

# Table of contents

- General theory
  - Point processes
  - Parametric models
  - Goodness-of-fit test
- Algorithm specific theory
  - Log-likelihood
  - Estimation
  - Distances
  - Simulation
  - Time transform
  - Fit test

## General theory

### Point processes

The tools available here are based on simple, unmarked, point processes on the real line. These are generally used to model instants when some event occurs. A simple example can provide some intuition.

Imagine a single lane tunnel with a sensor that sends a signal when a vehicle enters it. There are (at least) two ways we could monitor it. The first is to let a counter running, so each time a signal is received, we add 1 to the counter. The second is to maintain a list of times, adding a new line every time a signal is received. In any case, graphically we could represent time as the real line and events as points in it.

What are the specific properties that make this kind of point processes applicable in this situation? The first is that it is impossible to have an infinite number of cars entering the tunnel in some limited time interval. The second is that, since it is a single lane tunnel, it is impossible for 2 vehicles to enter it at the same time. These 2 properties make this a ‘simple’ process. Lastly, the signal we receive is always the same, so all we measure is the time it arrives. This refers to the ‘unmarked’ property.

There are two equivalent ways to formally model these processes.

- (1) A counting process  $(N(t)|t > 0)$ , which satisfies three properties: (i)  $N(0) = 0$ , (ii)  $N$  is a right continuous increasing step function with increments of 1, and (iii)  $N$  is almost surely finite. In this setting,  $N(t)$  is the number of events that occur during the interval  $[0, t]$ , therefore, for any  $s_1 < s_2$ ,  $N(s_2) - N(s_1)$  is the number of events observed in the interval  $(s_1, s_2]$ . Condition (i) means that the process starts at  $t = 0$ , (ii) ensures that two events cannot occur at the same time, and (iii) guarantees that only a finite amount of events can occur in any given interval. (the counter in the example)
- (2) A sequence of random variables  $\tau = (t_1, t_2, \dots)$  such that the probabilities of the variables being non-negative and non-decreasing is equal to 1 and, for any bounded time interval, the number of points in that interval is almost surely finite. In our application,  $t_i$  stands for the time of the  $i$ -th volcanic eruption in our data series. (the list in the example)

The most important concept related with point processes is that of the conditional intensity function (CIF). Although events occur randomly, the likelihood of them occurring might change over time. For example, it is impossible to know when a vehicle will enter the tunnel or even how many will enter in one hour, but if we record the signals for 24 hours, we might expect that more signals will arrive between 15:00 to 19:00 than from 2:00 to 6:00. Therefore we can say that the expected number of events per unit of time might vary (signals / hour). This variability is captured by the CIF, which, at each point, quantifies the ‘instantaneous likelihood’ that an event happens.

Formally, for some counting process  $N$ , define  $\mathcal{H}(t)$  as the history of the process in  $[0, t]$ . The CIF of the process is a function  $\lambda$  defined on  $[0, \infty)$  and is given by the identity

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{E[N(t+h) - N(t) | \mathcal{H}(t)]}{h}$$

The CIF quantifies, at each point, how many events are expected to occur per unit of time, therefore the larger the CIF, the more events will tend to occur in this period, as shown in below. A twofold increase in the CIF in some interval translates to a twofold increase in the expected number of events. More generally, the expected number of events in any interval  $(s_1, s_2]$  can be calculated as  $E[N(s_2) - N(s_1)] = E[\int_{s_1}^{s_2} \lambda(t)dt]$ .

## Parametric models

Given some data, for example the list of times when signals arrived from the example above, the main interest here is to test different hypothesis on how the CIF changes over time. Four different possibilities are implemented here: (1) A (homogeneous) Poisson process, (2) an inhomogeneous Poisson process, (3) a Hawkes process, and (4) a Hawkes process with time-varying base intensity (we will call this a inhomogeneous Hawkes process for clarity, but this is not the standard).

### Poisson

In this model the CIF of the process that generated our observations are constant, which means that the rate of arrivals is always constant and, therefore, independent of any external factors. The CIF in this case contains only one positive parameter  $\mu$  which defines the function

$$\lambda_{HP}(t; \mu) = \mu$$

### Inhomogeneous Poisson

In this model there is some external factor that influences the rate of arrivals. Here it is assumed that there is some constant minimum level for the rate of arrivals, but it might be larger at some periods.

Let  $f : [0, T] \rightarrow R$  be the function corresponding to the external factor of interest. To facilitate the interpretation and the comparison between different models,  $f$  will always be normalized so that  $\min(f) = 0$  and  $\max(f) = 1$ . The CIF of the inhomogeneous Poisson process contains an additional parameter  $\gamma$  and is defined as

$$\lambda_{IP}(t; \mu, \gamma) = \mu + \gamma f$$

The parameter  $\mu$  represents the baseline rate at which events occur, regardless of the value of the external factor  $f$ , while  $\gamma$  quantifies how strong the dependency of the process on  $f$  is.

Because of the linearity of the integral, it is possible to determine what influence each of the terms has in the process. If  $N_{IP}(\cdot; \mu, \gamma)$  is the counting process associated to the process with CIF  $\lambda_{IP}(\cdot; \mu, \gamma)$ , then

$$\begin{aligned} E[N_{IP}(s_2; \mu, \gamma) - N_{IP}(s_1; \mu, \gamma)] &= \int_{s_1}^{s_2} \lambda_{IP}(t; \mu, \gamma) dt \\ &= \int_{s_1}^{s_2} (\mu + \gamma f(t)) dt = \mu(s_2 - s_1) + \gamma \int_{s_1}^{s_2} f(t) dt \end{aligned}$$

This means that, on average, this model will generate  $\int_0^T (\mu + \gamma f(t)) dt$  events on the interval  $[0, T]$ , and we expect  $\mu T$  of them to be caused by the baseline rate (therefore independently of  $f$ ) and  $\gamma \int_0^T f(t) dt$  by  $f$ .

### Hawkes

The third model is the Hawkes process. Similarly to the Poisson process, Hawkes processes do not depend on external factors, but they exhibit an internal dependency in which events occurring at a certain time can trigger the occurrence of more events in the future. This self-exciting component can be incorporated in the Poisson process by adding the term  $\sum_{t_i < t} \alpha e^{-\beta(t-t_i)}$  to the term in the definition of the CIF of the homogeneous Poisson process, where  $\alpha$  and  $\beta$  are parameters and  $t_i$  refers to events in the observation  $(t_1, \dots, t_N)$ . Hence,

$$\lambda_{HH}(t; \mu, \alpha, \beta) = \mu + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}$$

This new term causes each event to trigger an increase in the CIF (and, therefore, an increase in the probability of an event occurring) with amplitude  $\alpha$ , which decays exponentially with rate  $\beta$ .

### Inhomogeneous Hawkes

The model above is built simply by adding the self-exciting component to the homogeneous Poisson model. This last model is built analogously, but starting from the inhomogeneous version, i.e.,

$$\lambda_{IH}(t; \mu, \gamma, \alpha, \beta) = \mu + \gamma f + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}$$

In this last case, both the internal and the external dependencies are present.

Earthquakes provide a useful analogy for understanding the last two processes. The occurrence of main shocks is modeled by the terms corresponding to Poisson processes, which may be constant ( $\mu$ ) or dependent on other processes, such as tides or orbital changes ( $\mu + \gamma f(t)$ ). After a main shock, there is a probability that aftershocks are triggered ( $\alpha e^{-\beta(t-t_i)}$ ), which in turn might trigger even more events, until this part of the process eventually dies out.

### Goodness-of-fit test

Hypothesis testing for parametric models, such as the ones presented here, has two forms.

Consider the constant intensity model, which admits only one parameter  $\mu$ . In the first form, the hypothesis being tested is whether the eruption record was generated by a process with a specific intensity function  $\lambda(\cdot; \mu_0)$  for a given value of  $\mu_0$ . This type of testing is straightforward since the exact model is known, and distributions related to the process can be directly computed or simulated.

The problem addressed here, however, is of the second form, which is considerably more complex. Here, the hypothesis being tested is whether there exists a parameter  $\mu_0$  such that the eruption record was generated by a process with an intensity function  $\lambda(\cdot; \mu_0)$ . The difficulty in developing such a test arises from the fact that the null hypothesis only specifies a general form for the model, rather than specific parameter values.

It is important to stress what goodness-of-fit tests can and cannot answer. Given some general parametric model, either the hypothesis that the observations came from this model will be rejected, meaning that there is a very low probability that this is in fact the data-generating model, or not rejected, which is not the same as confirming it. Therefore, the question answered by such a test is not ‘were the events generated by this model?’, but instead, ‘is it reasonable to assume that the events were generated by this model?’. While for the former only one specific model should produce the answer ‘yes’, this could happen for multiple distinct ones in the latter case. This is especially true for both versions of the Hawkes process, because it might not be possible to determine if the clustering in the data is actually caused by the self-exciting component of the process or if it is simply emulating an unknown external factor that causes these clusters.

### Test for conditional intensity function with fixed parameters

As already mentioned, testing for CIF with fixed parameters is not the problem this work deals with, but it is the foundation of the statistical method and provides an intuitive understanding of how our goodness-of-fit test works. This procedure, known as the Kolmogorov-Smirnov test, is meant for assessing whether a specific hypothesized CIF  $\lambda_0$  is a good fit for observed events. In this setting  $\lambda_0$  might come directly from a specific function or from fixing parameters in one of the models in ‘parametric models’ Section.

If  $\lambda_0$  is the true intensity process, then the random variables  $X_i = \int_{t_{i-1}}^{t_i} \lambda_0(t) dt$  (with  $t_0 = 0$ ) follow a unit exponential distribution (Laub, Lee and Taimre 2021). Conversely, if the distribution of the  $X_i$ ’s deviates significantly from the exponential, then  $\lambda_0$  is probably not the true CIF. This deviation can be measured with the Kolmogorov-Smirnov distance (KS-distance, for short), which, for events  $(t_1, \dots, t_N)$  and CIF  $\lambda$ , is calculated as

$$KS((t_1, \dots, t_N, \lambda)) = \sup_x \left| F(x) - \frac{1}{N} \sum_{i=1}^N 1_{[0, x]}(X_i) \right|$$

Here,  $F$  is the cumulative distribution function of the unit exponential distribution, and  $1_{[0,x]}$  is the characteristic function of the interval  $[0, x]$ , therefore  $\frac{1}{N} \sum_{i=1}^N 1_{[0,x]}(X_i)$  is the empirical cumulative distribution function of the  $X_i$ 's.

The ‘distance’ between events  $(t_1, \dots, t_N)$  and the hypothesized CIF  $\lambda_0$  is, then, a number  $d = KS((t_1, \dots, t_N), \lambda_0)$  between 0 and 1. What is missing, however, is some reference to judge if this distance  $d$  deviates too much from what is expected of events stemming from  $\lambda_0$ . Although in principle it would be possible to calculate precisely how the KS-distance of processes generated by  $\lambda_0$  is distributed, a simpler solution is to simulate these processes and calculate their distance from  $\lambda_0$ , which would approximate the desired distribution.

A Laplace transform based distance is also provided in this package, so everything that is done here can also be done with a different definition of the distance between an observation and an intensity function.

### Test for conditional intensity function with estimated parameters

In the above example a specific function  $\lambda_0$  was thought to be the true intensity of the process, but that is not the case in our application. It will only be assumed that the intensity comes specifically from one of the models described in ‘parametric models’ section. More precisely, we assume that the true conditional intensity function depends on a set of parameters  $\theta$  (as an example, in the Inhomogeneous Poisson case,  $\theta = (\mu, \gamma)$ ) and that there exists some  $\theta_0$  such that the true intensity of the process is  $\lambda_0(\cdot) = \lambda(\cdot; \theta_0)$ .

The naive approach to this problem would be simply to do everything as in the previous section, but replacing  $\lambda_0$  by  $\lambda(\cdot; \hat{\theta})$ , where  $\hat{\theta}$  is an estimate of the true parameters  $\theta_0$ . However, it is well known that, because it is not possible to simulate processes from the presumed CIF, this produces unwanted bias due to the estimation of the parameters and may lead to incorrect results (Babu and Rao 2004). A modification is, therefore, needed. The bootstrap-based algorithm is as follows.

- Based on observations  $(t_1, \dots, t_N)$  and a candidate model  $\lambda$ , estimate  $\theta_0$  as  $\hat{\theta}$  and compute  $d = KS((t_1, \dots, t_N), \lambda(\cdot; \hat{\theta}))$ .
- Simulate  $B$  point processes  $((t_1^b, \dots, t_{N_b}^b), b = 1, \dots, B)$  with the intensity function  $\lambda(\cdot; \hat{\theta})$ .
- For each simulation, estimate  $\hat{\theta}$  as  $\theta_b^*$  ( $b = 1, 2, \dots, B$ ).
- For each  $b = 1, \dots, B$ , calculate  $d_b = KS((t_1^b, \dots, t_{N_b}^b), \lambda(\cdot; \theta_b^*))$ .
- If there are  $k$  distances such that  $d_b \geq d$ , then the desired p-value is  $\frac{k}{B}$ .

Notice that in items 3 and 4, instead of calculating the KS-distance with respect to the intensity estimated from the original observation, the distance must be calculated with respect to the intensity estimated from the simulation itself.

## Algorithm specific theory

We now present the theory for each of the algorithms implemented here. Each folder is responsible for some specific task, so the categories used here coincide with the different folders in the project.

### Log-likelihoods

For a general point process, given some event history  $\tau = (t_1, \dots, t_N)$  and CIF  $\lambda$  (possibly dependent on the events) on the interval  $[0, T]$ , the likelihood function is given by

$$L(\tau, \lambda) = \left[ \prod_{i=1}^N \lambda(t_i) \right] e^{-\int_0^T \lambda(t) dt}$$

The negative log-likelihood is, therefore,

$$\ell(\tau, \lambda) = \left( \int_0^T \lambda(t) dt \right) - \sum_{i=1}^N \log(\lambda(t_i))$$

By substituting  $\lambda$  with the appropriate parametrization, we get the corresponding equations for the negative log-likelihood of each model with respect to the parameters.

Additional information in P. Laub, Y. Lee and T. Taimre - The elements of Hawkes processes and T. Ozaki - Maximum likelihood estimation of Hawkes' self-exciting point processes.

### Poisson

For the Poisson process. Given some parameters  $\mu$ , we have  $\lambda(t; \mu) = \mu$ . Therefore

$$\ell(\tau, \lambda(\cdot; \mu)) = \mu T - \sum_{i=1}^N \log(\mu) = \mu T - \log(\mu)N$$

### Inhomogeneous Poisson

In the inhomogeneous Poisson process, given parameters  $\mu$  and  $\gamma$ , and a function  $f$  on which the process depends, the CIF is given by  $\lambda(t; \mu, \gamma) = \mu + \gamma f(t)$ . Consequently,

$$\ell(\tau, \lambda(\cdot; \mu, \gamma)) = \left( \mu T + \gamma \int_0^T f(t) dt \right) - \sum_{i=1}^N \log(\mu + \gamma f(t_i))$$

### Hawkes process

For the Hawkes process we have  $\lambda(t; \mu, \alpha, \beta) = \mu + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}$ . The negative likelihood is

$$\ell(t; \mu, \alpha, \beta) = \mu T - \sum_{i=1}^N \frac{\alpha}{\beta} (e^{-\beta(t-t_i)} - 1) - \sum_{i=1}^N \log(\mu + \alpha A(i))$$

where  $A(i) = \sum_{t_j < t_i} e^{-\beta(t_i-t_j)}$  for  $i \geq 2$  and  $A(1) = 0$ .

### Inhomogeneous Hawkes

The CIF is  $\lambda(t; \mu, \gamma, \alpha, \beta) = \mu + \gamma f(t) + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}$  and the negative log-likelihood

$$\ell(t; \mu, \gamma, \alpha, \beta) = \left( \mu T + \gamma \int_0^T f(t) dt \right) - \sum_{i=1}^N \frac{\alpha}{\beta} (e^{-\beta(t-t_i)} - 1) - \sum_{i=1}^N \log(\mu + \gamma f(t_i) + \alpha A(i))$$

where  $A(i) = \sum_{t_j < t_i} e^{-\beta(t_i-t_j)}$  for  $i \geq 2$  and  $A(1) = 0$ .

### Estimations

All estimations are based on maximum likelihood estimation (MLE). Here the event history  $\tau$  is fixed, and we want to find the parameters that make the likelihood function as large as possible.

In other words, the parameters estimated by the MLE are the ones that make the observed events as likely as possible.

### Poisson

This is the simplest case, as there is a simple closed formula for the MLE. If the process is over the interval  $[0, T]$  and  $N$  events were observed, then the maximum likelihood estimator  $\hat{\mu}$  for  $\mu$  is

$$\hat{\mu} = \frac{N}{T}$$

which is simply the average number of events per unit of time.

## Inhomogeneous Poisson

The CIF in this case has the form

$$\lambda_{IP}(t; \mu, \gamma) = \mu + \gamma f(t)$$

Here  $\mu$  is a constant term independent of  $f$ , while  $\gamma$  quantifies the dependence of the process on  $f$ .

There is no closed form solution for the MLE in this case, therefore we use MATLAB's **trust-region-reflective** algorithm to optimize the likelihood function of the process. For increased performance and precision, both the gradient and hessian of the objective function was provided.

More on the likelihood function of point process (more particularly of Hawkes processes), T. Ozaki - Maximum likelihood estimation of Hawkes' self-exciting point processes

## Hawkes process

Estimation by optimizing the likelihood function as with the inhomogeneous Poisson process runs into numerical problems, because the objective function around the maximum is extremely flat. The expectation-maximization (EM) algorithm is therefore used.

As mentioned in the previous section, in the Hawkes process we assume that an event can trigger others, but in our observation we do not have access to what generated each process. The idea behind this algorithm is that knowing that would make the calculation easier, so we try to estimate it. This is how it works

1. Start with a initial guess for the parameters  $\bar{\theta} = (\bar{\mu}, \bar{\psi}, \bar{\beta})$ , where  $\bar{\psi} = \bar{\alpha}/\bar{\beta}$ .
2. For each event  $(t_1, \dots, t_N)$ , calculate  $\lambda(t_i; \bar{\theta})$ .
3. For each pair  $t_i > t_j$ , calculate  $p_{i,j} = \alpha e^{-\beta(t_i - t_j)} / \lambda(t_i)$  and for each  $t_i$  define  $p_i = \mu / \lambda(t_i)$ .
4. Go to step 1 replacing the initial guess with  $\bar{\mu} = \sum_{i=1}^N p_i$ ,  $\bar{\psi} = \sum_{i>j} p_{i,j} / N$  and  $\bar{\beta} = \sum_{i>j} p_{i,j} / \sum_{i>j} (t_i - t_j) p_{i,j}$ .

The iteration is stopped when the distance between the parameters in step 1 and in step 4 are smaller then some prefixed threshold. For a thorough explanation and all necessary results, see E. Lewis and G. mohler - A nonparametric EM algorithm for multiscale Hawkes process.

For better numerical stability, instead of applying the algorithm to the observed event history directly, we use the time transformation  $t_i \rightarrow t_i(N/T)$ .

## Inhomogeneous Poisson

We use the same algorithm as used in the Hawkes process, but since the base intensity of the process is not constant anymore, we adapt the previous algorithm.

The initial guess is now  $\bar{\theta} = (\bar{\mu}, \bar{\gamma}, \bar{\psi}, \bar{\beta})$ . The values  $p_i$  must also include the contribution from  $f$ , so, we calculate  $p_{i,\mu} = \mu / \lambda(t_i)$  and  $p_{i,\gamma} = \gamma f(t_i) / \lambda(t_i)$ . Finally, the parameter updating rules remain the same for  $\psi$  and  $\beta$ , but now  $\bar{\mu} = \sum_{i=1}^N p_{i,\mu}$  and  $\bar{\gamma} = \sum_{i=1}^N p_{i,\gamma}$ .

## Distances

Let  $\tau = (t_1, \dots, t_N)$  be an event history and  $\lambda$  a CIF. The time transformed event history is defined as

$$\tau' = (t'_1, \dots, t'_N) \text{ where } t'_i = \int_0^{t_i} \lambda(t) dt \quad \forall i \in \{1, \dots, N\}$$

If  $\tau$  is a process generated by the CIF  $\lambda$ , then time transformed event history  $\tau'$  is an unit Poisson process, therefore the interval between events are exponentially distributed. Define, for  $i \in \{1, \dots, N-1\}$ , the random variables

$$X_i = t'_{i+1} - t'_i = \int_{t_i}^{t_{i+1}} \lambda(t) dt$$

If the hypothesis that  $\lambda$  is the true CIF of the process, then the random variables  $X_i$ ,  $i \in \{1, \dots, N-1\}$ , are i.i.d. and exponentially distributed. Both procedures presented below provide a notion of distance from the random variables  $X_i$  to the exponential distribution, which, in this setting, translates to how much the observed event history deviates from what would be expected from a process stemming from  $\lambda$ .

### KSdist

The Kolmogorov-Smirnov distance from the random variables  $X_i$  to a unit exponential. If  $\tau$  is some event history and  $\lambda$  is a CIF, then

$$KSdist(\tau, \lambda) = \max_{x \in [0, \infty)} \left| (1 - e^{-x}) - \left( \frac{1}{N-1} \sum_{i=1}^{N-1} 1_{[0, x]}(X_i) \right) \right|$$

The first term is the cumulative distribution of a unit exponential, while the second is the empirical cumulative distribution of the  $X_i$ 's.

### LPdist

This distance is based on the Laplace transform of the distributions. Given some random variable  $X$ , the Laplace transform of  $X$  is defined as

$$\psi_X(t) = E[e^{(-tX)}]$$

If  $X$  is the unit exponential distribution, then  $\psi_X(t) = \frac{1}{1+t}$ . We want to compare this to the empirical Laplace transform of the  $X_i$ 's, which is defined as

$$\phi_N(t) = \frac{1}{N} \sum_{j=1}^N e^{-tX_j}$$

Notice that  $\phi_N$  is simply an average, which is the best guess we have given the observations.

Instead of taking the supremum over the difference of the empirical and theoretical functions, we will use a scaling of the  $L^2$  distance between them, specifically

$$LPdist(\tau, \lambda) = \int_0^\infty \left( \phi_N(t) - \left( \frac{1}{1+t} \right) \right)^2 e^{-at} dt$$

Notice that setting the free variable to  $a$  to 0 reduces the identity to the  $L^2$  distance. The term  $e^{-at}$  can be set to different values to make the test more or less sensitive to specific alternative distributions.

Although the calculation of this distance is more complex than the previous one, it has the advantage of being differentiable, which makes it more tractable in theoretical results.

The use of this procedure is based on N. Henze and S. Meintanis - Tests of fit for exponentiality based on the empirical Laplace transform

## Simulations

### Poisson

Suppose we want to simulate a process with parameter  $\mu$  on the interval  $[0, T]$ . The interval between two consecutive events follows an exponential distribution  $E$  with pdf  $\mu e^{-t\mu}$ . The algorithm is as follows.

1.  $t_1 \approx E$ . If  $t_i > T$ .
2. For every  $n \in 2, N$ ,  $t_n \approx t_{n-1} + E$ .
3. Return  $(t_1, \dots, t_N)$  where  $N = \max(\{i \in \mathbb{N} | t_i \leq T\})$

## Inhomogeneous Poisson

For inhomogeneous Poisson processes we use the thinning algorithm.

Suppose we have a process with CIF  $\lambda_0$  and we also have a function  $\lambda_1$  with  $\lambda_0 \geq \lambda_1$ . From the first process, we build a second one as such.

1.  $(t_1, \dots, t_N)$  follow the CIF  $\lambda_0$ .
2. For each  $i \in \{1, \dots, N\}$ , define the values  $u_i$  sampled from a uniform distribution on  $[0, 1]$ .
3. Define  $(t'_1, \dots, t'_M)$  by keeping only the events in  $(t_1, \dots, t_N)$  for which  $u_i \geq \lambda_1(t_i)/\lambda_0(t_i)$ .

A process defined like this is the same as a process following  $\lambda_1$ .

So, given the parameters  $\mu$  and  $\gamma$ , we simulate first a Poisson process with parameter  $\max(\mu + \gamma f(t))_{t \in [0, T]}$ . Now we apply steps 2 and 3 to this simulation with  $\lambda_0(t) = \max(\mu + \gamma f(t))_{t \in [0, T]}$  and  $\lambda_1(t) = \mu + \gamma f(t)$ .

Reference: P. Lewis and G. Shedler - Simulation of nonhomogeneous Poisson processes by thinning

## Hawkes

First we simulate the baseline process  $(t_1, \dots, t_M)$ , which is a Poisson process with parameter  $\mu$ . For each events  $t_i$ , we simulate a process on the interval  $[t_i, T]$  with CIF  $\alpha e^{-\beta t}$ . To simulate this last process we use the inverse time transform technique.

Define  $\Lambda(t) = \int_0^t \alpha e^{-\beta s} ds$ , the integral of the CIF of the process we want to simulate. First, simulate a unitary Poisson process on the interval  $[0, \Lambda(T - t_i)]$ . If  $(s_1, \dots, s_n)$  is the event history simulated, the  $(\Lambda^{-1}(s_1), \Lambda^{-1}(s_2), \dots, \Lambda^{-1}(s_n))$  is a realization of a process with CIF  $\alpha e^{\beta t}$ .

Do this recursively for every new event generated. The union of all events is a Hawkes process.

Reference: Møller and Rasmussen - Perfect simulation of Hawkes processes

## Inhomogeneous Hawkes

The same algorithm used in the previous case, but instead of starting with a Poisson process, we start with an inhomogeneous Poisson process with parameters  $\mu$  and  $\gamma$ .

## Time transform

Suppose we have a point process  $N$  with conditional intensity function (CIF)  $\lambda$ , which can be deterministic, as in both versions of the Poisson process, or random, as in the Hawkes processes. The time transformation  $\Lambda$  is defined as

$$\Lambda(t) = \int_0^t \lambda(s) ds$$

The time transformed process is defined as  $\Lambda \circ N$ . If  $(t_1, t_2, \dots)$  is an outcome of  $N$ , then the outcome of  $\Lambda \circ N$  is  $(\Lambda(t_1), \Lambda(t_2), \dots)$ . The importance of this transformation is that  $\Lambda \circ N$  is a Poisson process with rate equal to 1.

Notice that  $\Lambda \circ N$  is a unitary Poisson process only if the true CIF of the process is  $\lambda$ . Phrasing it the other way around, if  $\lambda$  is not the true CIF, then  $\Lambda \circ N$  is not a unitary Poisson process.

More detailed information can be found in Y. Ogata - Statistical models for earthquake occurrences and residual analysis for point processes or P. Laub, Y. Lee and T. Taimre - The elements of Hawkes processes.

## Fit tests

Here is the goodness-of-fit test implemented. The algorithm is as follows

- Based on observations  $(t_1, \dots, t_N)$  and a candidate model  $\lambda$ , estimate  $\theta_0$  as  $\hat{\theta}$  and compute  $d = D((t_1, \dots, t_N), \lambda(\cdot; \hat{\theta}))$ .
- Simulate  $B$  point processes  $((t_1^b, \dots, t_{N_b}^b), b = 1, \dots, B)$  with the intensity function  $\lambda(\cdot; \hat{\theta})$ .
- For each simulation, estimate  $\hat{\theta}$  as  $\theta_b^*$  ( $b = 1, 2, \dots, B$ ).
- For each  $b = 1, \dots, B$ , calculate  $d_b = KS((t_1^b, \dots, t_{N_b}^b), \lambda(\cdot; \theta_b^*))$ .



- If there are  $k$  distances such that  $d_b \geq d$ , then the desired p-value is  $\frac{k}{B}$ .

The candidate models can be any of the 4 models previously described and  $D$  is either the Kolmogorov-Smirnov distance or the Laplace transform based distance.