *canonical probability space* for a point process on $\mathcal{X}$. Every distinct probability measure on $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#}))$ defines a distinct point process.

As in the finite case, specific examples of point processes are commonly specified by their *finite-dimensional distributions*, or *fidi* distributions for short. These can no longer be defined globally, as was done through the Janossy measures for a finite point process, but are introduced by specifying consistent joint distributions

$$P_k(A_1, \ldots, A_k; n_1, \ldots, n_k) = \Pr\{N(A_1) = n_1, \ldots, N(A_k) = n_k\} \qquad (6.1.1)$$

for the number of points in finite families of bounded Borel sets. Indeed, this was the way we introduced the Poisson process in Chapter 2.

Consistency here combines conditions of two types: first, the usual conditions (analogous to those for any stochastic process) for consistency of marginal distributions and invariance under simultaneous permutation of the sets and the numbers falling into them; second, conditions to ensure that the realizations are almost surely measures, namely that

$$N(A \cup B) = N(A) + N(B) \quad \text{a.s.} \qquad \text{and} \qquad N(A_n) \to 0 \quad \text{a.s.} \qquad (6.1.2)$$

[1] In this edition, we use $\mathcal{M}_{\mathcal{X}}^{\#}$ (and $\mathcal{N}_{\mathcal{X}}^{\#}$) to denote spaces of boundedly finite (counting) measures on $\mathcal{X}$ where in the first edition we used $\widehat{\mathcal{M}}_{\mathcal{X}}$ (and $\widehat{\mathcal{N}}_{\mathcal{X}}$), respectively.

for (respectively) all disjoint Borel sets A, B, and all sequences $\{A_n\}$ of Borel sets with $A_n \downarrow \emptyset$. These two conditions reduce to the requirements on the fidi distributions that, for all finite families of *disjoint* bounded Borel sets, $(A_1, \ldots, A_k)$,

$$\sum_{r=0}^{n} P_k(A_1, A_2, A_3, \ldots, A_k;\, n-r, r, n_3, \ldots, n_k) = P_{k-1}(A_1 \cup A_2, A_3, \ldots, A_k;\, n, n_3, \ldots, n_k), \tag{6.1.3}$$

and

$$P_1(A_k; 0) \to 1 \tag{6.1.4}$$

for all sequences of bounded Borel sets $\{A_k\}$ with $A_k \downarrow \emptyset$. Moreover, for point processes defined on Euclidean spaces, it is enough for these relationships to hold when the sets are bounded intervals.

Example 6.1(a) *Simple Poisson process on* $\mathbb{R}$. Recall equation (2.2.1):

$$\Pr\{N(a_i, b_i] = n_i,\ i = 1, \ldots, k\} = \prod_{i=1}^{k} \frac{[\lambda(b_i - a_i)]^{n_i}}{n_i!}\, \mathrm{e}^{-\lambda(b_i - a_i)}. \tag{6.1.5}$$

Consistency of the marginals means that if one of the variables, say $N(a_1, b_1]$, is integrated out (by summing over n_1), the resulting quantity is the joint probability corresponding to the remaining variables. Invariance under permutations of the variables means that if the sets and the number of points falling into them are written down in a different order, the resulting probability is not affected. In the present example, both conditions are obvious from the product form of the joint distributions. The additivity requirement (6.1.3) comes from the additivity property of the Poisson distribution: for Poisson random variables N_1 and N_2 that are independent (as is implied here by the product form of the distributions), their sum again has a Poisson distribution. Finally, (6.1.4) follows from the property $\mathrm{e}^{-\delta_n} \to 1$ when $\delta_n \to 0$. □

Moment measures, factorial moment measures, and probability generating functionals can be defined as in Sections 5.4 and 5.5. The main differences are that in defining the moment measures we should restrict ourselves to bounded sets and that in defining the p.g.fl. we should confine ourselves to functions h in $\mathcal{V}(\mathcal{X})$, the space of nonnegative, measurable functions bounded by unity and such that $1 - h(x)$ vanishes outside some bounded set. Within these constraints, the relations between generating functionals, moment measures, and all the various quantities derived from these in Chapter 5 hold much as they did there. A more detailed account, examining existence and convergence conditions, is given in Chapter 9.

For many of the examples that we consider, the point processes will be defined on a Euclidean space and *stationary*, meaning that their fidi distributions are invariant under simultaneous shifts of their arguments: writing

$A + u = \{x + u, x \in A\}$, stationarity means that, for all real u,

$$P_k(A_1, \ldots, A_k; n_1 \ldots n_k) = P_k(A_1 + u, \ldots, A_k + u; n_1, \ldots, n_k). \qquad (6.1.6)$$

The full consequences of this assumption are quite profound (see the foretaste in Chapter 3), but for the present it is enough to note the following.

Proposition 6.1.I (Stationarity Properties)**.**

(i) *A point process with p.g.fl. $G[h]$ is stationary if and only if for all real u, $G[(S_u h)] = G[h]$, where $(S_u h)(x) = h(x - u)$.*

(ii) *If a point process is stationary and the first-moment measure M_1 exists, then M_1 reduces to a multiple of the uniform measure (Lebesgue measure), $M_1(\mathrm{d}x) = m\,\ell(\mathrm{d}x) = m\,\mathrm{d}x$, say.*

(iii) *If a point process is stationary and the second-moment measure M_2 exists, then M_2 reduces to the product of a Lebesgue component along the diagonal $x = y$ and a reduced component*[2]*, $\breve{M}_2(\mathrm{d}u)$ say, where $u = x - y$, orthogonal to the diagonal.*

Proof. The fidi distributions as above are determined by the p.g.fl. and can be evaluated by taking h to be the sum of simple functions on disjoint sets; conversely, the fidi distributions determine the p.g.fl., which has the shift-invariance properties under stationarity. Property (ii) can be proved from Cauchy's functional equation (see Section 3.6), while property (iii) is the measure analogue of the familiar fact that the covariance function of a stationary time series is a function of the difference in the arguments only:

$$c(x, y) = \breve{c}(x - y).$$

Similar expressions for the moment densities follow from property (iii) whenever the moment measures have densities, but in general they have a singular component along the diagonal $x = y$, which reappears as an atom at the origin in the reduced measure $\breve{M}_2(\cdot)$ (see also Section 8.1). General routes to these reduced measures are provided by the factorization theorems in Section A2.7 or by the disintegration theory outlined in Section A1.4 (see Chapter 8 for further discussion and examples). Estimation of these reduced moment measures and their Fourier transforms (spectral measures) is a key issue in the statistical analysis of point process data and will be taken further in Chapter 8 and in more detail in Chapter 12. □

We shall also need the idea of a *random measure*, so we note some elementary properties. The general theory of random measures is so closely interwoven with point process theory that the two can hardly be separated. Point processes are indeed only a special class (integer-valued) of the former,

[2] In this edition, we use $\breve{M}_2(\cdot)$ and $\breve{C}_2(\cdot)$ to denote reduced second moment and covariance measures (and $\breve{m}$ and $\breve{c}$ for their densities) where in the first edition we wrote $\widehat{M}_2(\cdot)$ and $\widehat{C}(\cdot)$, etc.

and much of the general theory runs in parallel for both cases, a fact exploited more systematically in Chapter 9. Here we provide just sufficient background to handle some simple applications.

The formal definition of a random measure $\xi(\cdot)$ proceeds much as in the discussion for point processes given above. Once again, the realizations $\xi(\cdot)$ are required to be a.s. *boundedly finite* and *countably additive*, and their distributional properties are completely specified by their finite-dimensional distributions. Since the values of the measure are no longer integer-valued in general (although still nonnegative), these take the more general form

$$F_k(A_1, \ldots, A_k; x_1, \ldots, x_k) = \Pr\{\xi(A_i) \le x_i,\ i = 1, \ldots, k\}. \tag{6.1.7}$$

The moment measures are defined as for point processes, although the special role played by the factorial moment measures is not sustained, particularly when the realizations are continuous. In place of the p.g.fl., the most useful transform is the *Laplace functional*, defined for $f \in \mathrm{BM}_+(\mathcal{X})$, the space of all nonnegative $f \in \mathrm{BM}(\mathcal{X})$, by

$$L[f] \equiv L_\xi[f] = \mathrm{E}\big[\exp\big(-\textstyle\int_{\mathcal{X}} f(x)\,\xi(\mathrm{d}x)\big)\big]. \tag{6.1.8}$$

[We sometimes write L_ξ as a reminder of the random measure ξ to which the Laplace functional L relates and $\int f\,\mathrm{d}\xi$ as shorthand for the integral in (6.1.8).] Of course, the Laplace functional can also be defined for point processes and is therefore the natural tool when both are discussed together.

Although L_ξ defines (the fidi distributions of) a random measure ξ uniquely, via appropriate inversion theorems, there is no easy counterpart to the expansion of the p.g.fl. about the zero function as in equations (5.5.3). There is, however, a Taylor series expansion for the Laplace functional about $f \equiv 0$, corresponding to the p.g.fl. expansion about $h \equiv 1$. It takes the form

$$\begin{aligned} L[sf] = 1 - s\int_{\mathcal{X}} f(x)\,M_1(\mathrm{d}x) + \frac{s^2}{2!}\int_{\mathcal{X}^{(2)}} f(x_1)f(x_2)\,M_2(\mathrm{d}x_1 \times \mathrm{d}x_2) - \cdots \\ + \frac{(-s)^r}{r!}\int_{\mathcal{X}^{(r)}} f(x_1)\ldots f(x_r)\,M_r(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_r) + \cdots. \end{aligned} \tag{6.1.9}$$

This expression is just the expectation of the expansion of the ordinary Laplace transform of the linear functional $Y = \int_{\mathcal{X}} f(x)\,\xi(\mathrm{d}x)$. Its validity depends first on the existence of all moments of the random measure ξ and second on the convergence, typically in a disk around the origin $s = 0$ with radius determined by the length of the largest interval $(0, r)$ within which the Laplace transform is analytic. Finite Taylor series expansions, when just a limited number of moment measures exist, are possible for imaginary values of s, corresponding to the use of the *characteristic functional*, and are set out in Chapter 9.

EXAMPLE 6.1(b) *Gamma random measures* (*stationary case*). Suppose that the random variables $\xi(A_i)$ in (6.1.7) are independent for disjoint Borel sets A_i in $\mathbb{R}^d$ and have the gamma distributions with Laplace–Stieltjes transforms

$$\mathrm{E}(\mathrm{e}^{-s\xi(A_i)}) = \psi(A_i, s) = (1 + \lambda s)^{-\alpha\ell(A_i)} \qquad (\lambda > 0,\ \alpha > 0,\ \mathrm{Re}(s) \geq 0), \tag{6.1.10}$$

where $\ell(\cdot)$ denotes Lebesgue measure. By inspection, $\psi(A_i, s) \to 1$ as $s \to 0$, showing that $\xi(A)$ is a.s. finite for any fixed bounded set A. Then, since $\mathcal{X}$ is separable, it can be represented as a denumerable union $\bigcup A_i$ of such sets and

$$\Pr\{\text{at least one } \xi(A_i) \text{ is infinite}\} \leq \sum_{i=1}^{\infty} \Pr\{\xi(A_i) = \infty\} = 0.$$

As in the case of a Poisson process, additivity of ξ is a consequence of independence and the additivity property of the gamma distribution. Also, $\psi(A_i, s) \to 1$ as $\ell(A_i) \to 0$, implying the equivalent of (6.1.4), which guarantees *countable* additivity for ξ and is equivalent to stochastic continuity of the cumulative process $\xi((0,t])$ when the process is on $\mathbb{R}^1$.

The Laplace functional of ξ can be found by extending (6.1.10) to the case where f is a linear combination of indicator functions and generalizing: it takes the form

$$L[f] = \exp\left(-\int_{\mathcal{X}} \log[1 + \lambda f(x)]\, \alpha\, \ell(\mathrm{d}x) \right).$$

Expanding this expression as in (6.1.9) and examining the first and second coefficients, we find

$$\begin{aligned} \mathrm{E}\big(\xi(\mathrm{d}x)\big) &= \lambda\alpha\, \ell(\mathrm{d}x), \\ \mathrm{E}\big(\xi(\mathrm{d}x)\, \xi(\mathrm{d}y)\big) &= \lambda^2\alpha^2\, \ell(\mathrm{d}x)\, \ell(\mathrm{d}y) + \delta(x-y)\lambda^2\alpha\, \ell(\mathrm{d}x). \end{aligned} \tag{6.1.11}$$

Thus, the covariance measure for $\xi(\cdot)$ vanishes except for the diagonal component along $x = y$, or, equivalently, the reduced covariance measure is just an atom of mass $\lambda^2\alpha$ at the origin. These features are consequences of the independence of the increments and the purely atomic nature of the sample paths $\xi(\cdot)$, equivalent when $\mathcal{X} = \mathbb{R}^1$ to the pure jump character of the cumulative process (see Section 8.3 for further discussion). From these results, we can also confirm the expressions for the moments as follow directly from (6.1.10), namely

$$\mathrm{E}\xi(A) = \lambda\alpha\, \ell(A) \qquad \text{and} \qquad \mathrm{var}\, \xi(A) = \lambda^2\alpha\, \ell(A).$$

Exercise 6.1.1 gives a more general version of a gamma random measure. □

EXAMPLE 6.1(c) *Quadratic random measure*. Let $Z(t)$ be a Gaussian process with a.s. continuous trajectories, and consider, for any Borel set A, the set function

$$\xi(A) = \int_A Z^2(u)\, \mathrm{d}u.$$

Since Z is a.s. continuous, so is Z^2, so the integral is a.s. well defined and is additive on disjoint sets. In particular, when Z has zero mean, each value $Z^2(t)$ is proportional to a chi-square random variable, so $\xi(A)$ for suitably 'small' sets A is also approximately a chi-square r.v. Generally, $\xi(A)$ can be defined (being an integral) as a limit of linear combinations of $Z^2(t_i)$ for points t_i that become dense in A, and this is quadratic in the Z, hence the name. The random measure properties of ξ are discussed in more detail in Chapter 9. See Exercise 6.1.3 for the first two moments of ξ. □

The next example has a long history. It was originally introduced in early work by Campbell (1909) to describe the properties of thermionic noise in vacuum tubes. Moran (1968, pp. 417–423) gives further details and references. In his work, Campbell developed formulae for the moments, such as

$$\mathrm{E}\left[\int g(x)\, N(\mathrm{d}x)\right] = \int g(x)\, M(\mathrm{d}x),$$

which led Matthes *et al.* (1978) to adopt the term *Campbell measure* for the concept that underlies their treatment of moments and Palm distributions (see also Chapter 13). Since that time, the ideas have appeared repeatedly in applications [see e.g. Vere-Jones and Davies (1966), where the model is referred to as a 'trigger process' and used to describe earthquake clustering]. Here we introduce it as a prelude to the major theme of this chapter. It is, like the other models in the chapter, a two-stage model, for which we consider here only the first stage.

Example 6.1(d) *Intensity of a shot-noise process.* A model for a shot-noise process is that the observations are those of a Poisson point process with a random intensity $\lambda(\cdot)$ with the following structure. A stochastic process $\lambda(t)$ is formed as a filtered version of a simple stationary Poisson process $N(\cdot)$ on $\mathbb{R}$ at rate ν with typical realization $\{t_i\}$, the filtering being effected by

(1) a nonnegative function g that integrates to unity and vanishes on $(-\infty, 0]$, and
(2) random 'multiplier' effects, $\{Y_i\}$, a series of i.i.d. nonnegative random variables with common distribution $F(\cdot)$.

We then define $\lambda(t)$ by

$$\lambda(t) = \sum_{i:t_i<t} Y_i g(t - t_i) = \int_0^\infty Y(u) g(t-u)\, N(\mathrm{d}u), \tag{6.1.12}$$

where $Y(u)$ is a (fictitious) process of i.i.d. variables with distribution F.

Since $\lambda(t)$, when finite, is stationary in t and is measurable, it is locally integrable: indeed, since its arguments are nonnegative, if it has finite expectation it must be finite a.s. For Borel sets A, the integral

$$\xi(A) \equiv \int_A \lambda(u)\, \mathrm{d}u = \sum_i Y_i \int_{A+t_i} g(u)\, \mathrm{d}u$$

is then well defined, though possibly infinite (see Exercise 6.1.4).

The Laplace functional of ξ can be evaluated as follows. We require

$$L[f] = \mathrm{E}\Big[\exp\Big(-\int_{\mathbb{R}} f(u)\lambda(u)\,\mathrm{d}u\Big)\Big].$$

Now, from (6.1.12), the integral can be written as a sum of terms

$$\int_{\mathbb{R}} f(u)\lambda(u)\,\mathrm{d}u = \sum_i Y_i \int_{\mathbb{R}} f(u)g(u-t_i)\,\mathrm{d}u \equiv \sum_i Z_i, \quad \text{say.}$$

If the points t_i are treated as given (i.e. fixed), then the Z_i are independent and, with $\phi(\cdot)$ denoting the common Laplace–Stieltjes transform of the Y_i, $Z_i = Y_i \int_{\mathbb{R}} f(u)g(u-t_i)\,\mathrm{d}u$ has the transform

$$\begin{aligned}\mathrm{E}(\mathrm{e}^{-Z_i}) &= \mathrm{E}\big[\exp\big(-Y_i \textstyle\int_{\mathbb{R}} f(u)g(u-t_i)\,\mathrm{d}u\big)\big] \\ &= \phi\big[\textstyle\int_{\mathbb{R}} f(u)g(u-t_i)\,\mathrm{d}u\big] \;\equiv\; \zeta(t_i),\end{aligned}$$

say, which lies in $(0,1]$ because f, g and the Y_i are all nonnegative. Proceeding formally, the last three equations give us

$$\begin{aligned}L[f] &= \mathrm{E}\big[\textstyle\prod_{t_i\in N}\zeta(t_i)\big] \\ &= G_N[\zeta], \qquad\qquad \text{by definition of a p.g.fl.,} \\ &= \exp\big[\nu\textstyle\int_{\mathbb{R}}[\zeta(t)-1]\,\mathrm{d}t\big], \qquad G_N \text{ is the p.g.fl. of a Poisson process,} \\ &= \exp\big\{\nu\textstyle\int_{\mathbb{R}}\big[\phi\big(\int_{\mathbb{R}} f(u)g(u-t)\,\mathrm{d}u\big)-1\big]\,\mathrm{d}t\big\}.\end{aligned}$$

It is clear from the random measure analogue of Proposition 6.1.I that the random measure $\xi(\cdot)$ here is stationary (we can easily check that $L[S_u f] = L[f]$). With a view to applying the expansion (6.1.9), we find after some manipulation that $L[f]-1$ equals

$$\begin{aligned}&\nu\int\Big[-\mu_1\int f(u)g(u-t)\,\mathrm{d}u + \tfrac{1}{2}\mu_2\int f(u)g(u-t)\,\mathrm{d}u\int f(v)g(v-t)\,\mathrm{d}v - \cdots\Big]\mathrm{d}t \\ &\quad + \tfrac{1}{2}\nu^2\iint\Big[\mu_1^2\int f(u)g(u-t)\,\mathrm{d}u\int f(v)g(v-s)\,\mathrm{d}v + \cdots\Big]\,\mathrm{d}t\,\mathrm{d}s + \cdots,\end{aligned}$$

where $\mu_j = \mathrm{E}(Y^j)$ for $j = 1, 2$. Collect terms, identify the measures associated with first and second powers of $f(\cdot)$, and recall that $\int_{-\infty}^{\infty} g(u)\,\mathrm{d}u = 1$ and $g(u) = 0$ for $u < 0$; then

$$\begin{aligned}M_1(\mathrm{d}t) &= \nu\mu_1\,\mathrm{d}t, \\ M_2(\mathrm{d}s\times\mathrm{d}t) &= \Big[\nu^2\mu_1^2 + \nu\mu_2\int_{-\infty}^{\min(s,t)} g(s-u)g(t-u)\,\mathrm{d}u\Big]\,\mathrm{d}s\,\mathrm{d}t,\end{aligned}$$

so that M_1 has constant density $\nu\mu_1$ and M_2 has the density

$$m(s,t) = \breve{m}_2(v) = \nu^2\mu_1^2 + \nu\mu_2\int_0^{\infty} g(y)g(y+|v|)\,\mathrm{d}y, \qquad \text{where } v = s-t.$$

The fact that M_2 is absolutely continuous stems from the absolute continuity of the trajectories. The appearance of the reduced density $\breve{m}_2$ here is characteristic of the stationary form of the moment measures (see Proposition 8.1.I and onward). □

While these arguments appear intuitively reasonable, to make them rigorous we must check two further points. First, we must establish that the random measure ξ is well defined in the sense that, despite the infinite sums in the definition, the realizations are a.s. boundedly finite; see Exercise 6.1.4.

Second, the implicit conditioning step, consisting here of being given a realization $\{t_i\}$ of the Poisson process and then taking expectations over such realizations, needs to be justified. In a more general context, this task hinges on the technical concept of measurability and is the subject of the next proposition; it appears repeatedly in this and later chapters.

As in Example 6.1(d), the models considered in this chapter are defined in two steps: first, an initial process is laid down and then a secondary process is defined, with distributions conditional on the realization of the initial process. The existence and other properties of such processes depend on extensions of standard theorems concerning the structure of bivariate distributions. Because a realization of a point process (or indeed a more general random measure) can be thought of as a point in a metric space, the same basic apparatus for describing the distributions conditional on the realization of a random measure is available as for dealing with bivariate distributions in $\mathbb{R}^2$. A general discussion of conditions for a bivariate random system in which each component takes its value in a c.s.m.s. is in Proposition A1.5.II. To apply the concepts in a point process context, the key idea we utilize is that of a *measurable family* of point processes or random measures.

Suppose there is given a family $\{N(\cdot \mid y): y \in \mathcal{Y}\}$ of point processes taking their values in the c.s.m.s. $\mathcal{X}$ and indexed by the elements y of the c.s.m.s. $\mathcal{Y}$. This family forms a *measurable family* if, for each set A in $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#})$, the function $\mathcal{P}(A \mid y)$ is $\mathcal{B}(\mathcal{Y})$-measurable, where

$$\mathcal{P}(A \mid y) = \Pr\{N(\cdot \mid y) \in A\}. \tag{6.1.13}$$

As in Proposition A1.5.II, we average across a measurable family of point processes to form a new point process as a mixture of the originals.

Proposition 6.1.II. *Suppose there is given*

(a) *a measurable family of point processes $\mathcal{P}(A \mid y)$, defined on the c.s.m.s. $\mathcal{X}$ and indexed by elements of $\mathcal{Y}$, and*

(b) *a $\mathcal{Y}$-valued random variable Y with distribution Π on $\mathcal{B}(\mathcal{Y})$.*

Then the integrals

$$\mathcal{P}(A) = \mathrm{E}[\mathcal{P}(A \mid Y)] = \int_{\mathcal{Y}} \mathcal{P}(A \mid y)\, \Pi(\mathrm{d}y) \tag{6.1.14}$$

define a probability measure $\mathcal{P}$ on $\mathcal{B}(\mathcal{X})$ and hence a point process on $\mathcal{X}$.

Corresponding concepts can readily be defined for random measures and are set out in Exercise 6.1.5.

The next lemma gives simple sufficient conditions for checking whether an indexed family of point processes forms a measurable family.

Lemma 6.1.III. *Each of the following conditions is necessary and sufficient to define a measurable family of point processes on a Euclidean space:*
(a) *for all choices of positive integer k, finite unions of disjoint intervals $(B_1, \ldots, B_k)$, and nonnegative integers $(n_1, \ldots, n_k)$, the fidi probabilities $P_k(B_1 \ldots, B_k; n_1, \ldots, n_k \mid y)$ are $\mathcal{B}(\mathcal{Y})$-measurable functions of y;*
(b) *for all functions h in the space $\mathcal{V}(\mathcal{X})$, the p.g.fl. $G[h \mid y]$ is a $\mathcal{B}(\mathcal{Y})$-measurable function of y.*

Proof. Denote by $\mathcal{A}$ the class of subsets A of $\mathcal{N}_\mathcal{X}$ for which $\mathcal{P}(A \mid y)$ is measurable in y with respect to $\mathcal{B}(\mathcal{Y})$. If (a) holds, then $\mathcal{A}$ contains the cylinder sets used in defining the fidi probabilities. It follows from the closure properties of families of measurable functions (see Appendix A1.4) that the class $\mathcal{A}$ is closed under monotone limits and therefore contains the σ-field of all subsets of $\mathcal{X}$ generated by the cylinder sets; that is, $\mathcal{A} \supseteq \mathcal{B}(\mathcal{X})$. Hence the given family of point processes forms a measurable family. If, alternatively, (b) holds, then by taking h to be a linear combination of indicator functions and differentiating, we can recover the fidi distributions. Differentiation and the other operations involved preserve measurability so that the result follows from (a). The necessity of (a) is obvious, and that of (b) follows on observing that $G[h \mid y]$ for a general $h \in \mathcal{V}(\mathcal{X})$ can be obtained from the case where h is a linear combination of indicator functions by operations that preserve the measurability in y. □

We can immediately apply this lemma to give sufficient conditions that are simpler to check than those of Proposition 6.1.II.

Corollary 6.1.IV. *Suppose there is given a $\mathcal{Y}$-valued random variable Y with distribution Π on $\mathcal{B}(\mathcal{Y})$ and either*
(a) *a family of fidi probabilities $P_k(B_1 \ldots, B_k; n_1, \ldots, n_k \mid y)$ satisfying condition* (a) *of Lemma* 6.1.III *or*
(b) *a family of p.g.fl.s $G[h \mid y]$ satisfying condition* (b) *of Lemma* 6.1.III.
For each of these cases, there exists a well-defined point process on $\mathcal{X}$ for which in case (a) *the fidi probabilities are given by*

$$P_k(B_1, \ldots, B_k; n_1, \ldots, n_k) = \mathrm{E}\big[P_k(B_1, \ldots, B_k; n_1, \ldots, n_k \mid Y)\big] = \int_\mathcal{Y} P_k(B_1, \ldots, B_k; n_1, \ldots, n_k \mid y)\, \Pi(\mathrm{d}y) \tag{6.1.15a}$$

and in case (b) *the p.g.fl. is given by*

$$G[h] = \mathrm{E}\big(G[h \mid Y]\big) = \int_\mathcal{Y} G[h \mid y]\, \Pi(\mathrm{d}y). \tag{6.1.15b}$$

The following is perhaps the simplest example to which these ideas apply; their applications will be explored more systematically in the next two sections.

Example 6.1(e) *Mixed Poisson process.* Take the distributions (6.1.5) as a candidate for a measurable family, with the role of y played by λ and that of $\mathcal{Y}$ played by the half-line $\mathbb{R}_+ = [0, \infty)$. For a fixed set of half-open intervals, the function (6.1.5) is a continuous and hence a measurable function of λ so that condition (a) of Lemma 6.1.III is satisfied. Thus, the simple Poisson processes form a measurable family with respect to the real variable λ. Consequently, we can mix (average) them with respect to a distribution Π for λ to obtain the fidi distributions of a new point process. If, for example, Π is the exponential distribution with density $\mu e^{-\mu\lambda}\,d\lambda$, then the number of points falling into any given set A has a geometric distribution $p_n = qp^n$ with parameter $p = \mu/(\mu + |A|)$, $q = 1 - p$. Moreover, the locations of the points in A, given the number of events in A, are uniformly distributed over A.

Alternatively, we could work from the p.g.fl. for the Poisson process, namely $G[h] = \exp\big(-\lambda \int [1 - h(u)]\,du\big)$, and take expectations over λ using condition (b) of the lemma and Corollary 6.1.IV. The resultant process has p.g.fl.

$$G[h] = \int_0^\infty \exp\big(-\lambda \textstyle\int [1 - h(u)]\,du\big)\,\Pi(d\lambda) = \Pi^*\big(\textstyle\int [1 - h(u)]\,du\big), \quad (6.1.16)$$

where $\Pi^*(\theta) = E(e^{-\theta Y})$ is the Laplace–Stieltjes transform of an r.v. Y with distribution Π. In particular, when Π is exponential with mean $1/\mu$, the p.g.fl. reduces to

$$G[h] = \frac{\mu}{\mu + \int [1 - h(u)]\,du}.$$

This reduces to the p.g.f. $\mu/\big[\mu + |A|(1 - z)\big]$ of the geometric distribution described above when we set $h(u) = 1 - (1 - z)I_A(u)$. □

Exercises and Complements to Section 6.1

6.1.1 A *general gamma random measure* on the c.s.m.s. $\mathcal{X}$ can be constructed as a process with independent nonnegative increments for which the increment $\xi(A)$ on the bounded Borel set A has a gamma distribution with Laplace transform

$$E(e^{-s\xi(A)}) = (1 + \lambda s)^{-\alpha(A)},$$

where the *scale* parameter λ is finite and positive and the *shape* parameter measure $\alpha(\cdot)$ is a boundedly finite measure on $\mathcal{B}_\mathcal{X}$.

(a) Verify that these marginal distributions, coupled with the independent increment property, lead to a well-defined random measure.

(b) In the case $\mathcal{X} = \mathbb{R}$, show that $\xi(\cdot)$ may be regarded as the increments of an underlying nondecreasing stochastic process $X(t)$, which with positive probability is discontinuous at t if and only if $\alpha(\{t\}) > 0$.

(c) Show that ξ has as its Laplace functional

$$L[f] = \exp\left(-\int_\mathcal{X} \log(1 + \lambda f(x))\,\alpha(dx)\right) \qquad (f \in \mathrm{BM}_+(\mathcal{X})).$$

[*Hint*: See Chapter 9 for more detail, especially parts (b) and (c).]

6.1.2 *Stable random measure.* Consider a random measure ξ for which $\mathrm{E}(\mathrm{e}^{-s\xi(A)}) = (1+[\exp(-s^\alpha)])^{-\Lambda(A)}$ for some fixed measure $\Lambda(\cdot)$ and that has independence properties as in Example 6.1(a). Verify that for $0 < \alpha < 1$, there is a well-defined random measure with marginal distributions as stated.

6.1.3 Let ξ be the quadratic random measure of Example 6.1(c) in which the Gaussian process Z is stationary with zero mean, variance σ^2 and $\mathrm{cov}(Z(s), Z(t)) = c(s-t)$. Show that for bounded Borel sets A and B,

$$\mathrm{E}[\xi(A)] = \sigma^2 \ell(A),$$

$$\mathrm{cov}(\xi(A), \xi(B)) = 2\int_A \int_B c^2(u-t)\,\mathrm{d}u\,\mathrm{d}t.$$

6.1.4 *Random measure and shot noise.* Denote by $\{x_i\}$ the points of a stationary Poisson process on $\mathbb{R}$ with rate parameter ν, and let $\{Y_j : j = 0, \pm 1, \ldots\}$ denote a sequence of i.i.d. r.v.s independent of $\{x_j\}$. Let the function g be as in Example 6.1(d). Investigate conditions under which the formally defined process

$$Y(t) = \sum_{x_j \le t} Y_j g(t - x_j)$$

is indeed well defined (e.g. by demanding that the series is absolutely convergent a.s.). Show that sufficient conditions are that

(a) $\mathrm{E}|Y| < \infty$, or else

(b) $g(\cdot)$ is nonincreasing on $\mathbb{R}_+$ and there is an increasing nonnegative function $\tilde{g}(\cdot)$ with $\tilde{g}(t) \to \infty$ as $t \to \infty$ such that $\int_0^\infty \tilde{g}(t)g(t)\,\mathrm{d}t < \infty$ and whose inverse $\tilde{g}^{-1}(\cdot)$ satisfies $\mathrm{E}\tilde{g}^{-1}(|Y|) < \infty$ [see also Daley (1981)].

6.1.5 Write down conditions, analogous to (6.1.13), for a measurable family of random measures, and establish the analogue of Proposition 6.1.II for random measures. Frame sufficient conditions for the existence of a two-stage process similar to those in Lemma 6.1.III and Corollary 6.1.IV but using the Laplace functional in place of the p.g.fl.

6.1.6 Let ξ be a random measure on $\mathcal{X} = \mathbb{R}^d$. For a nonnegative bounded measurable function g, define $G(A) = \int_A g(x)\,\ell(\mathrm{d}x)$ $(A \in \mathcal{B}_\mathcal{X})$, where ℓ denotes Lebesgue measure on $\mathbb{R}^d$, and

$$\eta(A) = \int_\mathcal{X} G(A - x)\,\xi(\mathrm{d}x).$$

(a) Show that $\eta(A)$ is an a.s. finite-valued r.v. for bounded $A \in \mathcal{B}_\mathcal{X}$ and that it is a.s. countably additive on $\mathcal{B}_\mathcal{X}$. Then, the existence theorems in Chapter 9 can be invoked to show that η is a well-defined random measure.

(b) Show that if ξ has moment measures up to order k, so does η, and find the relation between them. Verify that the kth moment measure of η is absolutely continuous with respect to Lebesgue measure on $(\mathbb{R}^d)^{(k)}$.

(c) Denoting the characteristic functionals of ξ and η by $\Phi_\xi[\cdot]$ and $\Phi_\eta[\cdot]$, show that, for $f \in \mathrm{BM}_+(\mathcal{X})$,

$$h(x) = \int_\mathcal{X} f(y)g(y - x)\,\mathrm{d}y$$

is also in $\mathrm{BM}_+(\mathcal{X})$, and $\Phi_\eta[f] = \Phi_\xi[h]$.

6.1.7 (*Continuation*). By its very definition, η is a.s. absolutely continuous with respect to Lebesgue measure, and when ξ is completely random, its density

$$Y(t) \equiv \int_{\mathcal{X}} g(t-x)\,\xi(\mathrm{d}x)$$

is called a *linear process*. [The shot-noise process noted in (6.1.12) is an example; for other references, see e.g. Westcott (1970).] Find the characteristic functional of Y when ξ is a stationary gamma random measure.

6.2. Cox (Doubly Stochastic Poisson) Processes

The doubly stochastic Poisson process—or, more briefly, the *Cox process*, so named in recognition of its appearance in a seminal paper of Cox (1955)—is obtained by randomizing the parameter measure in a Poisson process. It is thus a direct generalization of the mixed Poisson process in Example 6.1(e). We first give a definition, then discuss the consequences of the structural features it incorporates, and finally in Proposition 6.2.II give a more mathematical definition together with a list of properties.

Definition 6.2.I. *Let ξ be a random measure on $\mathcal{X}$. A point process N on $\mathcal{X}$ is a* Cox process directed by ξ *when, conditional on ξ, realizations of N are those of a Poisson process $N(\cdot \mid \xi)$ on $\mathcal{X}$ with parameter measure ξ.*

We must check that such a process is indeed well defined. The probabilities in the Poisson process $N(\cdot \mid \xi)$ are readily seen to be measurable functions of ξ; for example, $P(A;n) = [\xi(A)]^n \mathrm{e}^{-\xi(A)}/n!$ is a measurable function of $\xi(A)$, which in turn is a measurable function of ξ as an element in the metric space $\mathcal{M}_{\mathcal{X}}^{\#}$ of boundedly finite measures on $\mathcal{X}$; hence, we can apply Corollary 6.1.IV(a) and take expectations with respect to the distribution of ξ to obtain a well-defined 'mixed' point process on $\mathcal{X}$.

The finite-dimensional (i.e. fidi) distributions are easily obtained in terms of the distributions of the underlying directing measure ξ and are all of mixed Poisson type. Thus, for example,

$$P(A;k) = \Pr\{N(A)=k\} = \mathrm{E}\left(\frac{[\xi(A)]^k}{k!}\,\mathrm{e}^{-\xi(A)}\right) = \int_0^\infty \frac{x^k}{k!}\,\mathrm{e}^{-x}\,F_A(\mathrm{d}x), \tag{6.2.1}$$

where F_A is the distribution function for the random mass $\xi(A)$.

The factorial moment measures of the Cox process turn out to be the ordinary moment measures of the directing measure; this is because the factorial moment measures for the Poisson process are powers of the directing measure. Thus, denoting by μ_k and γ_k the ordinary moment and cumulant measures for ξ, we have for $k=2$,

$$M_{[2]}(A\times A) = \mathrm{E}\big(\mathrm{E}[N(A)(N(A)-1)\mid\xi]\big) = \mathrm{E}\big([\xi(A)]^2\big) = \mu_2(A\times A)\,,$$

and similarly for the covariance measures

$$C_{[2]}(A \times A) = \gamma_2(A \times A)\,.$$

The algebraic details are most easily handled via the p.g.fl. approach outlined in Corollary 6.1.IV(b). As a function of the parameter measure ξ, the p.g.fl. of the Poisson process can be written, for $h \in \mathcal{V}(\mathcal{X})$, as

$$G[h \mid \xi] = \exp\big(-\textstyle\int_{\mathcal{X}}[1-h(x)]\,\xi(\mathrm{d}x)\big). \tag{6.2.2}$$

For fixed h, this is a measurable function of ξ as an element of $\mathcal{M}_{\mathcal{X}}$. Thus, the family of p.g.fl.s (6.2.2) is a measurable family in the sense of Corollary 6.1.IV(b), which implies that we can indeed construct the p.g.fl of a point process by taking expectations in (6.2.2) with respect to any probability measure for ξ in $\mathcal{M}_{\mathcal{X}}$. The expectation

$$\mathrm{E}\left[\exp\left(-\int_{\mathcal{X}}[1-h(x)]\,\xi(\mathrm{d}x)\right)\right],$$

however, can be identified with the Laplace functional [see (6.1.8)] of the random measure ξ, evaluated at the function $[1-h(x)]$. This establishes the first part of the proposition below. The remaining parts are illustrated above for particular cases and are left for the reader to check in general.

Proposition 6.2.II. *Let ξ be a random measure on the c.s.m.s. $\mathcal{X}$ and L_ξ its Laplace functional. Then, the p.g.fl. of the Cox process directed by the random measure ξ is given by*

$$G[h] = \mathrm{E}\big[\exp\big(\textstyle\int_{\mathcal{X}}[h(x)-1]\,\xi(\mathrm{d}x)\big)\big] = L_\xi[1-h]. \tag{6.2.3}$$

The fidi distributions of a Cox process are of mixed Poisson type, as in (6.2.1); *its moment measures exist up to order n if and only if the same is true for ξ. When finite, the kth factorial moment measure $M_{[k]}$ for the Cox process equals the corresponding ordinary moment measure μ_k for ξ. Similarly, the kth factorial cumulant measure $C_{[k]}$ of the Cox process equals the corresponding ordinary cumulant measure γ_k for ξ.*

Note that this last result implies that the second cumulant measure of a Cox process is nonnegative-definite (see Chapter 8). Also, for bounded $A \in \mathcal{B}_{\mathcal{X}}$,

$$\begin{aligned}\operatorname{var} N(A) &= M_{[1]}(A) + C_{[2]}(A \times A)\\ &= M_{[1]}(A) + \operatorname{var}\big(\xi(A)\big) \;\geq\; M_{[1]}(A) \;=\; \mathrm{E}N(A),\end{aligned}$$

so a Cox process, like a Poisson cluster process, is overdispersed relative to the Poisson process.

Example 6.2(a) *Shot-noise or trigger process* [see Example 6.1(d) and Lowen and Teich (1990)]. We continue the discussion of this example by supposing the (random) function

$$\lambda(t) = \sum_{i:x_i<t} Y_i g(t-x_i) \tag{6.2.4}$$

to be the density of the random measure directing the observed Poisson process. In more picturesque language, the epochs $\{x_i\}$ are trigger events with respective sizes (or weights) $\{Y_i\}$ that decay according to the function g. Note that in the definition it is not necessary to assume that g decays monotonically: integrability is sufficient (see also Exercise 6.1.4).

Now we use the generating function formalism to obtain some elementary properties of the shot-noise process. Conditional on the sequence $\{(x_i, Y_i)\}$, we can appeal to (6.2.2) and write

$$G[h \mid \{(x_i, Y_i)\}] = \exp\left(\sum_i -Y_i \int_{x_i}^{\infty} [1 - h(t)]\, g(t - x_i)\, \mathrm{d}t \right). \tag{6.2.5}$$

Write $\phi(\theta) = \mathrm{E}(\mathrm{e}^{-\theta Y_1})$ for the common Laplace–Stieltjes transform of the $\{Y_i\}$. Taking expectations in (6.2.5) first with respect to $\{Y_i\}$ and then with respect to $\{x_i\}$, we have for the p.g.fl. of the process

$$\begin{aligned} G[h] &= \mathrm{E}\left(\prod_i \phi\Big(\textstyle\int_{x_i}^{\infty} [1 - h(t)]\, g(t - x_i)\, \mathrm{d}t \Big) \right) \\ &= \exp\left(\nu \int_{\mathbb{R}} \Big[\phi\Big(\textstyle\int_x^{\infty} [1 - h(t)]\, g(t - x)\, \mathrm{d}t \Big) - 1 \Big]\, \mathrm{d}x \right). \end{aligned} \tag{6.2.6}$$

By taking logarithms in this expression and expanding, it follows that the point process has factorial cumulant measures existing to as many orders as the r.v.s Y_i have finite moments, as is consistent with Proposition 6.2.II. It also follows that these moment measures are absolutely continuous with densities

$$\begin{aligned} m_1 &= \nu\mu_1 \int_0^{\infty} g(u)\, \mathrm{d}u, \\ c_{[2]}(t_1, t_2) &= \breve{c}_{[2]}(t_1 - t_2) \equiv \breve{c}_{[2]}(t_1') = \nu\mu_2 \int_0^{\infty} g(u) g(t_1' + u)\, \mathrm{d}u, \\ c_{[k]}(t_1, \ldots, t_k) &= \breve{c}_{[k]}(t_1', \ldots, t_{k-1}') \\ &= \nu\mu_k \int_0^{\infty} g(u) g(t_1' + u) \cdots g(t_{k-1}' + u)\, \mathrm{d}u, \end{aligned}$$

where $t_j' = t_j - t_k$ $(j = 1, \ldots, k-1)$ and $\mu_k = \mathrm{E}(Y^k)$. These relations are analogues of Campbell's formulae in the theory of shot noise (see references preceding Example 6.1(c)), while the first two illustrate the proposition insofar as the right-hand sides represent the ordinary cumulants of the directing shot-noise process. The fact that they are absolutely continuous reflects the same property in the realizations of ξ.

The representation (6.2.6) shows that the process can equally be regarded as a Neyman–Scott Poisson cluster process [see Example 6.3(a)]. The fact that the shot-noise process and the associated Neyman–Scott process have

the same p.g.fl. means that they are identical as point processes: no measurements on the point process can distinguish the clustering and doubly stochastic (or Cox) interpretations. This ambiguity of interpretation is an extension of the corresponding ambiguity concerning the dual interpretation of contagious distributions alluded to in Exercise 1.2.3. The possibility of such dual interpretations is not restricted to cluster processes: for example, Exercise 6.2.1 sketches a nontrivial characterization of the class of renewal processes that can be represented as Cox processes. □

EXAMPLE 6.2(b) *Boson processes* (Macchi, 1971a, 1975) [see Example 5.4(c)]. In optical problems concerning light beams of low density, the particulate aspects of light are important, and the emission or reception of individual photons (or more generally bosons) can be treated as a point process in time, or space, or both. A standard approach to modelling this situation is to treat the photon process as a Cox process directed by the fluctuating intensity of the light beam, with this latter phenomenon modelled as the squared modulus of a complex Gaussian process. Thus, for the (density of the) random intensity, we take the function

$$\lambda(t) = \lambda|X(t)|^2 \qquad (\lambda > 0), \tag{6.2.7}$$

where $X(\cdot)$ is a complex Gaussian process with zero mean and complex covariance function $C(s,t)$. The process $\lambda(\cdot)$ is similar to the quadratic random measure discussed in Example 6.1(c) with appropriate attention given to the conventions regarding a complex Gaussian process. These require that $X(t) = U(t) + iV(t)$, where $U(\cdot)$ and $V(\cdot)$ are real Gaussian processes such that

$$\begin{aligned}
\mathrm{E}\big(U(s)U(t)\big) &= \mathrm{E}\big(V(s)V(t)\big) = C_1(s,t),\\
\mathrm{E}\big(U(s)V(t)\big) &= -\mathrm{E}\big(U(t)V(s)\big) = C_2(s,t),\\
C(s,t) &= \mathrm{E}\big(\overline{X}(s)X(t)\big) = 2\big(C_1(s,t) + iC_2(s,t)\big).
\end{aligned}$$

Here it is to be understood that C_1 is real, symmetric, and nonnegative-definite, while C_2 is antisymmetric (so, in particular, $C_2(s,s) = 0$, and $\mathrm{E}[X(s)X(t)] = 0$ for all s,t).

The moments of the process $\lambda(\cdot)$ are given by a classical result concerning the even moments of a complex Gaussian process (see e.g. Goodman and Dubman, 1969)

$$\begin{aligned}
\mathrm{E}\big(\overline{X}(s_1)\cdots\overline{X}(s_k)X(t_1)\cdots X(t_k)\big) &= {}^{+}\begin{vmatrix} C(s_1,t_1) & \cdots & C(s_1,t_k)\\ \vdots & \ddots & \vdots\\ C(s_k,t_1) & \cdots & C(s_k,t_k)\end{vmatrix}^{+}\\
&= C^{+}\begin{pmatrix} s_1,\ldots,s_k\\ t_1,\ldots,t_k\end{pmatrix},
\end{aligned} \tag{6.2.8}$$

where the permanent $\mathrm{per}\, B \equiv {}^{+}|B|^{+}$ of a matrix B contains the same terms as the corresponding determinant $\det B$ but with constant positive signs for each

product of matrix elements in place of the alternating positive and negative signs of the determinant, so, for example,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}^{+} = ad + bc.$$

It can be shown (see Minc, 1978) that for any nonnegative-definite Hermitian matrix B, $\operatorname{per} B \geq \det B$.

Equations (6.2.7) and (6.2.8), taken together with Proposition 6.2.I, show that the factorial moment densities for the boson process are given by

$$m_{[k]}(t_1, \ldots, t_k) = \mathrm{E}\big(\lambda(t_1) \cdots \lambda(t_k)\big) = \lambda^k \, C^{+} \begin{pmatrix} t_1, \ldots, t_k \\ t_1, \ldots, t_k \end{pmatrix}. \tag{6.2.9}$$

This result paves the way for a discussion that exactly parallels the discussion of the fermion process of Example 5.4(c). In place of the expansion of the Fredholm determinant $d(\lambda)$ used there, we have here an analogous expansion of the function

$$d^{+}(\lambda) = 1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \int_A \cdots \int_A C^{+} \begin{pmatrix} u_1, \ldots, u_k \\ u_1, \ldots, u_k \end{pmatrix} \mathrm{d}u_1 \cdots \mathrm{d}u_k \,,$$

where as before the observation region A is a closed, bounded set in a general Euclidean space $\mathbb{R}^{\mathrm{d}}$. Corresponding to the expression (5.4.18) for the Fredholm minor is the expression

$$\begin{aligned} &\lambda^k R^{+}_{-\lambda} \begin{pmatrix} x_1, \ldots, x_k \\ y_1, \ldots, y_k \end{pmatrix} \\ &\quad = \frac{1}{d^{+}(\lambda)} \Bigg\{ \lambda^k C^{+} \begin{pmatrix} x_1, \ldots, x_k \\ y_1, \ldots, y_k \end{pmatrix} \\ &\qquad + \lambda^k \sum_{j=1}^{\infty} (-\lambda)^j \int_A \cdots \int_A C^{+} \begin{pmatrix} x_1, \ldots, x_k, \, u_1, \ldots, u_j \\ y_1, \ldots, y_k, \, u_1, \ldots, u_j \end{pmatrix} \mathrm{d}u_1 \cdots \mathrm{d}u_j \Bigg\}. \end{aligned} \tag{6.2.10}$$

This shows that the Janossy measures for the photon process have densities

$$j_k(x_1, \ldots, x_k) = \lambda^k d^{+}(\lambda) R^{+}_{-\lambda} \begin{pmatrix} x_1, \ldots, x_k \\ x_1, \ldots, x_k \end{pmatrix} \qquad (k = 1, 2, \ldots). \tag{6.2.11}$$

Macchi (1971a) established (6.2.11) directly by evaluating the expectation

$$j_k(x_1, \ldots, x_k) = \mathrm{E}\bigg(\lambda(x_1) \cdots \lambda(x_k) \exp\bigg(- \int_A \lambda(u) \, \mathrm{d}u \bigg) \bigg)$$

[see also Grandell (1976) and Exercises 6.2.5–6 for further discussion]. □

EXAMPLE 6.2(c) *A pseudo-Cox process: the Gauss–Poisson process.* The Gauss–Poisson process will be introduced as a two-point cluster process in Example 6.3(d) in the next section. Here we wish only to point out that the p.g.fl. $G[h]$ in (6.3.30) for such a process, if the measures Q_1 and Q_2 there are absolutely continuous with respect to Lebesgue measure, equals

$$\exp\left(\int_{\mathcal{X}}[1-h(x)]\,m(x)\,\mathrm{d}x - \tfrac{1}{2}\int_{\mathcal{X}}\int_{\mathcal{X}}[1-h(x)]\,[1-h(y)]\,c(x,y)\,\mathrm{d}x\,\mathrm{d}y\right),$$

where, in the notation of (6.3.30), in which $Q_2(\cdot)$ is symmetric,

$$m(x)\,\mathrm{d}x = Q_1(\mathrm{d}x) + 2Q_2(\mathrm{d}x\times\mathcal{X}) \quad \text{and} \quad c(x,y)\,\mathrm{d}x\,\mathrm{d}y = 2Q_2(\mathrm{d}x\times\mathrm{d}y).$$

This expression is identical in form with the expression $L^*[1-h]$ for the Laplace functional of a Gaussian process, $\{X(t)\colon t\in\mathbb{R}\}$ say, with mean $m(t) = \mathrm{E}X(t)$ and covariance $c(t,u) = \mathrm{cov}(X(t),X(u))$, provided only that the function $c(t,u)$ is positive-definite. On the other hand, the process is not an example of the construction described in Definition 6.2.I because, a.s., a realization of a Gaussian process takes both positive and negative values, so the notion of a Poisson process with parameter measure with density equal to the realization of such a Gaussian process is void. Newman (1970) coined the name 'Gauss–Poisson' because of this formal property of the p.g.fl. □

This example also serves to illustrate that while the conditions of 6.2.II are sufficient for a functional $L^*[1-h]$ to represent the p.g.fl. of a point process, they are not necessary because the functional displayed at the outset of Example 6.2(c) is not the Laplace functional of a random measure.

Exercises and Complements to Section 6.2

6.2.1 Let $\{I_n\} = \{(a_n,b_n]\colon n=1,2,\ldots\}$ be a sequence of random intervals on $\mathbb{R}_+$ of lengths $X_n = b_n - a_n > 0$ a.s. and having gaps $Y_n = a_{n+1} - b_n > 0$ a.s., with $\{X_n\}$ i.i.d. exponential r.v.s, $\{Y_n\}$ i.i.d. r.v.s independent of $\{X_n\}$ and with finite mean, and $a_1 = 0$. Let a Cox process N on $\mathbb{R}_+$ be directed by a random measure ξ, which has density λ on the set $\bigcup_{n=1}^{\infty} I_n$ and zero elsewhere. Show that $N(\cdot) + \delta_0(\cdot)$ is a renewal process.
[The points of the set $\{a_n, b_n\colon n=1,2,\ldots\}$ are those of an alternating renewal process with exponential lifetimes for one of the underlying lifetime distributions. Kingman (1964) showed, effectively, that any stationary Cox process that is also a stationary renewal process must be directed by the stationary version of the random measure described.]

6.2.2 *Discrete boson process.* Let $C \equiv (c_{ij})$ be a (real or complex) covariance matrix. The discrete counterpart of Example 5.4(c) and its associated exercises is the mixed Poisson process obtained by taking $N(i)$ $(i=1,\ldots,K)$ to be Poisson with random parameter $\lambda|Z_i|^2$, where $Z = (Z_1,\ldots,Z_K)$ has the multivariate normal distribution $N(0,C)$. For $K=1$, this reduces to a geometric distribution with p.g.f. $P(1+\eta) = 1/(1-\lambda c_{11}^2\eta)$. For $K>1$, the multivariate p.g.f. has the form

$$P(1+\eta_1,\ldots,1+\eta_K) = \frac{1}{\det(I-\lambda D_\eta C)}\,, \qquad (6.2.12)$$

where $D_\eta = \mathrm{diag}(\eta_1,\ldots,\eta_K)$.

The factorial moment relations corresponding to (6.2.9) may be written down as follows. For any $k > 0$, let $r_1, \ldots, r_K$ be nonnegative integers such that $r_1 + \cdots + r_K = k$; here, r_j is to be interpreted as the number of repetitions of the index j in defining the factorial moment

$$m_{[k]}(i_1, \ldots, i_k) = \mathrm{E}(N(1)^{[r_1]} \cdots N(K)^{[r_K]}),$$

where the set $(i_1, \ldots, i_k)$ consists of the index j repeated r_j times ($j = 1, \ldots, K$). We then have

$$m_{[k]}(i_1, \ldots, i_k) = \lambda^k C^+ \begin{pmatrix} i_1, \ldots, i_k \\ i_1, \ldots, i_k \end{pmatrix}. \tag{6.2.13}$$

6.2.3 (*Continuation*). The relations (6.2.12) and (6.2.13) of Exercise 6.2.2 are together equivalent to the identity for the reciprocal of the characteristic polynomial

$$\frac{1}{\det(I - \lambda D_\eta C)} = 1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{\text{perm}} C^+ \begin{pmatrix} i_1, \ldots, i_k \\ i_1, \ldots, i_k \end{pmatrix} \eta_{i_1} \cdots \eta_{i_k},$$

where the inner summation extends over all distinct permutations of k indices from the set $i_1, \ldots, i_k$ allowing repetitions [this is related to the Master Theorem of MacMahon (1915, Sections 63–66); see also Vere-Jones (1984, 1997)].

6.2.4 (*Continuation*). Using (6.2.12), we have also

$$P(z_1, \ldots, z_K) = d^+(\lambda) \det(I - \lambda D_z R_{-\lambda}),$$

where $R_{-\lambda} = C(I + \lambda C)^{-1}$ and $d^+(\lambda) = \det(I + \lambda C)$. From this p.g.f., we obtain the multivariate probabilities in the form (using the notation of preceding exercises)

$$\begin{aligned} \pi_k(i_1, \ldots, i_k) &= \Pr\{N(j) = r_j \ (j = 1, \ldots, K)\} \\ &= \lambda^k d^+(\lambda) \cdot \frac{R^+_{-\lambda} \begin{pmatrix} i_1, \ldots, i_k \\ i_1, \ldots, i_k \end{pmatrix}}{r_1! \cdots r_k!}. \end{aligned}$$

6.2.5 (*Continuation*). Derive the results of Example 6.2(b) by a suitable passage to the limit of the last three exercises.
[An alternative route to these results uses the expansion of $Z(t)$ in an orthogonal series over A: see Macchi (1971a) and Grandell (1976).]

6.2.6 (*Continuation*). When $C(s,t) = \sigma^2 \mathrm{e}^{-\alpha|s-t|}$ in Example 6.2(b), show that with $\beta = \sqrt{\alpha(\alpha - 2\sigma^2)}$,

$$\Pr\{N(0,T] = 0\} = \mathrm{e}^{\alpha T}\big(\cosh \beta T + (\alpha + 2\sigma^2)\beta^{-1} \sinh \beta T\big)^{-1}.$$

6.3. Cluster Processes

Cluster processes form one of the most important and widely used models in point process studies, whether applied or theoretical. They are natural

models for the locations of objects in the plane or in three-dimensional space, in a remarkable range of contexts: for example, plants, molecules, protozoa, human settlements, stars, galaxies, and earthquake epicentres. Along the time axis, they have been used to model photoelectric emissions, volcano eruptions, arrivals and departures at queueing systems, nerve signals, faults in computer systems, and many other phenomena. The cluster mechanism is also a natural way to describe the locations of individuals from consecutive generations of a branching process, an application with unexpectedly rich mathematical structure as well as its obvious practical applications.

The intuitive motivation of such processes involves two components: the locations of clusters and the locations of elements within a cluster. The superposition of the latter constitutes the 'observed' process. To model the cluster elements, we specify a countable family of point processes $N(\cdot \mid y_i)$ indexed by the cluster centres $\{y_i\}$ (a 'cluster field' in [MKM]). To model the cluster locations, we suppose there is given a process N_c of cluster centres, often unobserved, whose generic realization consists of the points $\{y_i\} \subset \mathcal{Y}$. More often than not, we have $\mathcal{Y} = \mathcal{X}$; it is useful to preserve the notational distinction as a reminder of the structure of the process. The centres y_i act as the germs (= ancestors in the branching process context) for the clusters they generate; it is supposed in general that there are no special features attaching to the points of a given cluster that would allow them to be distinguished from the points in some other cluster. More formally, we have the following definition.

Definition 6.3.I. *N is a* cluster process *on the c.s.m.s. $\mathcal{X}$, with centre process N_c on the c.s.m.s. $\mathcal{Y}$ and component processes the measurable family of point processes $\{N(\cdot \mid y): y \in \mathcal{Y}\}$, when for every bounded $A \in \mathcal{B}_{\mathcal{X}}$,*

$$N(A) = \int_{\mathcal{Y}} N(A \mid y)\, N_c(\mathrm{d}y) = \sum\nolimits_{y_i \in N_c(\cdot)} N(A \mid y_i) < \infty \quad \text{a.s.} \tag{6.3.1}$$

The definition requires the superposition of the clusters to be almost surely boundedly finite. There is, however, no requirement in general that the individual clusters must themselves be a.s. finite [i.e. the condition $N(\mathcal{X} \mid y) < \infty$ a.s. is not necessary], although it is a natural constraint in many examples. A general cluster random measure can be introduced in the same way by allowing the component processes to be random measures (see Exercise 6.3.1).

For the remainder of this section, we require the component processes to be mutually independent. We shall then speak of the component processes as coming from an *independent measurable family* and thereby defining an *independent cluster process*. In this definition, it is to be understood that multiple independent copies of $N(\cdot \mid y)$ are taken when $N_c\{y\} > 1$. If $\mathcal{Y} = \mathcal{X}$ (i.e. the cluster centre process and the component processes are all defined on the same space $\mathcal{X}$ and $\mathcal{X}$ admits translations), then the further constraint that the translated components $N(A - y \mid y)$ are identically distributed may be added, thus producing a natural candidate for a stationary version of the process.

Conditions for the existence of the resultant point process are not so easily obtained as for the Cox process, even though the superposition of the cluster member processes involves only operations that are clearly measurable. The difficulty revolves around the finiteness requirement embodied in equation (6.3.1). The number of clusters that are potentially able to contribute points to a given bounded set soars as the dimension of the state space increases, imposing delicate constraints that have to be met by any proposed existence theorem. For independent cluster processes, the finiteness condition can be rephrased somewhat more formally as follows.

Lemma 6.3.II. *An independent cluster process exists if and only if, for any bounded set $A \in \mathcal{B}_{\mathcal{X}}$,*

$$\int_{\mathcal{Y}} p_A(y)\, N_c(\mathrm{d}y) = \sum\nolimits_{y_i \in N_c} p_A(y_i) < \infty \quad \Pi_c\text{-a.s.}, \tag{6.3.2}$$

where $p_A(y) = \Pr\{N(A \mid y) > 0\}$ for $y \in \mathcal{Y}$ and $A \in \mathcal{B}_{\mathcal{X}}$, and Π_c is the probability measure for the process of cluster centres.

Proof. The sum (6.3.2) is required to converge a.s. as part of the definition of a cluster process. The converse, for given N_c, is an application of the second Borel–Cantelli lemma to the sequence of events

$$E_i = \{\text{cluster } i \text{ contributes at least one point to the set } A\}. \qquad \square$$

The condition of Lemma 6.3.II can alternatively be rephrased in terms of generating functionals (see Exercise 6.3.2). When the components of the process are stationary (i.e. their cluster centre process is stationary and the distribution of the cluster members depends only on their positions relative to the cluster centre), a simple sufficient condition for the resultant cluster process to exist is that the mean cluster size be finite; even in the Poisson case, however, this condition is not necessary (see Exercise 6.3.5 for details).

The moments are easier to handle. Thus, taking expectations conditional on the cluster centres yields

$$\mathrm{E}[N(A) \mid N_c] = \sum\nolimits_{y_i \in N_c} M_1(A \mid y_i) = \int_{\mathcal{Y}} M_1(A \mid y)\, N_c(\mathrm{d}y),$$

where $M_1(\cdot \mid y)$ denotes the expectation measure of the cluster member process with centre at y, assuming this latter exists. From the assumption that the cluster member processes form a measurable family, it follows also that whenever $M_1(A \mid y)$ exists, it defines a measurable kernel (a measure in A for each y and a measurable function of y for each fixed Borel set $A \in \mathcal{B}_{\mathcal{X}}$). Then we can take expectations with respect to the cluster centre process to obtain

$$\mathrm{E}[N(A)] = \int_{\mathcal{Y}} M_1(A \mid y)\, M^c(\mathrm{d}y), \tag{6.3.3}$$

finite or infinite, where $M^c(\cdot) = \mathrm{E}[N_c(\cdot)]$ is the expectation measure for the process of cluster centres. From this representation, it is clear that the first-

moment measure of the resultant process exists if and only if the integral in (6.3.3) is finite for all bounded Borel sets A.

Similar representations hold for the higher-order moment measures. In the case of the second factorial moment measure, for example, we need to consider all possible ways in which two distinct points from the superposition of clusters could fall into the product set $A \times B$ ($A, B \in \mathcal{B}_{\mathcal{X}}$). Here there are two possibilities: either both points come from the same cluster or they come from distinct clusters. Incorporating both cases, supposing the cluster centre process is given, we obtain

$$\begin{aligned}\mathrm{E}[N^{[2]}(A \times B \mid N_c)] = &\int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, N_c(\mathrm{d}y) \\ &+ \int_{\mathcal{Y}^{(2)}} M_1(A \mid y_1) M_1(B \mid y_2)\, N_c^{[2]}(\mathrm{d}y_1 \times \mathrm{d}y_2),\end{aligned}$$

where the superscript in $N^{[2]}$ denotes the process of distinct pairs from N and in the second integral we have used the assumption of *independent* clusters. Taking expectations with respect to the cluster centre process, we obtain for the second factorial moment of the cluster process

$$\begin{aligned}M_{[2]}(A \times B) = &\int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, M^c(\mathrm{d}y) \\ &+ \int_{\mathcal{Y}^{(2)}} M_1(A \mid y_1) M_1(A \mid y_2)\, M_{[2]}^c(\mathrm{d}y_1 \times \mathrm{d}y_2). \qquad (6.3.4)\end{aligned}$$

Again, the second factorial moment measure of the cluster process exists if and only if the component measures exist and the integrals in (6.3.4) converge. Restated in terms of the factorial cumulant measure, equation (6.3.4) reads

$$\begin{aligned}C_{[2]}(A \times B) = &\int_{\mathcal{Y}^{(2)}} M(A \mid y_1)\, M(B \mid y_2)\, C_{[2]}^c(\mathrm{d}y_1 \times \mathrm{d}y_2) \\ &+ \int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, M^c(\mathrm{d}y). \qquad (6.3.5)\end{aligned}$$

Many of these relationships are derived most easily, if somewhat mechanically, from the portmanteau relation for the probability generating functionals, which takes the form, for $h \in \mathcal{V}(\mathcal{X})$ and exploiting the independent cluster assumptions,

$$\begin{aligned}G[h] \;=\; \mathrm{E}\big(G[h \mid N_c]\big) \;&=\; \mathrm{E}\Big[\exp\Big(-\int_{\mathcal{Y}} \big(-\log G_m[h \mid y]\big)\, N_c(\mathrm{d}y)\Big)\Big] \\ &=\; G_c\big[G_m[h \mid \cdot\,]\big], \qquad (6.3.6)\end{aligned}$$

where $G_m[h \mid y]$ for $h \in \mathcal{V}(\mathcal{X})$ is the p.g.fl. of $N(\cdot \mid y)$, and

$$G[h \mid N_c] = \prod_{y_i \in N_c} G_m[h \mid y_i] = \exp\Big[-\int_{\mathcal{Y}} \big(-\log G_m[h \mid y]\big)\, N_c(\mathrm{d}y)\Big] \qquad (6.3.7)$$

is the conditional p.g.fl. of N given N_c. The a.s. convergence of the infinite product in (6.3.7) is equivalent to the a.s. convergence of the sum in Lemma 6.3.II by Exercise 6.3.2. The measurable family requirements of the family of p.g.fl.s for the cluster centres follow from the initial assumptions for the process. Thus, the p.g.fl. representation is valid whenever the cluster process exists.

One class of cluster processes occurs so frequently in applications, and is so important in the theory, that it warrants special attention. In this class, (1°) the cluster centres are the points of a Poisson process, and (2°) the clusters are independent and finite with probability 1. Whenever condition (1°) holds, we speak of a *Poisson cluster process*. The basic existence and moment results for Poisson cluster processes are summarized in the proposition below.

Proposition 6.3.III. *Suppose that the cluster centre process is Poisson with parameter measure $\mu_c(\cdot)$ and that the cluster member processes form an independent measurable family. Then, using the notation above,*

(i) *a necessary and sufficient condition for the existence of the resultant process is the convergence for each bounded $A \in \mathcal{B}_{\mathcal{X}}$ of the integrals*

$$\int_{\mathcal{Y}} p_A(y)\, \mu_c(\mathrm{d}y); \tag{6.3.8}$$

(ii) *when the process exists, its p.g.fl. is given by the expression*

$$G[h] = \exp\left(-\int_{\mathcal{Y}} \big(1 - G_m[h \mid y]\big)\, \mu_c(\mathrm{d}y)\right); \tag{6.3.9}$$

(iii) *the resultant process has first and second factorial moment measures and second factorial cumulant measure given, respectively, for $A, B \in \mathcal{B}_{\mathcal{X}}$, by*

$$M_1(A) = M_{[1]}(A) = \int_{\mathcal{Y}} M_{[1]}(A \mid y)\, \mu_c(\mathrm{d}y), \tag{6.3.10}$$

$$M_{[2]}(A \times B) = \int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, \mu_c(\mathrm{d}y) + M_1(A) M_1(B), \tag{6.3.11}$$

$$C_{[2]}(A \times B) = \int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, \mu_c(\mathrm{d}y); \tag{6.3.12}$$

(iv) *when $\mathcal{X} = \mathbb{R}^d$, the distribution function F of the distance from the origin to the nearest point of the process is given by*

$$1 - F(r) = \exp\left(-\int_{\mathcal{Y}} p_{S_r(0)}(y)\, \mu_c(\mathrm{d}y)\right), \tag{6.3.13}$$

where $S_r(0)$ is the sphere in $\mathcal{X} = \mathbb{R}^d$ of radius r and centre at 0.

Proof. Since $\mathrm{E}[N_c(\mathrm{d}y)] = M^c(\mathrm{d}y) = \mu_c(\mathrm{d}y)$ for a Poisson cluster process, condition (6.3.8) implies the a.s. convergence of (6.3.2) and hence the existence of the process. If the process exists, then since for $\bar{h} \in \mathcal{V}(\mathcal{Y})$, $G_c[\bar{h}] =$

$\exp\big(-\int[1-\bar{h}(y)]\,\mu_c(\mathrm{d}y)\big)$, equation (6.3.9) is just the appropriate special form of (6.3.6) with $\bar{h}(y) = G_m[h \mid y]$ for $h \in \mathcal{V}(\mathcal{X})$ and so it holds. Putting $h(x) = 1 - I_A(x)$, the integral in (6.3.9) reduces to

$$1 - G_m[1 - I_A(\cdot) \mid y] = p_A(y)\,,$$

from which the necessity of (6.3.8) is obvious.

The moment relations are just restatements of equations (6.3.3–5) for the special case of the Poisson process, where $M^c(\mathrm{d}y) = \mu_c(\mathrm{d}y)$ and $C_{[2]}(\mathrm{d}y_1 \times \mathrm{d}y_2) \equiv 0$. The final equation (6.3.13) is a consequence of the fact that if R is the distance from the origin to the nearest point of the process, then $R > r$ if and only if the sphere $S_r(0)$ contains no point of the process, which yields (6.3.13) as the special case of (6.3.9) with $h(x) = 1 - I_{S_r(0)}(x)$. □

If $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and the process is stationary, and the factorial measures entering into equations (6.3.10–12) have densities, then the latter equations simplify further. In this case, the cluster centre process reduces to a Poisson process with constant intensity μ_c, say, and the first-moment density for the cluster member process can be written

$$m_1(x \mid y) = m_1(x - y \mid 0) \equiv \rho_1(x - y), \quad \text{say.}$$

Similarly, the second factorial moment and cumulant densities can be written

$$\begin{aligned} m_{[2]}(x_1, x_2 \mid y) &= m_{[2]}(x_1 - y, x_2 - y) \equiv \rho_{[2]}(x_1 - y, x_2 - y), \\ c_{[2]}(x_1, x_2 \mid y) &= c_{[2]}(x_1 - y, x_2 - y) \equiv \gamma_{[2]}(x_1 - y, x_2 - y). \end{aligned}$$

Substituting, we obtain simplified forms for the corresponding densities of the cluster process:

$$\begin{aligned} m &= \mu_c \int_{\mathcal{X}} \rho_1(u)\,\mathrm{d}u = \mu_c M_1(\mathcal{X} \mid 0) = \mu_c \mathrm{E}[N_m(\mathcal{X} \mid 0)], \\ \breve{m}_{[2]}(u) &= m_{[2]}(y, y+u) = \mu_c \int_{\mathcal{X}} \rho_{[2]}(w, u+w)\,\mathrm{d}w + m^2, \qquad (6.3.14) \\ \breve{c}_{[2]}(u) &= \mu_c \int_{\mathcal{X}} \rho_{[2]}(w, u+w)\,\mathrm{d}w. \end{aligned}$$

A more systematic treatment of such reduced densities $\breve{m}_{[2]}$ and $\breve{c}_{[2]}$ is given in Section 8.1.

The particularly simple form of these expressions means that it is often possible to obtain explicit expressions for the second moments of the counting process in such examples. Note also that since the cumulant density $\breve{c}_{[2]}(u)$ is everywhere nonnegative, the resultant process is generally *overdispersed* relative to a Poisson process with the same first-moment measure (i.e. it shows greater variance in the number of counts). The alternative terms in the first line of (6.3.14) illustrate the sufficient condition for the existence of the process mentioned earlier and in Exercise 6.3.5: if the mean cluster size $M_1(\mathcal{X} \mid 0)$

is finite, then the first-moment measure of the resultant process exists, and *a fortiori* the resultant process itself exists.

Other aspects of the process, such as interval properties, are generally less easy to obtain. Nevertheless, some partial results may be obtained in this direction via equation (6.3.13). Suppose that $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. Then, from (6.3.13) but using the half-interval $(0, t)$ in place of the 'sphere' $(-t, t)$, the survivor function $S(t)$ [see below (2.1.3)] for the length of the interval from 0 to the first point of the process in $\mathbb{R}_+$ is given by

$$S(t) = \exp\left(-\int_{\mathbb{R}} p(t \mid y)\, \mu_c(\mathrm{d}y)\right), \tag{6.3.15}$$

where $p(t \mid y) = p_{(0,t)}(y)$, a special case of the function $p_A(y)$ in (6.3.2). Taking logarithms of (6.3.15) and differentiating, we see that the hazard function $r(t)$ for this first interval is given by

$$r(t) = -\int_{\mathbb{R}} \frac{\partial p(t \mid y)}{\partial t}\, \mu_c(\mathrm{d}y)\,.$$

When the process is stationary, a further differentiation gives the hazard function $q(\cdot)$ of the distribution of the interval between two consecutive points of the process, as in Exercise 3.4.2.

In higher dimensions, a similar approach may be used for the nearest-neighbour distributions, although explicit expressions here seem harder to determine (see Chapter 15).

In all of Examples 6.3(a)–(e) below, the spaces $\mathcal{X}$ and $\mathcal{Y}$ of Definition 6.3.I are the same.

Example 6.3(a) *The Neyman–Scott process: centre-satellite process; process of i.i.d. clusters* (Neyman and Scott, 1958, 1972; Thompson, 1955; Warren, 1962, 1971). Suppose that the individual cluster members are independently and identically distributed; that is, we are dealing with i.i.d. clusters as in Section 5.1 [see also Examples 5.3(a) and 5.5(a)]. Write $F(\mathrm{d}x \mid y)$ for the probability distribution of the cluster members with cluster centre at y and $Q(z \mid y)$ for the p.g.f. of the total cluster size (assumed finite). Then, the cluster member p.g.fl. is given by (5.5.12), which in the notation above becomes

$$G_m[h \mid y] = Q\left(\int_{\mathcal{X}} h(x)\, F(\mathrm{d}x \mid y) \;\middle|\; y\right), \tag{6.3.16}$$

while the corresponding factorial measures take the form

$$M_{[k]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k \mid y) = \mu_{[k]}(y) \prod_{i=1}^{k} F(\mathrm{d}x_i \mid y), \tag{6.3.17}$$

where $\mu_{[k]}(y)$ is the kth factorial moment for the cluster size distribution when the cluster centre is at y. Note that if F is degenerate at y, we obtain the compound Poisson process discussed in Example 2.1.10(b) and again in the next section, while if every cluster has exactly one point [so $Q(z \mid y) = z$], we have random translations, first mentioned above at Exercise 2.3.4(b).

In many practical applications with $\mathcal{X} = \mathbb{R}^d$, the cluster centre process is stationary Poisson at rate μ_c, $Q(z \mid y)$ and $\mu_{[k]}(y)$ are independent of y, and $F(\mathrm{d}x \mid y)$ is a function of the vector distance $x - y$ alone and has density function $f(x \mid y) = \breve{f}(x - y) = (\mathrm{d}/\mathrm{d}x)\breve{F}(x - y)$. With these simplifying assumptions, the resultant p.g.fl. takes the compact form

$$G[h] = \exp\left\{\mu_c \int_{\mathbb{R}^d} \left[Q\Big(\int_{\mathbb{R}^d} h(y+x)\,\breve{F}(\mathrm{d}x)\Big) - 1\right] \mathrm{d}y\right\}, \tag{6.3.18}$$

while from the densities in (6.3.14), the mean rate and second factorial cumulant measures for the resultant process are given by $m = \mu_c\mu_{[1]}$ and

$$\breve{c}_{[2]}(u) = \mu_c\mu_{[2]} \int_{\mathbb{R}^d} \breve{f}(y+u)\,\breve{f}(y)\,\mathrm{d}y, \tag{6.3.19}$$

respectively. Also, for the survivor function $S(t)$ of the interval to the first point in the case $d = 1$, we obtain

$$-\log S(t) = \mu_c \int_{\mathbb{R}} \big[1 - Q\big(1 - \breve{F}(y+t) + \breve{F}(y)\big)\big]\,\mathrm{d}y \tag{6.3.20}$$

with a pleasing simplification when $\breve{F}(\cdot)$ is the exponential distribution (see Exercise 6.3.7). Exercise 6.3.10 sketches a two-dimensional extension. □

EXAMPLE 6.3(b) *Bartlett–Lewis model: random walk cluster process; Poisson branching process* (Bartlett, 1963; Lewis, 1964a, b). In this example, we take $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and suppose that the points in a cluster are the successive end points in a finite random walk, starting from and including the cluster centre. The special case where the random walk has unidirectional steps in $\mathbb{R}^1$ (i.e. forms a finite renewal process), was used as a road traffic model in Bartlett (1963) and studied in depth by Lewis (1964a) as a model for computer failures.

A closed-form expression for $G_m[h \mid y]$ does not appear to exist, although for the special case where both the step lengths and the number of steps are independent of the positions of the cluster centre, it can be represented in the form

$$\begin{aligned} h(y)\Big(q_0 + q_1 \int_{\mathcal{X}} h(y+x_1)\,F(\mathrm{d}x_1) \\ + q_2 \int_{\mathcal{X}^{(2)}} h(y+x_1)h(y+x_1+x_2)\,F(\mathrm{d}x_1)\,F(\mathrm{d}x_2) + \cdots\Big), \end{aligned} \tag{6.3.21}$$

where q_j is the probability that the walk terminates after j steps and F is the common step-length distribution.

Assuming also a constant intensity μ_c for the Poisson process of cluster centres, the mean density takes the form

$$m = \mu_c \sum_{j=0}^{\infty} (j+1)q_j = \mu_c(1 + m_{[1]}), \tag{6.3.22}$$

while the reduced form for the second factorial cumulant measure is given by

$$\check{C}_{[2]}(\mathrm{d}u) = \mu_c \sum_{j=1}^{\infty} q_j \sum_{k=1}^{j} (j-k+1)\big(F^{k*}(\mathrm{d}u) + F^{k*}(-\mathrm{d}u)\big). \tag{6.3.23}$$

Expressions for the nearest point and nearest-neighbour distance can be obtained at least for the case $\mathcal{X} = \mathbb{R}$ and unidirectional $F(\cdot)$. Under these conditions, the probability $p(t \mid y)$ that a cluster with centre at y has a point in the interval $(0, t)$ is given by

$$p(t \mid y) = \begin{cases} 0 & \text{for } y > t, \\ 1 & \text{for } 0 \le y \le t, \\ \displaystyle\sum_{i=0}^{\infty} r_{i+1} \int_0^{|y|} \big[F(|y| + t - x) - F(|y| - x)\big]\, \mathrm{d}F^{i*}(x) & \text{for } y < 0, \end{cases}$$

where $r_i = \sum_{j=1}^{\infty} q_j$. Substituting in (6.3.17) and simplifying, we obtain for the log survivor and hazard functions

$$-\log S(t) = \mu_c t + \mu_c m_{[1]} \int_0^t [1 - F(x)]\, \mathrm{d}x = mt - \mu_c m_{[1]} \int_0^t F(x)\, \mathrm{d}x, \tag{6.3.24a}$$

$$r(t) = \mu_c + \mu_c m_{[1]}\big(1 - F(t)\big), \tag{6.3.24b}$$

where $1 + m_{[1]} = m/\mu_c$ as in (6.3.22) (see also Exercise 6.3.9). □

The next model, the Hawkes process, figures widely in applications of point processes to seismology, neurophysiology, epidemiology, and reliability. It is also an important model from the theoretical point of view and will figure repeatedly in later sections of this book. One reason for its versatility and popularity is that it combines in the one model both a cluster process representation and a simple conditional intensity representation, which is moreover linear. It comes closest to fulfilling, for point processes, the kind of role that the autoregressive model plays for conventional time series. However, the class of processes that can be approximated by Hawkes processes is more restricted than the class of time series models that can be approximated by autoregressive models. In particular, its representation as a cluster process means that the Hawkes process can only be used in situations that are overdispersed relative to the Poisson model.

In introducing the model, Hawkes (1971a, b, 1972) stressed the linear representation aspect from which the term 'self-exciting' derives. Here we derive its cluster process representation, following Hawkes and Oakes (1974), mainly because this approach leads directly to extensions in higher dimensional spaces but also because it simplifies study of the model.

Example 6.3(c) *Hawkes process: self-exciting process; infectivity model* [see also Examples 6.4(c) (marked Hawkes process), 7.2(b) (conditional intensity representation), 8.2(e) (Bartlett spectrum), 8.5(d) (mutually exciting point

processes) and 8.3(c) (linear prediction formulae)]. The points $\{x_i\}$ of a Hawkes process are of two types: 'immigrants' without extant parents in the process, and 'offspring' that are produced by existing points. An evolutionary construction of the points is as follows. Immigrants $\{y_j\}$, say, arrive according to a Poisson process at constant rate μ_c, while the offspring arise as elements of a finite Poisson process that is associated with some point already constructed. Any point of the process, located at x', say, has the potential to produce further points whose locations are those of a (finite) Poisson process with intensity measure $\mu(A - x')$; we assume that $\mu(\cdot)$ has total mass $\nu \equiv \mu(\mathcal{X}) < 1$ and that all these finite Poisson processes are mutually independent and, given the point that generates them, identically distributed (modulo the shift as noted) and independent of the immigrant process as well. Consequently, each immigrant has the potential to produce descendants whose numbers in successive generations constitute a Galton–Watson branching process with Poisson offspring distribution whose mean is ν. Since $\nu < 1$, this branching process is subcritical and therefore of finite total size with mean $1/(1-\nu) < \infty$ if we include the initial immigrant member. Regard the totality of all progeny of a given immigrant point y_j as a cluster; then the totality of all such immigrant points and their clusters constitutes a Hawkes process.

An important task is to find conditions that ensure the existence of a stationary Hawkes process (i.e. of realizations of point sets $\{x_i\}$ on the whole space $\mathcal{X} = \mathbb{R}^d$ having the structure above and with distributions invariant under translation). Since the immigrant process is stationary, a sufficient condition, by Exercise 6.3.5, is that the mean cluster size be finite [or else, since the immigrant process is Poisson, Proposition 6.3.III(i) can be invoked].

The cluster centres may be regarded as 'infected immigrants' from outside the system and the clusters they generate as the process of new infections they produce. Then, $\mu(\mathrm{d}x)$ is a measure of the infectivity at the point x due to an infected individual at the origin.

The key characteristics of any cluster are the first- and second-moment measures for the total progeny. From Exercise 5.5.6, the first of these is given by

$$M_1(A \mid 0) = \delta_0(A) + \mu(A) + \mu^{2*}(A) + \cdots \qquad (\text{bounded } A \in \mathcal{B}_{\mathcal{X}}),$$

while the second satisfies the integral equation

$$\begin{aligned}\int_{\mathcal{X}} M_{[2]}(\mathrm{d}y, y + A \mid 0) \\ = \int_{\mathcal{X}} M_1(y + A \mid 0)\, M_1(\mathrm{d}y \mid 0) - \delta_0(A) + \int_{\mathcal{X}} M_{[2]}(\mathrm{d}u, u + A \mid 0) \int_{\mathcal{X}} \mu(\mathrm{d}v),\end{aligned}$$

so that

$$(1-\nu) \int_{\mathcal{X}} M_{[2]}(\mathrm{d}y, y+A \mid 0) = \int_{\mathcal{X}} M_1(y+A \mid 0)\, M_1(\mathrm{d}y \mid 0) - \delta_0(A). \quad (6.3.25)$$

From the general results (6.3.10–12), it now follows that the mean density of the resultant cluster process is given by

$$m = \lambda M_1(\mathcal{X} \mid 0) = \mu_c/(1-\nu), \tag{6.3.26}$$

while for its factorial covariance measure we have

$$\begin{aligned} \breve{C}_{[2]}(A) &= \mu_c \int_{\mathcal{X}} M_{[2]}(\mathrm{d}y, y + A \mid 0) \\ &= \frac{\mu_c}{1-\nu}\left[\int_{\mathcal{X}} M_1(y + A \mid 0)\, M_1(\mathrm{d}y \mid 0) - \delta_0(A)\right]. \end{aligned} \tag{6.3.27}$$

This corresponds to the reduced density

$$\breve{c}_{[2]}(x) = \frac{\mu_c}{1-\nu}\left[\int_{\mathcal{X}} m_1(y)\, m_1(x+y)\, \mathrm{d}y - \delta_0(x)\right]$$

when $M_1(A \mid 0)$ is absolutely continuous with density $m_1(x)$, say, apart from the δ-function at the origin. An important feature of these formulae is that they lead to simple Fourier transforms, and we exploit this fact later in illustrating the spectral theory in Example 8.2(e).

For a parametric example, with $\mathcal{X} = \mathbb{R}$ and $\mu(\cdot)$ with support in $\mathbb{R}_+$, suppose that for some $\alpha > 0$ and $0 < \nu < 1$

$$\mu(\mathrm{d}x) = \begin{cases} \nu\alpha \mathrm{e}^{-\alpha x}\mathrm{d}x & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then $M_1(\cdot)$ is absolutely continuous apart from an atom at the origin; for its density $m_1(\cdot)$, we find on $x \geq 0$ that

$$m_1(x) = \delta(x) + \nu\alpha \mathrm{e}^{-\alpha(1-\nu)x}.$$

It follows that $\breve{C}_{[2]}(\cdot)$ is absolutely continuous also, and by substituting in (6.3.26) and (6.3.27), we find that the covariance density of the stationary process is given by

$$\breve{c}_{[2]}(y) = \frac{\mu_c \alpha\nu(1 - \frac{1}{2}\nu)}{(1-\nu)^2} \mathrm{e}^{-\alpha(1-\nu)|y|}. \tag{6.3.28}$$

□

Example 6.3(d) *The Gauss–Poisson process: process of correlated pairs* (Bol'shakov, 1969; Newman, 1970; Milne and Westcott, 1972). This process has the curious distinction of being simultaneously a Neyman–Scott process, a Bartlett–Lewis process, and a pseudo-Cox process [Example 6.2(c)]. Its essential characteristic is that the clusters contain either one or two points (so it exists if and only if the cluster centre process exists). Let one point be taken as the cluster centre, let $F(\mathrm{d}x \mid y)$ denote the distribution of the second point relative to the first, and let $q_1(y)$, $q_2(y)$ be the probabilities of 1 and 2 points, respectively, when the centre is at y. Then, we may regard the process as a special case of the Example 6.3(b) with

$$G_m[h \mid y] = q_1(y)h(y) + q_2(y)h(y) \int_{\mathcal{Y}} h(x)\, F(\mathrm{d}x \mid y)$$

so that for the resultant process (and recall that $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ here),

$$\log G[h] = \int_{\mathcal{X}} \big(h(y) - 1\big) q_1(y)\, \mu(\mathrm{d}y) + \int_{\mathcal{X}} \int_{\mathcal{X}} \big(h(x)h(y) - 1\big) q_2(y)\, \mu(\mathrm{d}y)\, F(\mathrm{d}x \mid y). \tag{6.3.29}$$

This is not quite in standard form because the measure $q_2(y)\, \mu(\mathrm{d}y)\, F(\mathrm{d}x \mid y)$ is not symmetric in general. However, the value of the p.g.fl. is unaltered when we replace this measure by its symmetrized form $Q_2(\mathrm{d}x \times \mathrm{d}y)$, say, so without loss of generality we may write the p.g.fl. in the form

$$\log G[h] = \int_{\mathcal{X}} \big(h(x) - 1\big)\, Q_1(\mathrm{d}x) + \int_{\mathcal{X}^{(2)}} \big(h(x)h(y) - 1\big)\, Q_2(\mathrm{d}x \times \mathrm{d}y), \tag{6.3.30}$$

where Q_1 and Q_2 are boundedly finite and Q_2 is symmetric with boundedly finite marginals. If now we define $\widetilde{Q}_2 = 2Q_2$ and substitute in (6.3.30), we obtain the standard form in (6.3.32) below using Khinchin measures.

Conversely, given any two such measures Q_1 and Q_2, any expression of the form (6.3.30) represents the p.g.fl. of a process of correlated points because we can first define a measure μ by

$$\mu(A) = Q_1(A) + Q_2(A \times \mathcal{X}),$$

then appeal to the Radon–Nikodym theorem to assert the existence μ-a.e. of nonnegative functions $q_1(\cdot)$, $q_2(\cdot)$ with $q_1(x) + q_2(x) = 1$ satisfying, for all bounded $A \in \mathcal{B}_{\mathcal{X}}$,

$$Q_1(A) = \int_A q_1(x)\, \mu(\mathrm{d}x) \quad \text{and} \quad Q_2(A \times \mathcal{X}) = \int_A q_2(x)\, \mu(\mathrm{d}x),$$

and finally use Proposition A1.5.III concerning regular conditional probabilities to define a family of probability measures $\{F(\cdot \mid x) : x \in \mathcal{X}\}$ by

$$Q_2(A \times B) = \int_A F(B \mid x)\, Q_2(\mathrm{d}x \times \mathcal{X}) = \int_A F(B \mid x) q_2(x)\, \mu(\mathrm{d}x)$$

for all bounded A and all $B \in \mathcal{B}_{\mathcal{X}}$. □

This discussion characterizes the p.g.fl. of such two-point cluster processes, but Milne and Westcott (1972) give the following stronger result.

Proposition 6.3.IV. *For* (6.3.30) *to represent the p.g.fl. of a point process, it is necessary and sufficient that*

(i) *Q_1 and Q_2 be nonnegative and boundedly finite, and*
(ii) *Q_2 have boundedly finite marginals.*

Proof. The additional point to be proved is that (6.3.30) fails to be a p.g.fl. if either Q_1 or Q_2 is a signed measure with nontrivial negative part. Exercise 6.3.11 sketches details [see also Example 6.2(c) and Exercises 6.3.12–13]. □

Observe that for the process with p.g.fl. given by (6.3.30), the expectation and second cumulant measures exist and are given, respectively, by

$$M(\mathrm{d}x) = Q_1(\mathrm{d}x) + Q_2(\mathrm{d}x \times \mathcal{X}) + Q_2(\mathcal{X} \times \mathrm{d}x), \quad (6.3.31a)$$

$$C_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2) = Q_2(\mathrm{d}x_1 \times \mathrm{d}x_2) + Q_2(\mathrm{d}x_2 \times \mathrm{d}x_1), \quad (6.3.31b)$$

the representation holding whether or not Q_2 is given in its symmetric version.

It appears to be an open problem to determine conditions similar to those in Proposition 6.3.IV for an expansion such as (6.3.30) with just k terms ($k \geq 3$) to represent the log p.g.fl. of a point process [see Milne and Westcott (1993) for discussion].

Example 6.3(e) *A bivariate Poisson process* [see also Examples 7.3(a) (intensity functions and associated martingales), 7.4(e) (random-time transformation to unit-rate Poisson process) and 8.3(a) (spectral properties), and Exercise 8.3.7 (joint forward recurrence time d.f.)]. A bivariate process can be represented as a process on the product space $\mathcal{X} \times \{1, 2\}$, where indices (or marks) $1, 2$ represent the two component processes. The p.g.fl. expansions are most conveniently written out with the integrals over each component space taken separately. Consider, in particular, a Poisson cluster process on $\mathcal{X} \times \{1, 2\}$ in which the clusters may be of three possible types only: a single point in process 1, a single point in process 2, and a pair of points, one from each process. Arguments analogous to those in the preceding example show that the joint p.g.fl. can be written in the form

$$\log G[h_1, h_2] = \int_{\mathcal{X}} \big(h_1(x) - 1\big)\, Q_1(\mathrm{d}x) + \int_{\mathcal{X}} \big(h_2(x) - 1\big)\, Q_2(\mathrm{d}x) + \int_{\mathcal{X}^{(2)}} \big(h_1(x_1)h_2(x_2) - 1\big)\, Q_3(\mathrm{d}x_1 \times \mathrm{d}x_2),$$

where Q_1, Q_2 and Q_3 are boundedly finite and Q_3 has boundedly finite marginals. The marginal p.g.fl. for process 1 can be found by setting $h_2 = 1$ and is therefore a Poisson process with parameter measure

$$\mu_1(\mathrm{d}x) = Q_1(\mathrm{d}x) + Q_3(\mathrm{d}x \times \mathcal{X});$$

similarly, the process with mark 2 is also Poisson with parameter measure

$$\mu_2(\mathrm{d}x) = Q_2(\mathrm{d}x) + Q_3(\mathcal{X} \times \mathrm{d}x).$$

Finally, the superposition of the two processes is of Gauss–Poisson type, with

$$\widetilde{Q}_1(\mathrm{d}x) = Q_1(\mathrm{d}x) + Q_2(\mathrm{d}x)$$

and (taking the symmetric form)

$$\widetilde{Q}_2(\mathrm{d}x_1 \times \mathrm{d}x_2) = \tfrac{1}{2}[Q_3(\mathrm{d}x_1 \times \mathrm{d}x_2) + Q_3(\mathrm{d}x_2 \times \mathrm{d}x_1)].$$

Evidently, this is the most general example of a bivariate Poisson cluster process with Poisson marginals since clusters of any higher order would introduce higher-order clusters in the marginals and hence destroy the Poisson property.

The resulting fidi distributions are infinitely divisible bivariate Poisson distributions of the kind studied by Holgate (1964) and Milne (1974); see also Griffiths, Milne and Wood (1979). The particular bivariate distribution studied by Dwass and Teicher (1957) corresponds to the situation where the pairs must occur for both processes at the same location x; the resultant process is then not only infinitely divisible but also has complete independence. □

Example 6.3(e) appears in many guises—for example as the joint process of the input and output streams of the M/M/∞ queue. It is closely related to the Gauss–Poisson process, which is nothing other than the 'ground process' (see Section 6.4) of the bivariate example above. We shall use it repeatedly to illustrate the structure of multivariate processes—their moments, spectra, conditional intensities, and compensators. See in particular Example 7.3(a).

There are, of course, many examples of bivariate Poisson processes that are not infinitely divisible; one class may be obtained by mixing over the relative proportions of pairs and single points in the example above (see Exercise 6.3.12). A queueing example is given in Daley (1972a).

The previous examples illustrate the point that the same process can be represented in several equivalent ways as a Poisson cluster process: the Gauss–Poisson process, for example, can be represented either as a Neyman–Scott process or as a Bartlett–Lewis type process for appropriately chosen special cases of those models. This same example also points the way to an intrinsic characterization of Poisson cluster processes. In the next result, the measures $K_k(\cdot)$ are extended versions of the Khinchin measures defined for finite processes by (5.5.5).

Proposition 6.3.V. *The p.g.fl. of every Poisson cluster process with a.s. finite clusters can be uniquely represented in the form*

$$\log G[h] = \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathcal{X}^{(k)}} \big(h(x_1) \dots h(x_k) - 1\big)\, K_k(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k), \quad (6.3.32)$$

where the $\{K_k\}$ form a family of symmetric, boundedly finite measures on $\mathcal{B}(\mathcal{X}^{(k)})$ such that each $K_k(\cdot)$ has boundedly finite marginals $K_k(\cdot \times \mathcal{X}^{(k-1)})$, and the sum

$$\sum_{k=1}^{\infty} \frac{1}{k!} \sum_{i=1}^{k} \binom{k}{i} K_k\big(A^{(i)} \times (A^c)^{(k-i)}\big) \quad (6.3.33)$$

is finite for bounded $A \in \mathcal{B}_{\mathcal{X}}$.

Conversely, given any such family of measures $\{K_k : k \geq 1\}$, the p.g.fl. (6.3.32) represents the p.g.fl. of a Poisson cluster process.

Proof. Suppose there is given a Poisson cluster process with cluster centres defined on the space $\mathcal{Y}$ and having parameter measure $\mu_c(\cdot)$. Suppose also

that the clusters are a.s. finite, so that they can be represented in terms of a family of Janossy measures $J_k(\cdot \mid y)$ (see Section 5.3), conditioned by the location y of the cluster centre. Note that by definition these measures are symmetric. Consequently, we consider the quantities $K_k(\cdot)$ defined by setting

$$K_k(B) = \int_{\mathcal{Y}} J_k(B \mid y)\, \mu_c(\mathrm{d}y) \qquad \big(B \in \mathcal{B}(\mathcal{X}^{(k)})\big)$$

and check that they are in fact boundedly finite measures. From Proposition 6.3.III, we know that the integral $\int_{\mathcal{Y}} p_A(y)\, \mu(\mathrm{d}y)$ converges for each bounded set $A \in \mathcal{B}_{\mathcal{X}}$. Here, $p_A(y)$ is just the sum over $k \geq 1$ of the probabilities that the cluster has k members of which at least one falls into the set A, so that, referring to (5.3.10), $p_A(y)$ equals

$$\sum_{k=1}^{\infty} \frac{J_k(\mathcal{X}^{(k)} \mid y) - J_k((A^c)^{(k)} \mid y)}{k!} = \sum_{k=1}^{\infty} \sum_{i=1}^{k} \binom{k}{i} \frac{J_k\big(A^{(i)} \times (A^c)^{(k-i)} \mid y\big)}{k!}\,.$$

The finiteness of $K_k(B)$ follows when B is of the form $A^{(k)}$ for bounded A. Similarly, by taking the term in the sum with $i = 1$, we deduce the bounded finiteness of the marginals. Finally, (6.3.33) is just a restatement of the necessary and sufficient condition that (6.3.8) be finite.

We can then obtain the representation (6.3.32) from the standard representation of a Poisson cluster p.g.fl.

$$\log G[h] = \int_{\mathcal{Y}} (G[h \mid y] - 1)\, \mu_c(\mathrm{d}y) \qquad \big(h \in \mathcal{V}(\mathcal{X})\big)$$

by expressing $G[h \mid y]$ in terms of the associated Janossy measures as in equation (5.5.3) and rearranging the integrations. Note that the term with $k = 0$ drops out of the summation. Uniqueness follows from standard results concerning uniqueness of the expression of the p.g.fl. and its logarithm about the origin.

Now suppose conversely that a family of measures K_k satisfying the stated conditions is given. We wish to construct at least one Poisson cluster process that has the p.g.fl. representation (6.3.32). Take $\mathcal{X} = \mathcal{Y}$, and let the measure $\mu_0(\cdot)$ be defined over bounded $A \in \mathcal{B}_{\mathcal{X}}$ by

$$\mu_0(A) = \sum_{k=1}^{\infty} K_k(A \times \mathcal{X}^{(k-1)})/k! \tag{6.3.34}$$

as the parameter measure for the cluster centre process. Note that the finiteness condition (6.3.33) entails the finiteness of (6.3.34) because

$$\begin{aligned} \sum_{i=1}^{k} \binom{k}{i} K_k\big(A^{(i)} \times (A^c)^{(k-i)}\big) &= \sum_{i=1}^{k} \frac{k}{i} \binom{k-1}{i-1} K_k\big(A \times A^{(i-1)} \times (A^c)^{k-i}\big) \\ &\geq K_k(A \times \mathcal{X}^{(k-1)}). \end{aligned}$$

As in the Gauss–Poisson case, we can define μ_0-a.e. a probability distribution $\{q_k(y)\}$ on $k = 1, 2, \ldots$ as the Radon–Nikodym derivatives in

$$\int_A q_k(y)\,\mu_0(\mathrm{d}y) = \frac{K_k(A \times \mathcal{X}^{(k-1)})}{k!}\,,$$

these probabilities $\{q_k(y)\}$ determining the number of points k in a cluster with centre y. The cluster member structure can be defined by taking one point as the cluster centre and locating the positions of the others relative to it through the distribution $P_{k-1}(B \mid y)$ defined μ_0-a.e. over $B \in \mathcal{B}(\mathcal{X}^{(k-1)})$ by

$$\int_A P_{k-1}(B \mid y)\, K_k(\mathrm{d}y \times \mathcal{X}^{(k-1)}) = K_k(A \times B),$$

appealing again to the existence of regular conditional probabilities. We can now check that the process with these components has the p.g.fl. representation (6.3.32) and that the existence condition (6.3.33) is satisfied. □

Note that there are many other processes that could be constructed from the same ingredients. In particular (see below Theorem 2.2.II), we can introduce an arbitrary probability $\tilde{q}_0(y)$ of empty clusters with $0 \le \tilde{q}_0(y) < 1$ (all y) by redefining

$$\tilde{q}_k(y) = \big(1 - \tilde{q}_0(y)\big) q_k(y) \qquad (k = 1, 2, \ldots)$$

and setting

$$\tilde{\mu}_c(\mathrm{d}y) = \big(1 - \tilde{q}_0(y)\big)^{-1} \mu_c(\mathrm{d}y).$$

The p.g.fl. is unaltered by this transformation, and the resultant processes are equivalent; we record this formally.

Corollary 6.3.VI. *The probability of a zero cluster is not an estimable parameter in any Poisson cluster model.*

A similar range of possibilities exists for the way the cluster centre x is defined relative to the joint distributions $P_k(\cdot)$ of the points in the cluster. In the construction above, we have chosen to fix the centre at an arbitrary point of the cluster. The measures $J_k(\,\cdot \mid y)$ are then related to the $P_k(\,\cdot \mid y)$ by $J_1(A) = P_1(A)$ and, for $k \ge 2$, the symmetrization relations

$$J_k(A_1 \times A_2 \times \cdots \times A_k \mid y) = k^{-1} \sum_{\mathrm{sym}} \delta_y(A_1) P_{k-1}(A_2 \times \cdots \times A_k \mid y).$$

Alternatively, we might prefer to locate the cluster centre at the multivariate centre of mass of the distribution (assuming this to be defined) or else in some other manner. This can be done without altering the final form of the p.g.fl. If it is necessary to select one particular form of representation for the process, we shall choose that used in the proof above and refer to it as the *regular representation* of the given process. The proposition implies that there is a one-to-one correspondence between measures on $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ induced by Poisson cluster processes and the elements in their regular representations.

Exercises and Complements to Section 6.3

6.3.1 *LeCam's precipitation process.* Formulate a definition for a general cluster random measure ζ analogous to Definition 6.3.I by replacing $\{N(\cdot \mid y)\}$ by a measurable family of random measures $\{\xi(\cdot \mid y)\}$. When these components are independent and $L_\xi[f \mid y]$ denotes the Laplace functional of $\xi(\cdot \mid y)$ defined over $f \in \mathrm{BM}_+(\mathcal{X})$ [see around (6.1.8)], the Laplace functional L_ζ of ζ is related to $\{L_\xi[f \mid y]\}$ and the p.g.fl. G_c of the cluster centre process by

$$L_\zeta = G_c[L_\xi[f \mid \cdot\,]\,]$$

provided ζ is well defined. [This model is discussed in LeCam (1961), who was motivated by the problem of modelling precipitation.]

6.3.2 Show that an independent cluster process exists if and only if, for each $h \in \mathcal{V}(\mathcal{X})$, the infinite product $G[h \mid N_c] = \prod_i G_m[h \mid y_i]$ converges Π_c-a.s.

6.3.3 Frequently, it may be desired specifically to include the cluster centres with the points generated by the cluster member processes with p.g.fl. $G_m[h \mid y]$. Show that the modified process has p.g.fl. $G_c[h(\cdot)G_m[h \mid \cdot\,]\,]$.

6.3.4 *Moment measures for a cluster process.* For a cluster process, the r.v. $X_f \equiv \int_{\mathcal{X}} f(y)\, N(\mathrm{d}y)$ can be expressed as the sum $\sum_i Y_f(y_i)$, where the y_i are the cluster centres and $Y_f(y) = \int_{\mathcal{X}} f(x)\, N_m(\mathrm{d}x \mid y)$ is the potential contribution to X_f from a cluster member with centre at y. Assume that for $f \in \mathrm{BM}_+(\mathcal{X})$

$$M_{1,f}(y) \equiv \mathrm{E}[Y_f(y)] = \int_{\mathcal{X}} f(x)\, M_1(\mathrm{d}x \mid y) < \infty,$$
$$M_{2,f}(y) \equiv \mathrm{E}[Y_f^2(y)] = \int_{\mathcal{X}^{(2)}} f(x_1) f(x_2)\, M_2(\mathrm{d}x_1 \times \mathrm{d}x_2 \mid y) < \infty.$$

Use a conditioning argument to obtain the basic relations

$$\begin{aligned}
\mathrm{E}X_f &= \int_{\mathcal{Y}} \mathrm{E}[Y_f(y)]\, M^c(\mathrm{d}y) = \int_{\mathcal{Y}} M_{1,f}(y)\, M^c(\mathrm{d}y) \\
&= \int_{\mathcal{Y}} \int_{\mathcal{X}} f(x)\, M_1(\mathrm{d}x \mid y)\, M^c(\mathrm{d}y), \\
\mathrm{E}X_f^2 &= \int_{\mathcal{Y}} V_2(y)\, M^c(\mathrm{d}y) + \int_{\mathcal{Y}^{(2)}} M_{1,f}(y)\, M_{1,f}(z)\, M_2^c(\mathrm{d}y \times \mathrm{d}z), \\
\operatorname{var} X_f &= \int_{\mathcal{Y}} V_2(y)\, M^c(\mathrm{d}y) + \int_{\mathcal{Y}^{(2)}} M_{1,f}(y) M_{1,f}(z)\, C_2^c(\mathrm{d}y \times \mathrm{d}z),
\end{aligned}$$

where $V_2(y) = M_{2,f}(y) - (M_{1,f}(y))^2 = \operatorname{var} Y_f(y)$. Derive equations (6.3.3–5) by considering also $\operatorname{cov}(X_f, X_g)$ and setting $f(\cdot) = I_A(\cdot)$, $g(\cdot) = I_B(\cdot)$.
[*Hint*: Take care in passing from ordinary to factorial moments.]

6.3.5 (a) Show that a sufficient condition for the existence of a stationary cluster process is that the mean cluster size be finite.

(b) Show by counterexample that the condition is not necessary, even for a Poisson cluster process.

[*Hint*: For part (a), show first that in the stationary case,

$$M_1(A) = \mu_c \int_{\mathcal{X}} M_1(A \mid x)\,\mathrm{d}x = \mu_c \int_{\mathcal{X}} M_1(A - x \mid 0)\,\mathrm{d}x = m\ell(A),$$

and then observe that $p(A \mid x) \le M_1(A \mid x)$. For part (b), consider a compound Poisson process with infinite mean batch size.]

6.3.6 (a) Show that a stationary Poisson cluster process is simple if and only if each cluster member process is simple.

(b) When this condition is satisfied, show that the d.f. F corresponding to an interval between successive points of the process has coefficient of variation ≥ 1. [*Hint*: Show that $R(t) \equiv -\log S(t)$ in (6.3.8) is subadditive in $t > 0$ and hence that $S(t) \ge \exp(-R'(0+)t)$. Use Korolyuk's theorem to identify $1/R'(0+)$ as the first moment of F, and use a hazard function argument (see Exercise 3.4.2) to identify the second moment of F with $(2/R'(0+)) \int_0^\infty S(t)\,\mathrm{d}t$. Exercise 6.3.9(b) below gives a special case.]

6.3.7 For a Neyman–Scott Poisson cluster process as around (6.3.20) with $\mathcal{Y} = \mathcal{X} = \mathbb{R}$, suppose $F(x)$ has an exponential distribution. Use (6.3.20) to show (see Vere-Jones, 1970) that the hazard function below (6.3.15) for the distance from the origin to the nearest point of the process is given by

$$r(t) = \frac{\mu_c(1 - Q(\mathrm{e}^{-\lambda t}))}{1 - \mathrm{e}^{-\lambda t}}\,.$$

6.3.8 Consider a Neyman–Scott cluster process with cluster centres y_i the points of a Poisson process at rate μ_c and for each such point a Poisson-distributed random number n_i of points, with mean Y_i for an i.i.d. sequence of r.v.s $\{Y_i\}$, are located at $\{y_i + x_{ij} : j = 1, \dots, n_i\}$, where the x_{ij} are i.i.d. with probability density $g(\cdot)$. Show that such a process $\{y_i + x_{ij} : i = 1, \dots, n_i, \text{ all } i\}$ is identical with the shot-noise process of Example 6.2(a).

6.3.9 (a) Evaluate the first-moment measure of the interval $(0, t]$ for a cluster with centre y in a Bartlett–Lewis process as

$$M_c((0,t] \mid y) = \begin{cases} 0 & y > t, \\ 1 + \sum_{i=1}^\infty r_i F^{i*}(t - y) & 0 < y \le t, \\ \sum_{i=1}^\infty r_i [F^{i*}(t + |y|) - F^{i*}(|y|)] & y \le 0. \end{cases}$$

(b) Show that the hazard function for the interval distribution in the process corresponding to (6.3.24) is

$$r(t) = \mu_c + \mu_c m_{[1]}(1 - F(t)) - \frac{m_{[1]} f(t)}{1 + m_{[1]}(1 - F(t))}\,,$$

where $f(t)$ is the density corresponding to $F(t)$. Now verify Exercise 6.3.6(b): the interval distribution has coefficient of variation ≥ 1 (Lewis, 1964a).

6.3.10 Suppose the common d.f. in a Neyman–Scott type process in $\mathbb{R}^2$ is circular normal with density $f(x, y) = (2\pi)^{-1} \exp[-\frac{1}{2}(x^2 + y^2)]$. Show that the probability that a particular point of a given cluster falls in the circle of radius r and centre at the origin, when the cluster centre is at a distance ρ from the

origin, equals

$$P(r \mid \rho) \equiv \mathrm{e}^{-\rho^2/2} \int_0^t u\mathrm{e}^{-u^2/2} I_0(u\rho)\,\mathrm{d}u,$$

where I_0 is the modified Bessel function of zero order. Then the log survivor function of the distance from the origin to the nearest point of such a Neyman–Scott Poisson cluster process, with cluster p.g.f. $Q(z)$, is given by

$$-\log S(r) = 2\pi\mu_c \int_0^\infty [1 - Q(1 - P(r \mid \rho))]\,\rho\,\mathrm{d}\rho.$$

In particular, if the number in each cluster has a Poisson distribution with mean λ,

$$-\log S(r) = 2\pi\mu_c \int_0^\infty (1 - \mathrm{e}^{-\lambda P(r|\rho)})\,\rho\,\mathrm{d}\rho.$$

6.3.11 Show that $P(z) = \exp\{q_1(z-1) + q_2(z^2-1)\}$ is a univariate p.g.f. if and only if $q_1 \geq 0$, $q_2 \geq 0$, and hence complete the proof of Proposition 6.3.IV. [*Hint*: To be a p.g.f., $P(z)$ must have nonnegative coefficients as a power series in z, while by virtue of its representation, $P(z)$ is an entire function. Hence, show that $\log P(z)$ must be well defined and nondecreasing on the whole positive half-line $z > 0$, and deduce that both q_1 and $q_2 \geq 0$.]

6.3.12 Show that a point process N is Gauss–Poisson if and only if the first two Khinchin measures are nonnegative with boundedly finite marginals and all remaining Khinchin measures vanish. [This is a rephrasing of Proposition 6.3.IV and Examples 6.2(c) and 6.3(d).]

6.3.13 Show that the functional of (possibly signed) measures $\overline{Q}_1(\cdot)$ and $\overline{Q}_2(\cdot \times \cdot)$

$$\int_{\mathcal{X}} [h(x) - 1]\,\overline{Q}_1(\mathrm{d}x) + \tfrac{1}{2} \int_{\mathcal{X}^{(2)}} [h(x) - 1]\,[h(y) - 1]\,\overline{Q}_2(\mathrm{d}x \times \mathrm{d}y)$$

equals the logarithm of the p.g.fl. of some point process if and only if $\overline{Q}_1$ is nonnegative and the symmetrized version

$$\overline{Q}_2^s(A \times B) = \tfrac{1}{2}(\overline{Q}_2(A \times B) + \overline{Q}_2(B \times A))$$

is nonnegative and bounded as in $\overline{Q}_2^s(A \times B) \leq \min(\overline{Q}_1(A), \overline{Q}_1(B))$ for bounded A, $B \in \mathcal{B}_{\mathcal{X}}$. [*Hint*: Reduce the functional above to the form of (6.3.30) and appeal to Proposition 6.3.IV. See also Example 6.2(d).]

6.3.14 Proposition 6.3.V represents a Poisson cluster process with a.s. finite clusters. Realize a cluster of size k and choose one of its points, Y say, at random. Show that

$$\Pr\{Y \in A\} = \frac{K_k(A \times \mathcal{Y}^{(k-1)})}{K_k(\mathcal{Y}^{(k)})},$$

but

$$\Pr\left\{\begin{array}{l}\text{a cluster realization of}\\ \text{size } k \text{ has a point in } A\end{array}\right\} = \sum_{i=1}^{k} \binom{k}{i} \frac{K_k(A^{(i)} \times (A^c)^{(k-i)})}{K_k(\mathcal{Y}^{(k)})}.$$

6.3.15 The factorial cumulant measures $C_{[k]}$ of a Gauss–Poisson process vanish for $k = 3, 4, \ldots$. Show in general that for a Poisson cluster process with clusters of size not exceeding k_0, $C_{[k]}$ vanishes for $k > k_0$. [*Hint*: Use (6.3.32) and write $1 + h$ for h.]

6.4. Marked Point Processes

In many stochastic process models, a point process arises not as the primary object of study but as a component of a more complex model; often, the point process is the component that carries the information about the locations in time or space of objects that may themselves have a stochastic structure and stochastic dependency relations. From the point of view of point process theory, many such models can be subsumed under the heading of *marked point processes*. In this section, we provide an initial study of such processes, particularly those with links to the Cox and cluster processes described in the two preceding sections.

For any marked point process, the locations $\{x_i\}$ where the events occur constitute an important process in their own right (the x_i may denote times but could also be two- or three-dimensional, for example). We shall refer to this process as the *ground process* and accordingly denote it by N_{g}.

Definitions 6.4.I.

(a) *A* marked point process (MPP), *with locations in the c.s.m.s.* $\mathcal{X}$ *and marks in the c.s.m.s.* $\mathcal{K}$, *is a point process* $\{(x_i, \kappa_i)\}$ *on* $\mathcal{X} \times \mathcal{K}$ *with the additional property that the ground process* $N_{\mathrm{g}}(\cdot)$ *is itself a point process; i.e. for bounded* $A \in \mathcal{B}_{\mathcal{X}}$, $N_{\mathrm{g}}(A) = N(A \times \mathcal{K}) < \infty$.

(b) *A* multivariate (*or* multitype) point process *is a marked point process with mark space the finite set* $\{1, \ldots, m\}$ *for some finite integer* m.

If a marked point process N is regarded as a process on the product space $\mathcal{X} \times \mathcal{K}$, then the ground process N_{g} is the *marginal process of locations*. However, it is a consequence of Definition 6.4.I(a) that not all point processes on product spaces are marked point processes. For example, the bivariate Poisson process on $\mathbb{R}^2$ with parameter measure $\mu\,\mathrm{d}x\,\mathrm{d}y$ cannot be represented as an MPP on $\mathbb{R} \times \mathbb{R}$ because such a Poisson process has $N(A \times \mathbb{R}) = \infty$ a.s. for Borel sets A of positive Lebesgue measure. However, in the special case of a multivariate point process, the extra condition is redundant since the finiteness of the mark space immediately implies that each *component process* $N_i(\cdot) = N(\cdot \times \{i\})$ is boundedly finite and we can write

$$N_{\mathrm{g}}(\cdot) = N(\cdot \times \{1, \ldots, m\}) = \sum_{i=1}^{m} N_i(\cdot). \tag{6.4.1}$$

In general, an MPP can be regarded either as a point process in the product space $\mathcal{X} \times \mathcal{K}$ subject to the finiteness constraint on the ground process N_{g} as set out above, or as an ordinary (not necessarily simple) point process in $\mathcal{X}$, $\{x_i\}$ say, with an associated sequence of random variables $\{\kappa_i\}$ taking their values in $\mathcal{K}$. Either approach leads to the representation of the MPP as a set of pairs $\{(x_i, \kappa_i)\}$ in the product space. They are equivalent whenever it can be shown that the marks κ_i in an MPP are well-defined random variables, which is certainly the case when the ground process has finite intensity, but there are subtleties in general: see Section 8.3 and Chapter 9 for further discussion.

The class of MPPs is a great deal richer than might at first appear. This is due to the great variety of forms that can be taken by the marks and the variety of dependence relations that can exist between the marks themselves and their locations. When $\mathcal{X} = \mathbb{R}$, for example, many remarkable results can be obtained by taking the mark at an event x_i to represent some feature from the history of the process up to x_i. A careful study of such MPPs lies at the heart of the fundamental researches of Matthes, Mecke, and co-workers.

Extending the concepts of earlier chapters, we define for MPPs the following two classes of point processes.

Definition 6.4.II.
(a) *The MPP N is* simple *if the ground process N_{g} is simple.*
(b) *The MPP N on $\mathcal{X} = \mathbb{R}^d$ is* stationary (homogeneous) *if the probability structure of the process is invariant under shifts in $\mathcal{X}$.*

The structure of an MPP may be spelled out in a variety of ways. If the ground process N_{g} is not necessarily simple, it can be thought of as a cluster process in which the cluster centres x_i are the distinct locations in $\mathcal{X}$ and the cluster members are all pairs in $\mathcal{X} \times \mathcal{K}$ of the form (x_i, κ_{ij}), where the κ_{ij} are the marks of the points with common location x_i. Equally, however, the family κ_{ij} could be thought of as a single, compound mark in the space $\mathcal{K}^{\cup}$ defined as in (5.3.8). This last comment implies that *by suitably redefining the marks, any MPP on $\mathcal{X}$ can be represented as an MPP on $\mathcal{X}$ for which the ground process N_{g} is simple.* For many applications, though not for all, we may therefore assume that the MPPs we encounter are simple.

The next pair of definitions characterize two important types of independence relating to the mark structure of MPPs. Observe in part (b) that a crucial feature is the role of order in the location space: it reflects the evolutionary property that we associate with a time-like dimension.

Definition 6.4.III (Independent marks and unpredictable marks)**.** *Let the MPP $N = \{(x_i, \kappa_i)\}$ on $\mathcal{X} \times \mathcal{K}$ be given.*
(a) *N has* independent marks *if, given the ground process $N_{\mathrm{g}} = \{x_i\}$, the $\{\kappa_i\}$ are mutually independent random variables such that the distribution of κ_i depends only on the corresponding location x_i.*
(b) *For $\mathcal{X} = \mathbb{R}$, N has* unpredictable marks *if the distribution of the mark at x_i is independent of locations and marks $\{(x_j, \kappa_j)\}$ for which $x_j < x_i$.*

The most common case of an MPP with independent marks occurs when the κ_i are in fact i.i.d. Similarly, the most common case of a process with unpredictable marks occurs when the marks are conditionally i.i.d. given the past of the process (but the marks may influence the future of N_{g}).

The next proposition outlines the basic structure of processes with independent marks, introducing in particular the *mark kernel* $F(\cdot \mid \cdot)$ at a specified location. P.g.fl.s for MPPs are defined over the space $\mathcal{V}(\mathcal{X} \times \mathcal{K})$ of measurable functions $h(x, \kappa)$ that lie between 0 and 1 and for some bounded set A, $h(x, \kappa) = 1$ for all $\kappa \in \mathcal{K}$ and $x \notin A$.

Proposition 6.4.IV (Structure of MPP with independent marks)**.** *Let N be an MPP with independent marks.*

(a) *The probability structure of N is completely defined by the distribution of the ground process N_{g} and the* mark kernel $\{F(K \mid x)\colon K \in \mathcal{B}(\mathcal{K}),\ x \in \mathcal{X}\}$, *representing the conditional distribution of the mark, given the location x.*

(b) *The p.g.fl. for N takes the form*

$$G[h] = G_{\mathrm{g}}[h_F] \qquad (h \in \mathcal{V}(\mathcal{X} \times \mathcal{K})), \tag{6.4.2}$$

where G_{g} is the p.g.fl. of N_{g} and $h_F(x) = \int_{\mathcal{K}} h(x,\kappa)\, F(\mathrm{d}\kappa \mid x)$.

(c) *The moment measure M_k of order k for N exists if and only if the corresponding moment measure M_k^{g} exists for the ground process N_{g}, in which case*

$$\begin{aligned} &M_k(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k \times \mathrm{d}\kappa_1 \times \cdots \times \mathrm{d}\kappa_k) \\ &\qquad = M_k^{\mathrm{g}}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k) \prod_{i=1}^{k} F(\mathrm{d}\kappa_i \mid x_i). \end{aligned} \tag{6.4.3}$$

Similar representations hold for factorial and cumulant measures.

Proof. All the statements above are corollaries of the general results for conditional point processes outlined in Section 6.1. In the present case, we deduce statements for the process of pairs $\{(x_i, \kappa_i)\}$ from their distribution conditional on the process of locations $\{x_i\}$ using the conditional independence of the κ_i.

Because of the independence properties, it is easiest to approach the statements via the p.g.fl. Given the locations x_i, the p.g.fl. of the pairs (x_i, κ_i) takes the form

$$G[h(x,\kappa) \mid N_{\mathrm{g}}] = \prod_i \left[\int_{\mathcal{K}} h(x_i,\kappa)\, F(\mathrm{d}\kappa \mid x_i) \right] = \prod_i h_F(x_i). \tag{6.4.4}$$

Note that $h_F \in \mathcal{V}(\mathcal{X})$ when $h \in \mathcal{V}(\mathcal{X} \times \mathcal{K})$ because for some bounded set A, $h(x,\kappa) = 1$ for $x \notin A$ and all $\kappa \in \mathcal{K}$, and hence for such x, $h_F(x) = \int_{\mathcal{K}} h(x,\kappa)\, F(\mathrm{d}\kappa \mid x) = 1$. Provided then that N_{g} exists, the final product is well defined for $h \in \mathcal{V}(\mathcal{X} \times \mathcal{K})$ and defines a measurable function of N_{g}. We thus have a measurable family satisfying Lemma 6.1.III(b); taking expectations over the locations, we obtain (6.4.2). Since the p.g.fl. is well defined, so are the fidi distributions and hence the probability structure of the process.

To justify the expressions for the moment measures, consider an integral of the form $\int h(x_1, \ldots, x_k, \kappa_1, \ldots, \kappa_k)\, N(\mathrm{d}x_1 \times \mathrm{d}\kappa_1) \cdots N(\mathrm{d}x_k \times \mathrm{d}\kappa_k)$. Conditional on the locations $\{x_i\}$, its expectation can be written

$$\int_{\mathcal{K}} \cdots \int_{\mathcal{K}} h(x_1, \ldots, x_k, \kappa_1, \ldots, \kappa_k)\, F(\mathrm{d}\kappa_1 \mid x_1) \cdots F(\mathrm{d}\kappa_k \mid x_k). \tag{6.4.5}$$

Now taking expectations over the locations, assuming the moment measure to exist for the ground process, we obtain (6.4.3), finite or infinite according to whether the integrals converge. But convergence of the integrals for all appropriate h is the necessary and sufficient condition for the existence of the moment measures, so statement (c) follows. □

In many applications, $\mathcal{K} = \mathbb{R}_+$ and interest centres on the random measure defined by

$$\xi(A) = \int_{A\times\mathcal{K}} \kappa\, N(\mathrm{d}x \times \mathrm{d}\kappa) = \sum_{x_i \in A} \kappa_i\,. \tag{6.4.6}$$

Its properties when ξ has independent marks are summarized below. Observe that if $\kappa_i = \kappa$ a.s. for all i, then $\xi(A) = \kappa N_{\mathrm{g}}(A)$.

Proposition 6.4.V. *If $\mathcal{K} = \mathbb{R}_+$ and the MPP N has independent marks, ξ in* (6.4.6) *defines a purely atomic random measure on $\mathcal{X}$ with only finitely many atoms on any bounded set $A \in \mathcal{B}_{\mathcal{X}}$. It has Laplace functional*

$$L_\xi[h] = G_{\mathrm{g}}[\phi_h] \qquad (h \in \mathrm{BM}_+(\mathcal{X})), \tag{6.4.7}$$

where $\phi_h(x) = \int_{\mathcal{K}} \mathrm{e}^{-\kappa h(x)}\, F(\mathrm{d}\kappa \mid x)$ and G_{g} is as in (6.4.2).

The moment measure M_k^ξ of order k for ξ exists if

(i) *the moment measure M_k^{g} of order k exists for the ground process N_{g},*

(ii) *the kth moment of the mark distribution, $\mu_k(x) = \int_{\mathbb{R}_+} \kappa^k\, F(\mathrm{d}\kappa \mid x)$ exists M_1^{g}-a.e., and*

(iii) *the integrals defining M_k^ξ in terms of μ_r and M_s^{g} for $r, s = 1, \dots, k$, converge.*

When they exist, the first- and second-moment measures are given, for bounded $A, B \in \mathcal{B}_{\mathcal{X}}$, by

$$M_1^\xi(B) = \int_B \mu_1(x)\, M_1^{\mathrm{g}}(\mathrm{d}x)\,, \tag{6.4.8}$$

$$M_2^\xi(A\times B) = \int_{A\times B} \mu_1(x_1)\mu_1(x_2)\, M_{[2]}^{\mathrm{g}}(\mathrm{d}x_1 \times \mathrm{d}x_2) + \int_{A\cap B} \mu_2(x)\, M_1^{\mathrm{g}}(\mathrm{d}x)\,. \tag{6.4.9}$$

Proof. The statements follow from reasoning similar to that used in Proposition 6.4.IV. The integral in (6.4.6) is a.s. finite when A is bounded (since the sum is then over an a.s. finite number of terms) and is easily seen to have the additivity properties required of a random measure. Its Laplace functional and moment measures can again be found by first conditioning on the locations. Thus, $L_\xi(h \mid N_{\mathrm{g}})$ equals

$$\mathrm{E}\left[\exp\left(-\int_{\mathbb{R}_+} h(x)\,\xi(\mathrm{d}x)\right) \,\middle|\, N_{\mathrm{g}}\right] = \prod_i \left[\int_{\mathbb{R}_+} \mathrm{e}^{-\kappa h(x_i)}\, F(\mathrm{d}\kappa \mid x_i)\right].$$

Equation (6.4.7) follows on taking expectations over the locations. Note that when $h \in \mathrm{BM}_+(\mathcal{X})$, the Laplace–Stieltjes transform $\phi_h \in \mathcal{V}(\mathcal{X})$, as is required for a p.g.fl. Equation (6.4.8) is derived similarly.

To obtain (6.4.9), we have to condition on the location of pairs (x_i, x_j) defined by the product counting measure $N_{\mathrm{g}} \times N_{\mathrm{g}}$. Note the special attention given to the diagonal pairs (x_i, x_i): $M_2^{\xi}(A \times B)$ equals

$$\mathrm{E}\Big[\int_A \int_B \Big(\int_{\mathcal{K}} \int_{\mathcal{K}} \kappa_1\kappa_2\, F(\mathrm{d}\kappa_1 \mid x_1)\, F(\mathrm{d}\kappa_2 \mid x_2)\Big)\, N_{\mathrm{g}}(\mathrm{d}x_1)\, N_{\mathrm{g}}(\mathrm{d}x_2) + \int_{A\cap B} \Big(\int_{\mathcal{K}} \kappa^2\, F(\mathrm{d}\kappa \mid x)\Big)\, N_{\mathrm{g}}(\mathrm{d}x)\Big]$$
$$= \int_{A\times B} \mu_1(x_1)\mu_1(x_2)\, M^{\mathrm{g}}_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2) + \int_{A\cap B} \mu_2(x)\, M^{\mathrm{g}}_1(\mathrm{d}x)\,.$$

These expressions can be checked by expanding the functionals and transforms concerned (see Exercise 6.4.1 for the case $k = 3$). □

As for cluster processes, the results simplify if the process is stationary, and the relevant factorial moment densities exist. Stationarity implies that the mark kernel is independent of x, $F(\cdot \mid x) = F(\cdot)$ say, so that ϕ_h in (6.4.7) becomes $\phi_h(x) = \int_{\mathcal{K}} \mathrm{e}^{-\kappa h(x)}\, F(\mathrm{d}\kappa)$, the usual Laplace–Stieltjes transform of the distribution F evaluated at $h(x) \in \mathrm{BM}_+(\mathcal{X})$. Given the existence of the reduced densities $\breve{m}^{\mathrm{g}}_{[2]}(\cdot)$ and $\breve{c}^{\mathrm{g}}_{[2]}(\cdot)$, and writing $\mu_k = \int_{\mathcal{K}} \kappa^k\, F(\mathrm{d}\kappa)$, (6.4.8) and (6.4.9) lead to

$$m = \mu_1 m_{\mathrm{g}}\,, \tag{6.4.10}$$
$$\breve{m}_2(u) = (\mu_1)^2 \breve{m}^{\mathrm{g}}_{[2]}(u) + \delta(u)\mu_2 m_{\mathrm{g}}\,, \tag{6.4.11a}$$
$$\breve{c}_2(u) = (\mu_1)^2 \breve{c}^{\mathrm{g}}_{[2]}(u) + \delta(u)\mu_2 m_{\mathrm{g}}\,. \tag{6.4.11b}$$

The appearance of the δ-function in (6.4.11) is a reminder that the ξ process, as well as the process N_{g}, is purely atomic and therefore has a diagonal concentration (see Section 8.1 below). Equation (6.4.11b) leads to the well-known expression for the variance of a random sum of i.i.d. r.v.s,

$$\operatorname{var} \xi(A) = [\mathrm{E}(\kappa)]^2 \operatorname{var} N_{\mathrm{g}}(A) + \mathrm{E}[N_{\mathrm{g}}(A)] \operatorname{var} \kappa. \tag{6.4.12}$$

Extension of the discussion above to the mark space $\mathcal{K} = \mathbb{R}$ is possible but leads to signed measures and requires the use of characteristic functionals in place of Laplace functionals; see Exercise 6.4.2.

An important special case arises when the ground process N_{g} is Poisson. We call such a process a *compound Poisson process*. As such, it extends the compound Poisson process introduced in Section 2.2, where $\mathcal{K} = \mathbb{Z}_+$. For this (generalized) compound Poisson process, the marks often represent a weight associated with the point, such as a monetary value in financial applications, an energy or seismic moment in seismology, a weight or volume in forestry or

geological prospecting, and so on. In such cases, ξ measures the total value, energy, weight, volume, etc., accumulating within a certain time interval or region. We give some examples shortly but first present a simple, important structural property that foreshadows results for more general classes of MPPs.

Lemma 6.4.VI. *A compound Poisson process that has mark kernel $F(\cdot \mid \cdot)$, and for which the Poisson ground process N_{g} has intensity measure $\mu(\cdot)$, is equivalent to a Poisson process on the product space $\mathcal{X} \times \mathcal{K}$ with intensity measure $\Lambda(\mathrm{d}x \times \mathrm{d}\kappa) = \mu(\mathrm{d}x)\, F(\mathrm{d}\kappa \mid x)$.*

Proof. We examine the p.g.fl.s. Substituting in (6.4.2) for the p.g.fl. of the Poisson process for N_{g} and rearranging, we have, using notation from (6.4.2),

$$G[h] = \exp\left(\int [h_F(x) - 1]\, \mu(\mathrm{d}x)\right) = \exp\left(\int\int [h(x,\kappa) - 1]\, F(\mathrm{d}\kappa \mid x)\, \mu(\mathrm{d}x)\right),$$

where the last expression can be identified with the p.g.fl. of the Poisson process on the product space. □

Many classical stochastic models are rooted in the compound Poisson process. One famous example is as follows.

Example 6.4(a) *Lundberg's collective risk model* (Lundberg, 1903; Cramér, 1930). Suppose that claims W_i against an insurer are made at times t_i. Let $\xi(t)$ represent the accumulated claims $\sum_{i:0<t_i<t} W_i$ over the period $(0,t)$. If the distribution of a generic claim W is supposed constant (independent of t) and the claim times t_i follow a Poisson process with constant intensity μ, then the pairs $\{(t_i, W_i)\}$ form a compound Poisson process. Typically, in this context, the claim distribution is chosen from the gamma, Weibull, or Pareto families or various modifications of these chosen to fit the specific application. From equations (6.4.10) and (6.4.11), we obtain the elegant special forms

$$\begin{aligned} \mathrm{E}[\xi(t)] &= \mu \mathrm{E}(W)\, t, \\ \operatorname{var} \xi(t) &= \mu \mathrm{E}(W^2)\, t. \end{aligned}$$

The crucial simplification underlying this elegance arises from the location process being Poisson, for then the covariance density $c_{[2]}(\cdot)$ vanishes.

If the insurance company has initial capital U_0 and it is assumed that gross premium income comes in at a constant rate α, then the financial reserve of the company after time t, excluding running costs, depreciation, inflation, income from investment, and other external factors, is equal to $U(t) = U_0 + \alpha t - \xi(t)$. The *classical ruin problem* consists in determining whether, and if so when, $U(t)$ first becomes zero. If $\alpha \leq \mu \mathrm{E}(W)$, ruin is certain, but the time to ruin may still be of importance. If $\alpha > \mu \mathrm{E}(W)$, ruin may be avoided and interest centres around estimating the *probability of ruin*, say η. In both cases, important information may be derived from the observation that, if

$\tau_i = t_i - t_{i-1}$, then the random variables $Z_i = W_i - \alpha\tau_i$ are independent, so that the process

$$U_n - U_0 = \sum_{i=1}^{n} Z_i = \alpha t_n - \xi t_n$$

constitutes a random walk. In particular, this observation, coupled to a standard martingale argument, leads to the classical Cramér bound on the probability of ultimate ruin. The argument is outlined in Exercise 6.4.3 (or else, see e.g. Embrechts *et al.*, 1997, Section 1.1). □

Example 6.4(b) *Negative binomial processes.* The negative binomial distribution is a common choice for the count random variables $N(A)$ in applications to processes $N(\cdot)$ where a clustering alternative is preferred to the Poisson process. It is somewhat surprising that the only known examples of processes yielding the negative binomial form for the distributions of $N(A)$ are both extreme cases: a compound Poisson process that has the complete independence property and in which all the clusters are concentrated at single points, and a mixed Poisson process in which the individual realizations are indistinguishable from those of a Poisson process. The usefulness of the negative binomial distribution in practice stems more from its relative simplicity and tractability than its link to organic physical models, although it will of course be true that for long time intervals, when the time scale of clustering is short relative to the time scale of observation, the compound Poisson model may be an adequate approximation. We describe these two models; see also Grégoire (1984) and the review article of Diggle and Milne (1983).

(i) *Compound Poisson process leading to negative binomial distributions.* Suppose there is given a compound Poisson process with constant intensity μ and discrete mark distribution that is independent of the location x. If $N(A)$ is to have a negative binomial distribution, then we know from Example 5.2(a) that the cluster size distribution should have the logarithmic form

$$\pi_n(x) = (\rho^n/n)\log[1/(1-\rho)].$$

Taking this as the mark distribution, we find that the p.g.fl. for the resulting random measure ξ, which in this case is again a point process but nonorderly, now has the form

$$G[h] = \exp\left(\int_{\mathcal{X}} \frac{\log\big([1-\rho h(x)]/(1-\rho)\big)}{\log(1-\rho)}\,\mu(\mathrm{d}x)\right) \qquad (h \in \mathcal{V}(\mathcal{X})).$$

This corresponds to the multivariate p.g.f. for the fidi distributions on disjoint sets $A_1, \ldots, A_k$,

$$P_k(A_1, \ldots, A_k;\, z_1, \ldots, z_k) = \prod_{i=1}^{k} \left[\frac{1-\rho}{1-\rho z_i}\right]^{-\mu(A_i)/\log(1-\rho)},$$

representing one simple type of multivariate negative binomial distribution. The factorial cumulant measures can be obtained from the expansion

$$\begin{aligned}\log G[1+\eta] &= \int_{\mathcal{X}} \frac{\log[1-\rho\eta(x)/(1-\rho)]}{\log(1-\rho)}\,\mu(\mathrm{d}x) \\ &= -\frac{1}{\log(1-\rho)}\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{\rho}{1-\rho}\right)^k \int_{\mathcal{X}}[\eta(x)]^k\,\mu(\mathrm{d}x)\end{aligned}$$

so that $C_{[k]}(\cdot)$ for $k \geq 2$ is a singular measure with a concentration $c_{[k]}\mu(\cdot)$ on the diagonal $x_1 = \cdots = x_k$, where $c_{[k]}$ is the kth factorial moment of the logarithmic distribution, or, equivalently, $c_{[k]}/\log[1/(1-\rho)]$ is the kth factorial cumulant of the negative binomial distribution.

Recall the p.g.f. of the negative binomial distribution in Example 5.2(a) and the p.g.fl. for a local process on a bounded Borel set A as in Example 5.5(b). The p.g.fl. for the type (i) negative binomial process applied to Example 5.5(b) gives us (since the integral over A^c vanishes)

$$G_A[1-I_A+h^*] = \exp\left(\frac{1}{\log(1-\rho)}\int_A \log\left(\frac{1-\rho h}{1-\rho}\right)\mu(\mathrm{d}x)\right),$$

where $h^*(x) = h(x)I_A(x)$. Thus, the localized process is still a negative binomial process. The local Janossy measures can be found from the expansion

$$\log\left(\frac{1-\rho h}{1-\rho}\right) = -\log(1-\rho) + \sum_{n=1}^{\infty}\frac{\rho^n}{n}\,h^{(n)},$$

from which we deduce that $p_0(A) = \exp[-\mu(A)]$ and

$$\begin{aligned}J_1(\mathrm{d}x \mid A) &= \rho p_0(A)\,\mu(\mathrm{d}x), \\ J_2(\mathrm{d}x_1 \times \mathrm{d}x_2 \mid A) &= \rho^2 p_0(A)[\mu(\mathrm{d}x_1)\mu(\mathrm{d}x_2) + \delta(x_1,x_2)\mu(\mathrm{d}x_1)],\end{aligned}$$

where the two terms in J_2 represent contributions from two single-point clusters at x_1 and x_2 ($x_1 \neq x_2$) and a two-point cluster at $x_1 = x_2$. □

(ii) *Mixed Poisson process leading to negative binomial distributions.* Take the mixing distribution Π with Laplace–Stieltjes transform Π^* as in (6.1.16), now generalized to the nonstationary case, to have the gamma distribution $\Gamma(\alpha,\lambda)$ with Laplace–Stieltjes transform $(1+s/\lambda)^{-\alpha}$. Then

$$G[h] = \Pi^*\left(\int_{\mathcal{X}}[1-h(x)]\,\mu(\mathrm{d}x)\right) = \left(1+\frac{1}{\lambda}\int_{\mathcal{X}}[1-h(x)]\,\mu(\mathrm{d}x)\right)^{-\alpha},$$

so that the multivariate p.g.f. has the form

$$P_k(A_1,\ldots,A_k;\,z_1,\ldots,z_k) = \left(1+\frac{1}{\lambda}\sum_{i=1}^{k}(1-z_i)\mu(A_i)\right)^{-\alpha}.$$

The factorial cumulants can be obtained from the expansion

$$\log G[1+\eta] = -\alpha \log\left(1 - \frac{1}{\lambda}\int_{\mathcal{X}} \eta(x)\,\mu(\mathrm{d}x)\right) = \alpha \sum_{k=1}^{\infty} \frac{1}{k}\left(\int_{\mathcal{X}} \frac{\eta(x)}{\lambda}\,\mu(\mathrm{d}x)\right)^k,$$

so

$$C_{[k]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k) = \alpha\lambda^{-k}(k-1)!\,\mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_k),$$

where we can recognize the coefficient of the product measure on the right-hand side as the kth cumulant measure of the negative binomial distribution. Note that Example 5.2(a) corresponds to the case where the measure $\mu(\cdot)$ is totally finite, in which case $\mu(\mathcal{X})/\lambda$ here equals the parameter μ there. □

Most of the examples of point processes that we have considered in earlier sections can be adorned with marks in a way similar to the Poisson process in Examples 6.4(a) and (b) above. The choice of underlying model will depend on the context and anticipated dependence structure. The most interesting extensions appear when we drop the assumption of completely independent marks and consider ways in which either the marks can influence the future development of the process or the current state of the process can influence the distribution of marks, or both. Using the Hawkes process of Example 6.3(c) as below illustrates some of the many possible issues that can arise.

Example 6.4(c) *Marked Hawkes process.* Marked versions of the Hawkes process of Example 6.3(c) are best known from Hawkes (1971b, 1972), who considered the multivariate case in detail, with an application in Hawkes and Adamopoulos (1973), though Kerstan (1964) considered them at length. We consider here the case of unpredictable marks; for a more general multivariate extension, see Example 8.3(c). Both extensions have important applications in seismology [see also Example 6.4(d) below], epidemiology, neurophysiology, and teletraffic (see e.g. Brémaud and Massoulié, 1996).

In extending the Hawkes process of Example 6.3(c) to an MPP $\{(x_i, \kappa_i)\}$, we interpret the marks κ_i as the 'type' of an individual in a multitype branching process. Recall that, in the branching process interpretation, points in a Hawkes process are either 'immigrants' without parents or 'offspring' of another point in the process. This (multitype) model now incorporates the following assumptions:

(i) immigrants arrive according to a compound Poisson process $N(\mathrm{d}y \times \mathrm{d}\kappa)$ with constant rate μ_c and fixed mark distribution $F(\mathrm{d}\kappa)$;

(ii) each individual in the process, whether an immigrant or not, has the potential to act as an ancestor and thereby yield first-generation offspring according to an ordinary Poisson process with intensity measure $\mu(\mathrm{d}u \mid \kappa) = \psi(\kappa)\,\mu(\mathrm{d}u)$ that depends only on the mark κ of the ancestor event and the distance u of the offspring from the ancestor; and

(iii) the marks of the offspring form an i.i.d. sequence with the same d.f. F as the immigrants.

The factor $\psi(\kappa)$ determines the relative average sizes of families with different marks, while the measure $\mu(\cdot)$ determines how the family members are spread out along the time axis. For a stable process, $\mu(\mathcal{X})$ must be finite, and for the sake of definiteness, we assume that $\mu(\mathcal{X}) = 1$ so that $\psi(\kappa)$ becomes the expected number of direct offspring with mark κ.

In principle, the analysis of such a process requires the general theory of multiple type branching processes with a continuous range of types. However, the assumption of i.i.d. marks (i.e. offspring types) greatly simplifies the analysis. Indeed, the assumptions above imply that the ground process N_g for this marked point process can be described as an ordinary Hawkes process with immigration rate μ_c and infectivity measure

$$\mu_g(\mathrm{d}u) = \rho\,\mu(\mathrm{d}u), \qquad \text{where } \rho = \mathrm{E}[\psi(\kappa)] = \int_{\mathcal{K}} \psi(\kappa)\,F(\mathrm{d}\kappa) < \infty.$$

If then $\rho < 1$, the total number of progeny is a.s. finite with finite mean $1/(1-\rho)$ so that the ground process is well defined and has a stationary version (see Exercise 6.3.5). Since the overall process may itself be regarded as a Poisson cluster process taking its values in $\mathcal{X} \times \mathcal{K}$, a second application of Exercise 6.3.5 implies that the overall process has a well-defined stationary version. We state this formally for reference.

Proposition 6.4.VII. *Using the notation above, sufficient conditions for the existence of a stationary version of the marked Hawkes process with unpredictable marks are*

(i) *the intensity measure $\mu(\cdot)$ is totally finite (and then taken to be a probability measure); and*
(ii) $\rho = \mathrm{E}[\psi(\kappa)] < 1$.

First- and second-order properties of the process can be obtained by combining results for branching processes with results for cluster processes and are given in Chapter 8. The p.g.fl. is difficult to obtain explicitly; one approach is suggested in Exercise 6.4.4.

Many variations and extensions of this model are possible. Example 7.3(b) will show that the conditional intensity for this process has a very simple and powerful linear form, which lends itself to various types of generalization. The mark can be expanded to include a spatial as well as a size component, as for the spatial ETAS model described below. The assumption of unpredictable marks can also be weakened in several ways, for example by allowing the distributions of the marks of the offspring to depend on either the mark of the ancestor or the offspring's distance from the ancestor, or both. See Example 8.3(e) for a somewhat simpler model illustrating such dependence.

If the branching structure is critical rather than subcritical (i.e. $\rho = 1$), further types of behaviour can occur. For example, if the infectivity function is sufficiently long-tailed, Brémaud and Massoulié (2001) provides examples of stationary Hawkes processes *without* immigration (i.e. of a Hawkes process

whose clusters overlap at such large distances that the process maintains a stationary regime). Further details are given in Chapter 10. □

Example 6.4(d) *Ordinary and spatial ETAS models.* Ogata (1988) introduced the ETAS (Epidemic Type After-Shock) model to describe earthquake occurrence, following earlier applications of the Hawkes model to this context by Hawkes and Adamopoulos (1973) and Vere-Jones and Ozaki (1982). It corresponds to the special case of the marked Hawkes process where $\mathcal{X} = \mathcal{K} = \mathbb{R}$, the x_i are interpreted as the occurrence times of the earthquakes and the κ_i as their magnitudes, and the following specific choices are made:

$$\begin{aligned}\psi(\kappa) &= A\mathrm{e}^{\alpha(\kappa-\kappa_0)} I_{\{\kappa>\kappa_0\}}(\kappa),\\ \mu(\mathrm{d}u) &= \frac{K}{(c+u)^{1+p}}\, I_{\{u>0\}}(u)\,\mathrm{d}u,\\ F(\mathrm{d}\kappa) &= \beta\mathrm{e}^{-\beta(\kappa-\kappa_0)} I_{\{\kappa>\kappa_0\}}(\kappa)\,\mathrm{d}\kappa.\end{aligned}$$

These choices are dictated largely by seismological considerations: thus, the mark distribution cited above corresponds to the Gutenberg–Richter frequency–magnitude law, while the power-law form for μ follows the empirical *Omori Law* for aftershock sequences. The free parameters are β, α, c, A and p. $K = p\,c^p$ is a normalizing constant chosen to ensure $\int_0^\infty \mu(\mathrm{d}u) = 1$.

In this case, sufficient conditions for a stationary process are that

$$p > 0, \qquad \beta > \alpha, \qquad \text{and} \qquad \rho = A\beta/(\beta-\alpha) < 1.$$

The last condition in particular is physically somewhat unrealistic since it is well known that the frequency–magnitude distribution cannot retain the pure exponential form indefinitely, but must drop to zero much more quickly for very large magnitudes.

An important extension involves adding locations to the description of the offspring so that the branching structure evolves in both space and time. Then, one obvious way of extending the model is to have the ground process include both space and time coordinates, retaining the same mark space $\mathcal{K}$.

From the computational point of view, however, and especially for the conditional intensity and likelihood analyses to be described in Chapter 7, there are advantages in keeping the ground process to the set of time points and regarding the spatial coordinates as additional dimensions of the mark. The weight (magnitude) component of the mark retains its unpredictable character (so the weights are i.i.d. given the past), but we allow the spatial component of the mark to be affected by the spatial location of its ancestor.

No matter which of these descriptions we adopt, the cluster structure evolves over both space and time, offspring events occurring at various distances away from the initial ancestor, just as they follow it in time. When the branching structure is spatially homogeneous, the infectivity measure $\mu(\mathrm{d}t \times \mathrm{d}x)$ depends both on the time delay $u = t - t_0$ and the displacement $y = x - x_0$ from the time and location of the ancestor (t_0, x_0).

Various branching mechanisms of this type have been proposed in the literature [see e.g. Ogata (1998) for a review]. Thus, Vere-Jones and Musmeci (1992) suggests a space–time diffusion with infectivity density

$$\mu(\mathrm{d}u \times \mathrm{d}y \times \mathrm{d}z) = \frac{\beta \mathrm{e}^{-\beta u}}{2\pi u \sigma_y \sigma_z} \exp\left[-\frac{1}{2u}\left(\frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2}\right)\right] \mathrm{d}u\,\mathrm{d}y\,\mathrm{d}z,$$

whereas Ogata's space–time ETAS model uses a simpler product form for the space and time terms. Many choices are possible for the components of the model without affecting the underlying cluster character. In some applications, the assumption of spatial homogeneity may not be appropriate, so the infectivity or mark distribution may depend on the absolute location of the offspring as well as its separation from the ancestor.

In all of this wide diversity of models, the basic sufficient condition for the existence of a stationary version of the model, essentially the subcriticality of the offspring branching process, is affected only insofar as the integral of the infectivity measure needs to be extended over space as well as time. □

We conclude this section with a preliminary foray into the fascinating and also practically important realm of *stochastic geometry*. Marked point processes play an important role here as models for finite or denumerable families of random geometrical objects. The objects may be of many kinds: triplets or quadruplets of points (then, the process would be a special case of a cluster process), circles, line segments, triangles, spheres, and so on.

Definition 6.4.VIII (Particle process)**.** *A* particle process *is a point process with state space $\Sigma_{\mathcal{X}}$ equal to the class of nonempty compact sets in $\mathcal{X}$.*

Thus, a typical realization of a particle process is a sequence, ordered in some way, of compact sets $\{K_1, K_2, \ldots\}$ from the c.s.m.s. $\mathcal{X}$. An underlying difficulty with such a definition is that of finding a convenient metric for the space $\Sigma_{\mathcal{X}}$. One possibility is the *Hausdorff metric* defined by

$$\rho(K_1, K_2) = \inf\{\epsilon \colon K_1 \subseteq K_2^{\epsilon} \text{ and } K_2 \subseteq K_1^{\epsilon}\},$$

where K^{ϵ} is the *halo set* $\bigcup_{x \in K} S_{\epsilon}(x)$ (see Appendix A2.2); for further references and discussion, see Stoyan *et al.* (1995), Stoyan and Stoyan (1994), and Molchanov (1997), amongst others. In special cases, when the elements are more specific geometrical objects such as spheres or line segments, this difficulty does not arise, as there are many suitable metrics at hand. Very often, interest centres on the *union set* or *coverage process*

$$\Xi = \bigcup S_i$$

(see Hall, 1988), which is then an example of a *random closed set* in $\mathcal{X}$.

Now let us suppose that $\mathcal{X} = \mathbb{R}^d$ and that for each compact set $S \subset \mathcal{X}$ we can identify a unique *centre* $y(S)$, for example its centre of gravity. Then, we

may introduce an equivalence relation among the sets in $\Sigma_{\mathcal{X}}$ by defining two compact sets to belong to the same equivalence class if they differ only by a translation. The sets in $\Sigma^o \equiv \Sigma^o_{\mathcal{X}}$, the compact subsets of $\mathcal{X}$ with their centres at the origin, index the equivalence classes so that every set $S \in \Sigma_{\mathcal{X}}$ can be represented as the pair (y, S^o), where $y \in \mathcal{X}$ and $S^o \in \Sigma^o$, and $S = y + S^o$ (set addition). This opens the way to defining the particle process as an MPP $\{y_i, S_i\}$, where the $\{y_i\}$ form a point process in $\mathcal{X}$ and the marks $\{S_i\}$ take their values in Σ^o. Once again, there is the problem of identifying a convenient metric on Σ^o, but this point aside, we have represented the original particle process as an example of a so-called *germ–grain model* in which the $\{y_i\}$ are the *germs* and the $\{S_i\}$ are the *grains*. The next example illustrates one of the most straightforward and widely used models of this type.

Example 6.4(e) *Boolean model.* This is the compound Poisson analogue for germ–grain models. We suppose that the locations $\{y_i\}$ form a Poisson process in $\mathcal{X}$ and that the compact sets S_i^o are i.i.d. and independent of the location process; write $S_i = y_i + S_i^o$. Two derived processes suggest themselves for special attention. One is the random measure $\Upsilon(\cdot)$ formed by superposing the compact sets S_i. With the addition of random weights W_i, this gives the bounded set A the (random) mass

$$\Upsilon(A) = \sum_i W_i \, \ell(A \cap S_i) \qquad (A \in \mathcal{B}_{\mathcal{X}}), \tag{6.4.13}$$

where $\ell(\cdot)$ is the reference measure on $\mathcal{X}$ (e.g. Lebesgue measure, or counting measure on a lattice). The other is the localized measure of the union set Ξ described above, which gives the bounded set A the (random) mass

$$\Psi(A) = \ell(A \cap \Xi) \equiv \ell\big\{ \textstyle\bigcup_i (A \cap S_i) \big\}. \tag{6.4.14}$$

For example, (6.4.13) might represent the total mass of ejected material falling within the set A from a series of volcanic eruptions at different locations; then (6.4.14) would represent the area of A covered by the ejected material.

In both cases, the processes can be represented in terms of densities forming random processes (random fields) on $\mathcal{X}$. Thus, (6.4.13) and (6.4.14) have respective densities

$$v(x) = \sum_i W_i I_{S_i}(x) \tag{6.4.15}$$

and

$$\psi(x) = I_{\{\cup_i S_i\}}(x). \tag{6.4.16}$$

Many aspects of these and related processes are studied in the stochastic geometry literature such as Mathéron (1975), Stoyan *et al.* (1995) and Molchanov (1997). Here we restrict ourselves to a consideration of the mean and covariance functions of (6.4.15) and (6.4.16) under the more explicit assumptions that $\mathcal{X} = \mathbb{R}^2$, that the location process N_{g} of centres $\{y(S_i)\} = \{y_i\}$ is a simple Poisson process with constant intensity λ, and that each S_i is a

disk of random radius R_i and has weight W_i that may depend on R_i but that the pairs (R_i, W_i) are mutually independent and independent also of the centres $\{y_i\}$. Consistent with our earlier description, we thus have an MPP on $\mathbb{R}^2$, with mark space $\mathbb{R}_+ \times \mathbb{R}_+$, and hence a point process N on $\mathbb{R}^2 \times \mathbb{R}_+^2$.

The mean and covariance function for $v(x)$ can be found by first conditioning on the ground process N_{g} as in earlier examples. Thus, writing $v(x)$ as

$$v(x) = \int_{\mathbb{R}^2 \times \mathbb{R}_+^2} w I_{\{r \geq \|y-x\|\}}(r, y)\, N(\mathrm{d}y \times \mathrm{d}r \times \mathrm{d}w) \tag{6.4.17}$$

and taking expectations, the independence assumptions coupled with the stationarity of the Poisson process yield

$$\begin{aligned} \mathrm{E}[v(x)] &= \lambda\, \mathrm{E}\Big[W \int_{\mathbb{R}^2} I_{\{R \geq \|y\|\}}(R, y)\, \mathrm{d}y \Big] = \lambda\, \mathrm{E}\Big[W \int_0^R \int_0^{2\pi} r\, \mathrm{d}r\, \mathrm{d}\theta \Big] \\ &= \lambda\, \pi\, \mathrm{E}(W R^2)\,. \end{aligned}$$

The second moment $\mathrm{E}[v(x_1)v(x_2)]$ can be found similarly by first conditioning on the $\{y_i\}$. Terms involving both pairs of distinct locations and coincident locations (arising from the diagonal term in the second-moment measure of the location process) are involved. However, as for Poisson cluster processes, we find that the covariance $\mathrm{cov}[v(x_1), v(x_2)]$ depends only on the term involving coincident locations: it equals

$$\begin{aligned} &\mathrm{E}\Big[\int_{\mathbb{R}^2 \times \mathbb{R}_+ \times \mathbb{R}_+} w^2 I_{\{r \geq \|y-x_1\|, r \geq \|y-x_2\|\}}(r, y)\, N(\mathrm{d}y \times \mathrm{d}r \times \mathrm{d}w) \Big] \\ &\qquad = \lambda \mathrm{E}\left[W^2 \int_{\mathbb{R}^2} I_{\{R \geq \max(\|y-x_1\|, \|y-x_2\|)\}}(R, y)\, \mathrm{d}y \right] \\ &\qquad = 2\lambda \mathrm{E}\big[W^2 \big(R^2 \arccos(u/R) - u\sqrt{R^2 - u^2} \big) I_{\{R \geq u\}}(R) \big], \end{aligned}$$

where $u = \frac{1}{2}\|x_1 - x_2\|$. Note that the first moment is independent of x and the covariance is a function only of $\|x_1 - x_2\|$, as we should expect from the stationary, isotropic character of the generating process. Note also that if the radius R is fixed, the covariance vanishes for $\|x_1 - x_2\| > 2R$.

The resemblance of these formulae to those for Poisson cluster processes is hardly coincidental. From a more general point of view, the process is a special case of LeCam's precipitation model in Exercise 6.3.1, where the Poisson cluster structure is generalized to cluster random measures. Some details and extensions are indicated in Exercise 6.4.6.

The corresponding formulae for the union process present quite different and, in general, much harder problems since we lose the additive structure for the independent contributions to the sum process. The first moment $\mathrm{E}[\psi(x)]$ represents the *volume fraction* of space (in this case area) occupied

by the union set Ξ. It can be approached by the following argument, which is characteristic for properties of the Boolean model. First, note that

$$1 - \mathrm{E}[\psi(x)] = 1 - \Pr\{\Xi \ni x\} = \Pr\{\Xi \not\ni x\} = \mathrm{E}\Big[\prod_i [1 - I_{S_i}(x)]\Big].$$

Conditioning on the locations $\{y_i\}$ (i.e. on the ground process N_{g}), we can write

$$\Pr\{\Xi \not\ni x \mid N_{\mathrm{g}}\} = \prod_i \Pr\{R_i < \|x - y_i\|\} = \prod_i h(y_i; x)\,,$$

say, where $h(y;x) = \mathrm{E}[I_{[0,\|y-x\|)}(R)]$ and R has the common distribution of the i.i.d. radii R_i. Removing the conditioning, we have

$$1 - \mathrm{E}[\psi(x)] = \mathrm{E}\Big[\prod_i h(y_i; x)\Big] = G_{\mathrm{g}}[h(\cdot\,; x)] = \exp\Big(-\lambda \int_{\mathbb{R}^2} [1 - h(y; x)]\,\mathrm{d}y\Big).$$

Substituting for $h(y;x)$ and simplifying, we obtain for the mean density the constant

$$p^* \equiv \mathrm{E}[\psi(x)] = 1 - \mathrm{e}^{-\lambda\,\mathrm{E}(\pi R^2)}. \tag{6.4.18}$$

For the second product moment, using similar reasoning, we have

$$\begin{aligned} m_2(x_1, x_2) = \mathrm{E}[\psi(x_1)\psi(x_2)] &= \Pr\{\Xi \ni x_1,\, \Xi \ni x_2\} \\ &= \Pr\{\Xi \ni x_1\} + \Pr\{\Xi \ni x_2\} - [1 - \Pr\{\Xi \not\ni x_1 \text{ or } x_2\}] \\ &= \mathrm{E}[\psi(x_1)] + \mathrm{E}[\psi(x_2)] - [1 - \Pr\{\Xi \not\ni x_1,\, \Xi \not\ni x_2\}] \\ &= 2p^* - 1 + G_{\mathrm{g}}[h(\cdot\,;\, x_1, x_2)], \end{aligned}$$

say, where $h(y; x_1, x_2) = \mathrm{E}[I_{[0,\min(\|y-x_1\|,\|y-x_2\|)]}(R)]$. Substituting for the p.g.fl. of the Poisson ground process, putting $u = \frac{1}{2}\|x_1 - x_2\|$ and simplifying, we find that $m(x_1, x_2)$ equals

$$2p^* - 1 + \exp\left(-\lambda \mathrm{E}\Big[\pi R^2 (1 + I_{\{R<u\}}) + 2\Big(R^2 \arcsin\frac{u}{R} + u\sqrt{R^2 - u^2}\Big) I_{\{R \ge u\}}\Big]\right).$$

Exercise 6.4.10 sketches an extension to higher-order product moments. □

Exercises and Complements to Section 6.4

6.4.1 For the atomic random measure ξ with independent marks as in Proposition 6.4.V, show that the third-order moment measure $M_3^{\xi}(A_1 \times A_2 \times A_3)$ equals

$$\begin{aligned} &\int_{A_1 \times A_2 \times A_3} \mu_1(x_1)\mu_1(x_2)\mu_1(x_3)\, M_{[3]}^{\mathrm{g}}(\mathrm{d}x_1 \times \mathrm{d}x_2 \times \mathrm{d}x_3) \\ &\quad + \Big[\int_{A_1 \times A_{23}} + \int_{A_2 \times A_{31}} + \int_{A_3 \times A_{12}}\Big] \mu_1(x_1)\mu_2(x_2)\, M_{[2]}^{\mathrm{g}}(\mathrm{d}x_1 \times \mathrm{d}x_2) \\ &\quad + \int_{A_1 \cap A_2 \cap A_3} \mu_3(x_1)\, M_1^{\mathrm{g}}(\mathrm{d}x_1), \end{aligned}$$

where $A_{ij} = A_i \cap A_j$ for $i \neq j$.

[*Hint*: Each side is the coefficient of $\frac{1}{6}s^3$ in the respective expansions of (6.4.7) with argument $sh(\cdot)$, using (6.1.9) for the Laplace functional, (5.5.4) [with $\eta(x) = \phi_{sh(x)} - 1 = -sh(x)\mu_1(x) + \frac{1}{2}s^2[h(x)]^2\mu_2(x) - \frac{1}{6}s^3[h(x)]^3\mu_3(x) + \cdots$ and $\mu_r(x) = \int_{\mathcal{K}} \kappa^r F(\mathrm{d}\kappa \mid x)$, $r = 1, 2, 3$] for the p.g.fl., and ϕ_{sh} as in (6.4.7). The general case now follows by appealing to the symmetry (invariance under permutations of the axes) of the moment measures.]

6.4.2 Develop formulae, analogous to those of Proposition 6.4.V, for characteristic functionals of MPPs with marks in $\mathbb{R}$. Use these to extend the results of Proposition 6.4.V to the case where ξ may be a signed measure.

6.4.3 *Cramér bound on probability of ruin.* For the compound risk process, verify the following results [with notation as for Example 6.4(a)].

(i) The sequence $U_n - U_0$ forms a random walk with mean $\alpha/\mu - \mathrm{E}(W)$.

(ii) If ruin occurs, then it does so at the first time point t_n for which $U_n < 0$.

(iii) If $\alpha \le \mu\mathrm{E}(W)$, then ruin is certain, but if $\alpha > \mu\mathrm{E}(W)$, then there is positive probability that ruin will never occur.

(iv) In the latter case, if the Laplace–Stieltjes transform $\mathrm{E}(\mathrm{e}^{-sW})$ is an entire function of s, then there exists positive real s^* such that $\mathrm{E}(\mathrm{e}^{-s^*W}) = 1$.

(v) The sequence $\{\zeta_n\} = \{\exp(-s^*U_n)\}$ constitutes a martingale for which the time of ruin is a stopping time.

(vi) Let p_M denote the probability that ruin occurs before the accumulated reserves reach a large number M. Deduce from the martingale property that

$$p_M\mathrm{E}[\exp(s^*\Delta_0) \mid 0] + (1 - p_M)\mathrm{E}[\exp(-s^*\Delta_M) \mid M] = \exp(-s^*U_0),$$

where $-\Delta_0$ and Δ_M are the respective overshoots at 0 and M.

(vii) Hence, obtain the Cramér bound for the probability of ultimate ruin

$$p = \lim_{M\to\infty} p_M \le \exp(-s^*U_0)\,.$$

6.4.4 Find first and second factorial moment measures for the ground processes of the marked and space–time Hawkes processes described in Example 6.4(c). [*Hint*: Use the cluster process representation much as in Example 6.3(c).]

6.4.5 Study the Laplace functional and moment measures for the random measure ξ for a Hawkes process with unpredictable marks. [*Hint*: Use the cluster representation to get a general form for the p.g.fl. of the process as a process on $\mathcal{X} \times \mathcal{K}$. From it, develop equations for the first and second moments.] Are explicit results available?

6.4.6 Formulate the process $\Upsilon(A)$ in (6.4.13) as an example of a LeCam process (see Exercise 6.3.1). Show that in the special case considered in (6.4.17), when the random sets are spheres [= disks in $\mathbb{R}^2$] with random radii we can write

$$L_\xi[f \mid x] = \mathrm{E}\left[\exp\left(-W\int_{\mathbb{R}^2} f(y)\, I_{\{R\ge\|x-y\|\}}(y)\,\mathrm{d}y\right)\right].$$

Derive expressions for the mean and covariance functions of $v(x)$ as corollaries.

6.4.7 *Higher-order moments of the union set.* In the context of the union set Ξ of the Boolean model of Example 6.4(e), show that the kth product moment

$$\mathrm{E}[\psi(x_1)\cdots\psi(x_k)] = \Pr\{\Xi \ni x_j \ (j = 1, \ldots, x_k)\},$$

for k distinct points $x_1, \ldots, x_k$ in $\mathcal{X} = \mathbb{R}^2$, equals

$$1 + \sum_{r=1}^{k} (-1)^r \sum\nolimits_r{}' q(x_{j_1}, \ldots, x_{j_r}),$$

where $\sum'$ denotes the sum over all distinct r-tuplets of the set $\{x_1, \ldots, x_k\}$, $q(x_1, \ldots, x_r) = G_{\mathrm{g}}[h(\cdot\,; x_1, \ldots, x_r)]$, and the function $h(y\,; x_1, \ldots, x_r) = \Pr\{R < \min_{1\le j\le r}\{\|x_j - y\|\}\}$. [*Hint*: The relation arises from taking expectations in the expansion of products of indicator random variables

$$\begin{aligned} I\{\Xi \ni \text{ all } x_j\} &= \prod\nolimits_j I\{\Xi \ni x_j\} \;=\; \prod\nolimits_j (1 - I\{\Xi \not\ni x_j\}) \\ &= 1 + \sum\nolimits_{r=1}^{k} (-1)^r \sum\nolimits_r{}' \prod\nolimits_{\ell=1}^{r} I\{\Xi \not\ni x_{j_\ell}\} \end{aligned}$$

and

$$\prod\nolimits_{\ell=1}^{r} I\{\Xi \not\ni x_{j_\ell}\} = \prod\nolimits_{\ell=1}^{r} \prod\nolimits_i I\{S_i \not\ni x_{j_\ell}\} = \prod\nolimits_i \prod\nolimits_{\ell=1}^{r} I\{S_i \not\ni x_{j_\ell}\},$$

and the conditional expectation of the last product, given the locations $\{y_i\}$, equals $h(y_i; x_{j_1}, \ldots, x_{j_r})$, as indicated.]