Temporal point processes: the conditional intensity function

Jakob Gulddahl Rasmussen

January 24, 2011

Contents

Events	Marks
Earthquakes	Magnitudes Locations
Arrivals at a server	Service time
Accidents	Insurance claims Type of Injury

Table 1: Examples of events and marks.

1 Introduction

A temporal point pattern is basically a list of times of events. Many real phenomena produce data that can be represented as a temporal point pattern; the left column of Table 1 shows a few examples. Common to these examples is that we do not know how many events will occur, or at what times they will occur. Usually complex mechanisms are behind these seemingly random times, for example earthquakes cause new earthquakes in the form of aftershocks. An essential tool for dealing with these mechanisms, for example in predicting future events, is a stochastic process modelling the point patterns: a *temporal point process*. The term point is used since we may think of an event as being instant and thus can represent it as a point on the time line. For the same reason the words point and event will be used interchangeably throughout this note.

Often there is more information available associated with an event. This information is known as marks. Examples are given in the right column of Table 1. The marks may be of separate interest or may simply be included to make a more realistic model of the event times. For example, it is of practical relevance to know the position and magnitude of an earthquake, not just its time. At the same time, the magnitude of an earthquake also influences how many aftershocks there will be, so a model not including magnitudes as marks may not be reliable at modelling the event times either.

In this note, familiarity with the Poisson process on the line as well as basic probability theory and statistics is assumed. On the other hand, measure theory is not assumed; for a much more thorough treatment with all the measure theoretical details, see Daley and Vere-Jones (2003) and Daley and Vere-Jones (2008).

2 Evolutionary point processes

There are many ways of treating (marked) temporal point processes. In this note we will explore one approach based on the so-called conditional intensity function.

To understand what this is, we first have to understand the concept of evolutionarity.

2.1 Evolutionarity

Usually we think of time as having an *evolutionary character*: what happens now may depend on what happened in the past, but not on what is going to happen in the future. This order of time is also a natural starting point for defining practically useful temporal point processes. Roughly speaking, we can define a point process by specifying a stochastic model for the time of the next event given we know all the times of previous events. The term *evolutionary point process* is used for processes defined in this way.

The past in a point process is captured by the concept of the *history* of the process. If we consider the time t, then the history $\mathcal{H}_t$ is the list of times of events $(\ldots, t_1, t_2, \ldots, t_n)$ up to but not including time t. Note that theoretically the point process may extend infinitely far back in time, but it does not have to do this. Note also that we assume that we have a *simple point process*, i.e. a point process where no points coincide, such that the points can be strictly ordered in time.

2.2 Interevent times

When specifying a temporal point process we can use many different approaches. In this note, we will consider two, where we specify the distribution of

(I) the time lengths between subsequent events, or

(II) the number of events occurring in an arbitrary time-interval.

We will start with (I).

The lengths of the time intervals between subsequent events are known as *interevent times*. We can define a temporal point process by specifying the distributions of these. Let $f^*(t) = f(t|\mathcal{H}_t)$ be the conditional density function of the time of the next event t_{n+1} given the history of previous events $(\ldots, t_{n-1}, t_n)$. Here we use the notation $*$ from Daley and Vere-Jones (2003) to remind ourselves that this density is conditional on the past, rather than writing explicitly that a function depends on the history. Note that the density function $f^*(t)$ specifies the distribution of all interevent times, one by one, starting in the past, and thus the distribution of all events is given by the joint density

$$f(\ldots, t_1, t_2, \ldots) = \prod_i f(t_i|\ldots, t_{i-2}, t_{i-1}) = \prod_i f^*(t_i)$$

in the same manner as the joint density for a bivariate random variable factorises into $p(x, y) = p(x)p(y|x)$. Let us consider a simple example of a point process specified by specifying $f^*(t)$:

Example 2.1 (Renewal process and Wold process) The simplest process we can define by specifying the distribution of the interevent times is the renewal process. This process is defined by letting the interevent times be i.i.d. stochastic variables, i.e. $f^*(t_n) = f(t_n - t_{n-1})$ where f is a density function for a distribution on $(0, \infty)$. A special case of this is the homogeneous Poisson process with intensity λ, where f is the density of the exponential distribution with inverse mean λ. Figure 1 shows simulations of three different renewal processes. A generalisation of the renewal process is the Wold process where, rather than being independent, the interevent times may depend on the previous interevent time (i.e. the interevent times are a first-order Markov chain).

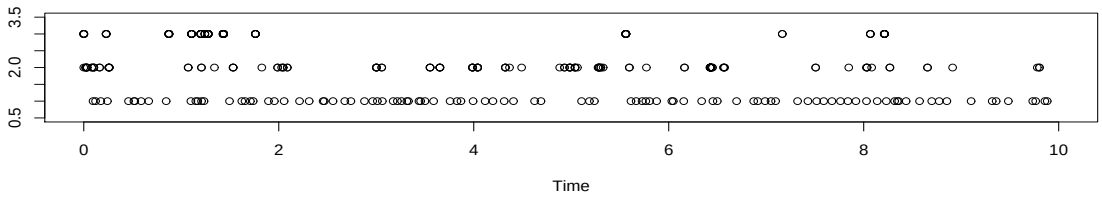

Figure 1: Three simulations of renewal processes with different interevent time distributions: Gamma(0.02,0.2) (upper), Gamma(0.1,1) (middle), Gamma(2,20) (lower). Note how the upper case is clustered and the lower case is regular compared to the middle case (which is a Poisson process). Also note that all the simulations have roughly 100 points for easy comparison (they are very densely packed together for the upper case).

The above examples show cases where the placement of t_n depends on t_{n-1} and, in the case of the Wold process, t_{n-2}. However, in general it may depend on the whole history, and it turns out that $f^*(t)$ is not the most convenient or intuitive way of specifying the general case.

2.3 Conditional intensity function

The conditional intensity function is a convenient and intuitive way of specifying how the present depends on the past in an evolutionary point process. Consider

the conditional density f^* and its corresponding cumulative distribution function F^*. Then the *conditional intensity function* (or hazard function) is defined by

$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)}. \tag{1}$$

The conditional intensity function can be interpreted heuristically in the following way: consider an infinitisemal interval around t, say $\mathrm{d}t$, then

$$\begin{aligned}
\lambda^*(t)\mathrm{d}t &= \frac{f^*(t)\mathrm{d}t}{1 - F^*(t)} \\
&= \frac{\mathbb{P}(\text{point in dt}|\mathcal{H}_t)}{\mathbb{P}(\text{point not before t}|\mathcal{H}_t)} \\
&= \frac{\mathbb{P}(\text{point in dt, point not before t}|\mathcal{H}_t)}{\mathbb{P}(\text{point not before t}|\mathcal{H}_t)} \\
&= \mathbb{P}(\text{point in dt}|\text{point not before t}, \mathcal{H}_t) \\
&= \mathbb{P}(\text{point in dt}|\mathcal{H}_t) \\
&= \mathbb{E}[N(\mathrm{d}t)|\mathcal{H}_t].
\end{aligned}$$

Here $N(A)$ denotes the number of points falling in an interval, and the last equality follows from the assumption that no points coincide, so that there is either zero or one point in an infinitisemal interval. In other words, the conditional intensity function specifies the mean number of events in a region conditional on the past.

We consider a few examples of point processes where the conditional intensity has particular functional forms:

Example 2.2 (Poisson process) The (inhomogeneous) Poisson process is among other things characterised by the number of points in disjoint sets being independent. The conditional intensity function inherets this independence. The Poisson process is quite simply the point process where the conditional intensity function is independent of the past, i.e. the conditional intensity function is equal to the intensity function of the Poisson process, $\lambda^*(t) = \lambda(t)$.

Example 2.3 (Hawkes process) Define a point process by the conditional intensity function

$$\lambda^*(t) = \mu + \alpha \sum_{t_i < t} \exp(-(t - t_i)), \tag{2}$$

where μ and α are positive parameters. Note that each time a new point arrives in this process, the conditional intensity grows by α and then decreases exponentially back towards μ. In other words, a point increases the chance of getting other points immediately after, and thus this is model for clustered point patterns. A

simulation of the process with parameters $(\mu, \alpha) = (0.5, 0.9)$ is shown in Figure 2 together with its conditional intensity function (in Section 4 we will learn how to make such a simulation). The so-called Hawkes process is a generalization of this process and has the conditional intensity function

$$\lambda^*(t) = \mu(t) + \alpha \sum_{t_i<t} \gamma(t - t_i; \beta),$$

where $\mu(t) \geq 0$, $\alpha > 0$, and $\gamma(t; \beta)$ is a density on $(0, \infty)$ depending on some parameter β. For more on the Hawkes process, see e.g. Hawkes (1971b,a, 1972); Hawkes and Oakes (1974).

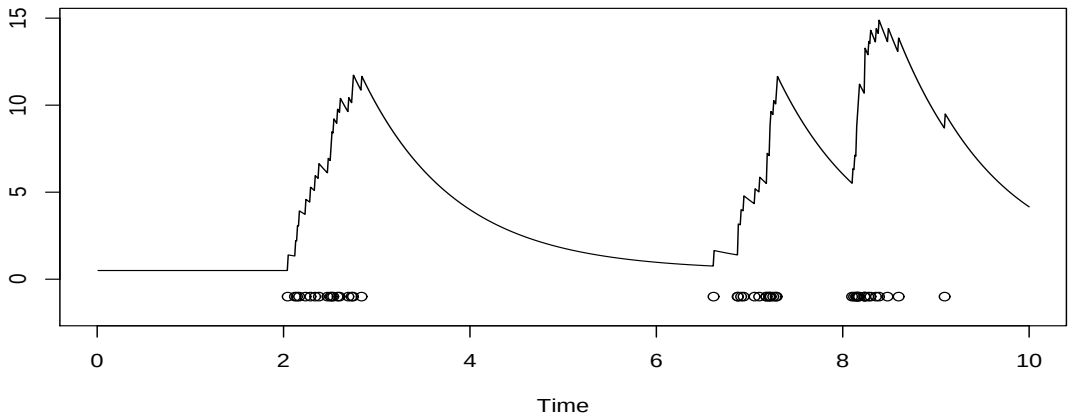

Figure 2: A simulation of the Hawkes process is shown at the bottom of this plot, and the corresponding conditional intensity function is shown in the top. Note that the point pattern is clustered.

Example 2.4 (Self-correcting process) What do we do if we want a point process for regular point patterns? Exchanging the plus for a minus in the Hawkes process will not work, since a conditional intensity function has to be non-negative. We can instead use

$$\lambda^*(t) = \exp\left(\mu t - \sum_{t_i<t} \alpha\right),$$

where μ and α are positive parameters. Now the intensity rises as time passes, but each time a new point appears we multiply by a constant $\exp(-\alpha) < 1$, and thus the chance of new points decreases immediately after a point has appeared; in other words, this is a regular point process. A simulation point pattern and the

conditional intensity function is shown in Figure 3. This process is a special case of the so-called self-correcting process (Isham and Westcott, 1979).

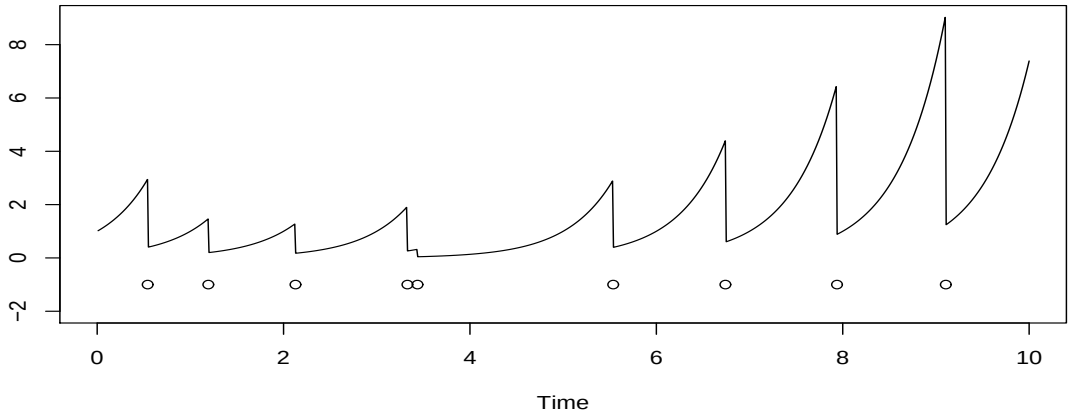

Figure 3: A simulation of a self-correcting process is shown at the bottom of this plot, and the corresponding conditional intensity function is shown in the top. Note that the point pattern is regular.

Note that the models in examples 2.3 and 2.4 are specified simply by choosing a particular form of the conditional intensity and interpreting this. A little creativity and common sense can be used to define many new models using the conditional intensity function. This, of course, depends on the fact that the conditional intensity function uniquely defines a point process. To prove this we first need to note that the definition of the conditional intensity function can also be reversed such that an expresion for the density or cumulative distribution function of the interevent times can be obtained:

Proposition 2.1 *The reverse relation of (1) is given by*

$$f^*(t) = \lambda^*(t) \exp\left(- \int_{t_n}^{t} \lambda^*(s) \mathrm{d}s \right), \tag{3}$$

or

$$F^*(t) = 1 - \exp\left(- \int_{t_n}^{t} \lambda^*(s) \mathrm{d}s \right), \tag{4}$$

where t_n is the last point before t.

Proof. By (1), we get that

$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)} = \frac{\frac{\mathrm{d}}{\mathrm{d}t}F^*(t)}{1 - F^*(t)} = -\frac{\mathrm{d}}{\mathrm{d}t}\log(1 - F^*(t)). \tag{5}$$

Integrating both sides, we get by the fundamental theorem of calculus that

$$\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s = -(\log(1 - F^*(t)) - \log(1 - F^*(t_n))) = -\log(1 - F^*(t)),$$

since $F^*(t_n) = 0$ here (point t_{n+1} falls on top of t_n with probability zero, since the point process is simple). Isolating $F^*(t)$ we get (4), and (3) then follows by differentiating $F^*(t)$ with respect to t, again using the fundamental theorem of calculus. □

Proposition 2.2 *A conditional intensity function $\lambda^*(t)$ uniquely defines a point process if it satisfies the following conditions for all t and all possible point patterns before t:*

1. *$\lambda^*(t)$ is well-defined and non-negative,*
2. *the integral $\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s$ is well-defined,*
3. *$\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s \to \infty$ for $t \to \infty$.*

Proof. The distribution of the point process is well-defined, if all interevent times have well-defined densities, i.e. $f^*(t)$ should be a density function for all t, or equivalently $F^*(t)$ should be a cumulative distribution function. From the three assumptions and (4) it follows that

- $0 \leq F^*(t) \leq 1$,
- $F^*(t)$ is a non-decreasing function of t,
- $F^*(t) \to 1$ for $t \to \infty$,

which means that the distributions of the interevent times are well-defined. Uniqueness follows from Proposition 2.1, since $F^*(t)$ is uniquely obtained from $\lambda^*(t)$ using (4). □

Note that item 3. in Proposition 2.2 implies that the point process continuous forever, a property which is often not desireable for practical use - luckily we can get rid of this assumption. If we remove this, the proof still holds except that item 3. in the proof has to be removed. Now $F^*(t) \to p$ for some probability $p < 1$, so we have to understand what it means when the cumulative distribution function

for the interevent time does not tend to one when time tends to infinity. Basically this means that there is only probability p of having one (or more) points in the rest of the process, and with probability $1-p$ the process terminates with no more points.

Example 2.5 (Two terminating point processes) Consider a unit-rate Poisson process on $[0,1]$. This has conditional intensity function $\lambda^*(t) = \mathbf{1}[t \in [0,1]]$. Thus starting at zero (with no points so far), we get that

$$F^*(t) = 1 - \exp\left(-\int_0^t \mathbf{1}[s \in [0,1]]\mathrm{d}s\right) = 1 - \exp\left(-\min\{t,1\}\right),$$

where $\mathbf{1}[\cdot]$ denotes the indicator function. For $t > 1$, this equals $1-\exp(-1) \approx 0.63$, so there is a probability of about 0.37 of having no points at all. If we do get a point there is an even smaller chance of getting another point after the first one. Another terminating unit-rate process could be a process that stops after getting n points. In this case

$$F^*(t) = (1 - \exp(-t))\mathbf{1}[N([0,t) < n].$$

Both these examples illustrate that assumption 3. in Proposition 2.2 is not necessary to get well-defined point processes.

2.4 The marked case

The conditional intensity function also generalises to the marked case, but before we get that far it is worth reminding ourselves that the mark space $\mathbb{M}$ can be many different types of spaces, often (a subset of) $\mathbb{R}$ or $\mathbb{N}$. We can specify the distribution of the mark κ associated with the point t by its conditional density function $f^*(\kappa|t)$, i.e. this specifies the distribution of the mark given t and the history $\mathcal{H}_t$, which now includes information of both times and marks of past events. Here the term density function is used in a broad sense: if the mark is a continuous random variable, this is the usual (conditional) density function, but if it is a discrete random variable, this is its (conditional) probability function.

We can now define the conditional intensity function for the marked case as

$$\lambda^*(t,\kappa) = \lambda^*(t) f^*(\kappa|t),$$

where $\lambda^*(t)$ is called the *ground intensity*, and is defined exactly as the conditional intensity function for the unmarked case, except that it is allowed to depend on the marks of the past events also (note the close resemblance of this formula with

$p(x, y) = p(x)p(y|x)$ for the relation between the joint, marginal and conditional distributions). Thus we can rewrite this expression to

$$\lambda^*(t, \kappa) = \lambda^*(t) f^*(\kappa|t) = \frac{f^*(t) f^*(\kappa|t)}{1 - F^*(t)} = \frac{f^*(t, \kappa)}{1 - F^*(t)},$$

where $f^*(t, \kappa)$ is the joint density of the time and the mark (again the word the density is used in a broad sense), and $F^*(t)$ is the cumulative distribution function of t conditional on the past times and marks. Therefore following the same arguments as in Section 2.3, the conditional intensity function $\lambda^*(t, \kappa)$ can now be interpreted for the case of discrete marks by

$$\lambda^*(t, \kappa)\mathrm{d}t = \mathbb{E}[N(\mathrm{d}t \times \kappa)|\mathcal{H}_t],$$

that is, the mean number of points in a small time interval $\mathrm{d}t$ with the mark κ. Similarly for the continuous case,

$$\lambda^*(t, \kappa)\mathrm{d}t\mathrm{d}\kappa = \mathbb{E}[N(\mathrm{d}t \times \mathrm{d}\kappa)|\mathcal{H}_t],$$

that is, the mean number of points in a small time interval $\mathrm{d}t$ with the mark in a small interval $\mathrm{d}\kappa$.

We revisit the Hawkes process from Example 2.3, now with marks:

Example 2.6 (marked Hawkes process) The ETAS (epidemic type aftershock sequence) model is a particular type of marked Hawkes process for modelling earthquakes times and magnitudes. Here $\kappa_i \in [0, \infty)$ denotes the magnitude of an earthquake occurring at time t_i. In its simplest form the ETAS model can be defined by its ground intensity

$$\lambda^*(t) = \mu + \alpha \sum_{t_i < t} e^{\beta \kappa_i} e^{-\gamma(t - t_i)},$$

where $\alpha, \beta, \gamma > 0$ are parameters, and an exponential distribution as its mark density

$$f^*(\kappa|t) = \delta e^{-\delta \kappa}.$$

Equivalently we could define it by its conditional intensity function including both marks and times

$$\lambda^*(t, \kappa) = \left(\mu + \alpha \sum_{t_i < t} e^{\beta \kappa_i} e^{-\gamma(t - t_i)} \right) \delta e^{-\delta \kappa}.$$

The idea behind using this model is that earthquakes cause aftershocks - this is reflected in the fact that every new earthquake increases the intensity by $\alpha e^{\beta \kappa_i}$. Note that large earthquakes increase the intensity more than small earthquakes. For more on the ETAS model, see e.g. Ogata (1988, 1998).

We sometimes make simplifying independence assumptions on the marks. An *unpredictable mark* is a mark that does not depend on the past (and therefore cannot be "predicted" using the information about the past, hence the term "unpredictable"). Example 2.6 has unpredictable marks, since $f^*(\kappa|t)$ does not depend on the past. An even stronger assumption is that of an *independent mark*, which means that κ_i is independent of everything else except maybe t_i. Example 2.6 does not have independent marks, since the ground intensity depends on the past marks (which is just another way of saying that the marks depend on the future events).

3 Inference

There are many possibilities for estimating the parameters in a process specified by a conditional intensity function. The likelihood function for such a process has a fairly simple expression, which usually means that maximum likelihood inference or Bayesian inference are good choices.

3.1 Likelihood function

Assume that we have observed a point pattern $(t_1, \dots, t_n)$ on $[0, T)$ for some given $T > 0$, and if we are in the marked case, also its accompanying marks $(\kappa_1, \dots, \kappa_n)$. Then the likelihood function is given by the following proposition.

Proposition 3.1 *Given an unmarked point pattern $(t_1, \dots, t_n)$ on an observation interval $[0, T)$, the likelihood function is given by*

$$L = \left(\prod_{i=1}^{n} \lambda^*(t_i) \right) \exp\left(- \int_0^T \lambda^*(s) \mathrm{d}s \right).$$

Given a marked point pattern $((t_1, \kappa_1), \dots, (t_n, \kappa_n))$ on $[0, T) \times \mathbb{M}$, the likelihood function is given by

$$L = \left(\prod_{i=1}^{n} \lambda^*(t_i, \kappa_i) \right) \exp\left(- \int_0^T \lambda^*(s) \mathrm{d}s \right).$$

Proof. The likelihood function is the joint density function of all the points in the observed point pattern $(t_1, \dots, t_n) \in [0, T)$, and can therefore be factorised into all the conditional densities of each points given all points before it. This yields

$$L = f^*(t_1) \cdots f^*(t_n)(1 - F^*(T)),$$

where the last term $(1 - F^*(T))$ appears since the unobserved point t_{n+1} must appear after the end of the observation interval. Using (1) and (3), we get that

$$\begin{aligned} L &= \left(\prod_{i=1}^{n} f^*(t_i)\right) \frac{f^*(T)}{\lambda^*(T)} \\ &= \left(\prod_{i=1}^{n} \lambda^*(t_i) \exp\left(-\int_{t_{i-1}}^{t_i} \lambda^*(s)\mathrm{d}s\right)\right) \exp\left(-\int_{t_n}^{T} \lambda^*(s)\mathrm{d}s\right) \\ &= \left(\prod_{i=1}^{n} \lambda^*(t_i)\right) \exp\left(-\int_{0}^{T} \lambda^*(s)\mathrm{d}s\right), \end{aligned}$$

where $t_0 = 0$. This proves the result for the unmarked case. To obtain the result for the marked case, start by the factorisation

$$L = f^*(t_1)f^*(\kappa_1|t_1)\cdots f^*(t_n)f^*(\kappa_n|t_n)(1 - F^*(T))$$

All the terms except the conditional mark densities are the same as in the unmarked case, so

$$\begin{aligned} L &= \left(\prod_{i=1}^{n} f^*(\kappa_i|t_i)\right)\left(\prod_{i=1}^{n} \lambda^*(t_i)\right) \exp\left(-\int_{0}^{T} \lambda^*(s)\mathrm{d}s\right) \\ &= \left(\prod_{i=1}^{n} \lambda^*(t_i, \kappa_i)\right) \exp\left(-\int_{0}^{T} \lambda^*(s)\mathrm{d}s\right), \end{aligned}$$

which establishes the result for the marked case. □

3.2 Estimation

Although Proposition 3.1 gives an explicit expression for the likelihood function, it is rarely simple enough that we can find the maximum likelihood estimate (MLE) analytically. One special case where we can find the MLE is the homogeneous Poisson process:

Example 3.1 (MLE for the homogeneous Poisson process) For the homogeneous Poisson process with intensity $\lambda^*(t) = \lambda$, the likelihood simplifies to

$$L = \lambda^n \exp(-\lambda t).$$

Differentiating this and equating to zero, we get that the MLE is given by

$$\hat{\lambda} = \frac{n}{t}.$$

Note that this expression does not depend on the times of the points, only the total number of points. However, this is not true for other processes.

For most other point processes we will require numerical methods to obtain estimates, such as Newton-Raphson for maximizing the likelihood, or Markov chain Monte Carlo for approximating the posterior in a Bayesian approach.

4 Simulation

Simulation turns out to be fairly easy when the conditional intensity function is specified. The conditional intensity function leads to two different approaches for simulating a point process: The inverse method and Ogata's modified thinning algorithm. Both are generalisations of similar methods for simulation of inhomogeneous Poisson processes.

4.1 Inverse method

The basic idea in the inverse method is that we simulate a unit-rate Poisson process (this is just a series of independent exponential random variables with mean one) and transform these into the desired point process.

Let the *integrated conditional intensity function* (or integrated hazard function) be given by

$$\Lambda^*(t) = \int_0^t \lambda^*(s)\mathrm{d}s.$$

The inverse of Λ^* is useful in transforming point patterns as stated by the following proposition.

Proposition 4.1 *If $(s_i)_{i\in\mathbb{Z}}$ is a unit rate Poisson process on $\mathbb{R}$, and $t_i = \Lambda^{*-1}(s_i)$, then $(t_i)_{i\in\mathbb{Z}}$ is a point process with intensity $\lambda^*(t_i)$.*

We skip the proof of this theorem.

Although the point process is defined on the whole of $\mathbb{R}$ in Theorem 4.1, this condition can be relaxed. If we instead use a Poisson process with $s_i \in [0, T]$, then we get a new point process with $t_i \in [0, \Lambda^{*-1}(T)]$, i.e. we also need to transform the final end point. This means we cannot simply simulate a Poisson process on the interval needed, since this interval changes during the transformation, so we need to simulate one exponential variable at a time, and then transform them to see if our simulation fills out the whole interval. The following algorithm does this.

Algorithm 4.1 (Simulation by inversion)

1. Set $t = 0$, $t_0 = 0$ and $n = 0$ (note that t_0 is not an event).
2. Repeat until $t > T$:

(a) Generate $s_n \sim \text{Exp}(1)$.

(b) Calculate t, where $t = \Lambda^{*-1}(s_n)$.

(c) If $t < T$, set $n = n + 1$ and $t_n = t$.

3. Output is $\{t_1, \ldots, t_n\}$.

The difficult part of this algorithm is of course calculating t in step 2(b) since this requires finding the inverse of the integrated conditional intensity function. Notice that since λ^* is non-negative, we get that Λ^* is non-decreasing. Strictly speaking, this means that Λ^* may not even be an invertible function, since it can be constant on intervals (corresponding to λ^* being zero in these intervals). However, any point s_i from the Poisson process will hit these points with probability zero, so we never need to evaluate Λ^{*-1}, where it is not well-defined.

Example 4.1 (Hawkes process, Inverse method) We revisit the special case of Hawkes process from Example 2.3 given by (2). For this we get the integrated conditional intensity function

$$\Lambda^*(t) = \mu t + \alpha \sum_{t_i < t} \left(1 - e^{-(t-t_i)}\right).$$

Looking at the expression, it seems to be hard solve this with respect to t, so an analytical expression for Λ^{*-1} is not available, meaning we will need to approximate this when we use Algorithm 4.1. A simple way of doing this is to calculate $\tilde{s}_i = \Lambda^*(\tilde{t}_i)$ starting at very small values of $\tilde{t}_i$ and then increase $\tilde{t}_i$ until $s_i \approx \Lambda^*(\tilde{t}_i)$, and then use $t_i = \tilde{t}_i$.

The easiest way to generalise this to the marked case is to simulate the associated mark to an event t_i just after we have transformed s_i to t_i (notice that we have all the information that this may depend on, since we have already simulated the past events and marks).

4.2 Ogata's modified thinning algorithm

Ogata's modified thinning algorithm (Ogata, 1981) is a thinning algorithm based on simulating homogeneous Poisson processes with too high intensities and then thin out the points that are too many according to the conditional intensity function. Since the conditional intensity function depends on the past, we have to do this starting in the past and follow the direction of time.

The basic idea behind the algorithm is that when we are at time t we need to find out where to place the next point $t_i > t$. To do this we simulate a homogeneous Poisson process on some interval $[t, t + l(t)]$ for some chosen function $l(t)$ (this is

the maximum distance we may go forward in time from t and it may be infinite). This Poisson process has a chosen constant intensity on $[t, t + l(t)]$, which fulfills

$$m(t) \geq \sup_{s \in [t, t+l(t)]} \lambda^*(s). \tag{6}$$

Actually we only need to simulate the first point t_i of this Poisson process. There are now two possibilities: If $t_i > l(t)$, then there is no point in $[t, t + l(t)]$, so we start again from $t + l(t)$, but if $t_i \leq l(t)$, there may be a point at t_i in $[t, t + l(t)]$. In the latter case we need to figure out whether to keep this point or not. By independent thinning, we keep it with probability $\lambda^*(t_i)/m(t)$. Whether or not we keep it, we start all over at t_i.

Algorithm 4.2 (*Ogata's modified thinning algorithm.*)

1. Set t=0 and n=0.
2. Repeat until $t > T$:
   (a) Compute $m(t)$ and $l(t)$.
   (b) Generate independent random variables $s \sim \text{Exp}(m(t))$ and $U \sim \text{Unif}([0, 1])$.
   (c) If $s > l(t)$, set $t = t + l(t)$.
   (d) Else if $t + s > T$ or $U > \lambda^*(t + s)/m(t)$, set $t = t + s$.
   (e) Otherwise, set $n = n + 1$, $t_n = t + s$, $t = t + s$.
3. Output is $\{t_1, \ldots, t_n\}$.

Proposition 4.2 *The output of Algorithm 4.2 is a realisation of a point process with conditional intensity function $\lambda^*(t)$.*

Proof. It follows from independent thinning that this process has the right conditional intensity function (essentially the explanation above the algorithm is the proof).

Example 4.2 (Hawkes process, Ogata's modified thinning algorithm) In order to use the algorithm we need to choose the $m(t)$ and $l(t)$, and the only requirement is that the inequality (6) is fulfilled at any possible step of the algorithm. Since

$$\lambda^*(t) = \mu + \alpha \sum_{t_i < t} \exp(-(t - t_i)),$$

is non-increasing (except when new points appear), we can choose $m(t) = \lambda(t)$ at every starting point t in the algorithm, and $l(t) = \infty$. This choice can be

used for any point process where $\lambda^*(t)$ only increases when new points arrive. So the Hawkes process can be simulated either by the inverse method or Ogata's modified thinning algorithm (but in fact there are simpler methods for simulating the Hawkes process, see e.g. Møller and Rasmussen (2005, 2006)).

It is easy to generalise the algorithm to the marked case: every time we keep a point t_i in the algorithm, we should simulate its marks from the mark distribution $f^*(\kappa_i|t_i)$ (just as for the inverse method we have the required knowledge of the past when we need to simulate this).

4.3 Why simulate a point process?

Simulations of point processes are useful for many things:

What does a point pattern typically look like? Simulating a point process a couple of times for a given model and a given set of parameters will provide valuable information on what a typical point pattern looks. Is it clustered or regular? Is it inhomogeneous or homogeneous? Does it look anything remotely like the data you are going to spend the next week fitting the model to?

Prediction: Given an observed past, what does the future hold? The specification of the conditional intensity function means that it is easy to include the already observed past, and then simulate the future.

Model checking: Prediction can also be used for model checking if we only use the data in the first half of the observation interval to fit a model, and then simulate predictions of the second half to see if this corresponds to the second half of the observed data. Or we can use all of the data, and compare with simulations of the whole dataset.

Summary statistics: Many quantities can be calculated explicitly from the conditional intensity function, such as the probability of getting no events in the next month or the mean time to the next event. However, particularly complicated summary statistics may not be available on closed form, but can instead be approximated by simulation. For example, the mean number of events in a given time interval may not be available on closed form for a complicated model, but we can then approximate it by the average number of points in a number of simulations.

5 Model checking

In addition to the model checking approaches mentioned in Section 4.3, there is a particular kind of model checking associated with the conditional intensity function known as residual analysis.

5.1 Residual analysis

Residual analysis (Ogata, 1988) is a type of model checking for point processes specified by a conditional intensity function. It is based on the reverse transformation than the one used in Proposition 4.1.

Proposition 5.1 *If $(t_i)_{i\in\mathbb{Z}}$ is a point process with intensity $\lambda^*(t_i)$, and $s_i = \Lambda^*(t_i)$, then $(s_i)_{i\in\mathbb{Z}}$ is a unit rate Poisson process.*

Thus if a point pattern is a realization of a point process with conditional intensity function λ^*, then the integrated conditional intensity function will transform the pattern into a realization of a unit rate Poisson process. In practice this means that if we have modelled an observed point pattern with a point process, and the type of point process is well-chosen, then the transformed pattern should closely resemble a unit-rate Poisson process. In other words, the model checking boils down to checking whether the interevent times are independent exponential variables with mean one.

If the model does not fit, residual analysis may provide important information on how it does not fit. For example, if the data contains an unrealistically large gap for the model between t_i and t_{i+1}, then the transformed data will contain a large gap between s_i and s_{i+1}, i.e. $s_{i+1} - s_i$ will be to large to realistically come from a unit rate exponential distribution. A bit of creativity in analysing the residuals can give us all kinds of information about the original point pattern.

6 Concluding remarks

We have now seen that the conditional intensity function is a valuable tool for point process modelling, and can be used at all stages of data analysis:

- Preliminary analysis (simulation of potential models)
- Model specification and interpretation.
- Parameter estimation (maximum likelihood or Bayesian estimation).
- Model checking (residual analysis or simulation based approaches).
- Prediction.

However, we should note that basing parameter estimation and model checking on the same functions of the data is usually considered bad practice. For example, if we fit a model using maximum likelihood estimation, we have essentially fitted the conditional intensity function as well as we can, and it should not come as a

surprise if the residuals fit rather well, since they are also based on the conditional intensity function. Here it would be more appropriate to base the model checking on other aspects of the model (such as the summary statistics given for example in Møller and Waagepetersen (2004)), which may not be caught so well by the conditional intensity function.

References

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Springer, New York, 2nd edition.

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*. Springer, New York, 2nd edition.

Hawkes, A. G. (1971a). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society Series B*, **33**, 438–443.

Hawkes, A. G. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, **58**(1), 83–90.

Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. In P. A. W. Lewis, editor, *Stochastic Point Processes*, pages 261–271. Wiley, New York,.

Hawkes, A. G. and Oakes, D. (1974). A cluster representation of a self-exciting process. *Journal of Applied Probability*, **11**, 493–503.

Isham, V. and Westcott, M. (1979). A self-correcting point process. *Stoch. Proc. Appl.*, **8**, 335–347.

Møller, J. and Rasmussen, J. G. (2005). Perfect simulation of hawkes processes. *Adv. in Appl. Probab.*, **37**(3), 629–646.

Møller, J. and Rasmussen, J. G. (2006). Approximate simulation of hawkes processes. *Methodol. Comput. Appl. Probab.*, **8**, 53–65.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall, Boca Raton, Florida.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, **IT-27**(1), 23–31.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, **83**(401), 9–27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, **50**(2), 379–402.