Stochastic Processes: Lecture 4

Andrew J. Holbrook

UCLA Biostatistics 270

Spring 2022

Point processes

- A point process is a random list of points $T_i \in \mathcal{T} \subset \mathbb{R}^D$.
- The total number of points $N(\mathcal{T})$ may be fixed or random.
- For $A \subset \mathcal{T}$, let $N(A)$ be the total number of points in A:

$$N(A) = \sum_{i=1}^{N(\mathcal{T})} 1\{T_i \in A\}\,.$$

- We are interested in *non-explosive* point processes, for which

$$\Pr(N(A) < \infty) = 1 \quad \text{when} \quad \mathrm{vol}(A) < \infty\,.$$

Poisson processes

- The points $T_i \in \mathcal{T}$ follow a homogeneous poisson process with intensity $\lambda > 0$ if

  $$N(A_j) \overset{\perp}{\sim} \text{Pois}(\lambda \cdot \text{vol}(A_j))$$

  for disjoint sets $A_j \subset \mathcal{T}$ that satisfy $\text{vol}(A_j) < \infty$.

- Define $\lambda(t) : \mathcal{T} \to [0, \infty)$ so that

  $$\int_A \lambda(t) dt < \infty \quad \text{whenever} \quad \text{vol}(A) < \infty \, .$$

  Then for a non-homogeneous point process on $\mathcal{T}$ with intensity function $\lambda(t)$

  $$N(A_j) \overset{\perp}{\sim} \text{Pois}\left(\int_{A_j} \lambda(t) dt \right)$$

  for disjoint sets $A_j \subset \mathcal{T}$ that satisfy $\text{vol}(A_j) < \infty$.

A sampling technique

Theorem 1
Let T_i be the points of a Poisson process on $\mathcal{T}$ with intensity function $\lambda(t) \geq 0$, where $\Lambda(\mathcal{T}) = \int_{\mathcal{T}} \lambda(t)dt$. Then T_i can be sampled by

1. *generating $N(\mathcal{T}) \sim Pois(\Lambda(\mathcal{T}))$ and*

2. *generating $N(\mathcal{T})$ independent T_i with probabilities*

$$Pr(T_i \in A) = \frac{1}{\Lambda(\mathcal{T})} \int_A \lambda(t)dt \,.$$

A sampling technique

Proof.
For $J \geq 1$, let $A_1, \ldots, A_J$ be disjoint subsets of $\mathcal{T}$ and define $A_0 = \{t \in \mathcal{T} | t \notin \cup_{j=1}^{J} A_j\}$. Let $n_j \geq 0$ for $j = 1, \ldots, J$. Let

$$\begin{aligned} P_* &= \Pr(N(A_1) = n_1, \ldots, N(A_J) = n_J) \\ &= \sum_{n_0=0}^{\infty} \Pr(N(A_0) = n_0, N(A_1) = n_1, \ldots, N(A_J) = n_J)\,. \end{aligned}$$

Set $n = n_0 + n_1 + \cdots + n_J$. Under this sampling scheme:

$$\begin{aligned} P_* &= \frac{n!}{n_0! n_1! \ldots n_J!} \sum_{n_0=0}^{\infty} \frac{e^{-\Lambda(\mathcal{T})} \Lambda(\mathcal{T})^n}{n!} \prod_{j=0}^{J} \left(\frac{\Lambda(A_j)}{\Lambda(\mathcal{T})} \right)^{n_j} \\ &= \sum_{n_0=0}^{\infty} \prod_{j=0}^{J} \frac{e^{-\Lambda(A_j)} \Lambda(A_j)^{n_j}}{n_j!} = \prod_{j=1}^{J} \frac{e^{-\Lambda(A_j)} \Lambda(A_j)^{n_j}}{n_j!}\,. \end{aligned}$$

□

A sampling technique

Corollary 1

Let T_i be the points of a homogeneous Poisson process on $\mathcal{T}$ with intensity $\lambda > 0$, where vol$(\mathcal{T}) < \infty$. Then we may sample the process by

1. *generating $N(\mathcal{T}) \sim$ Pois$(\Lambda(\mathcal{T}))$ and*
2. *generating $T_i \overset{iid}{\sim}$ Uni$(\mathcal{T})$, $i = 1, \ldots, N(\mathcal{T})$.*

Proof.

Apply the Theorem with constant $\lambda(t)$. Then

$$\Pr(T_i \in A) = \frac{\lambda \int_A dt}{\lambda \int_{\mathcal{T}} dt} = \mathrm{vol}(A)/\mathrm{vol}(\mathcal{T})\,.$$

□

Poisson processes on $[0, \infty)$

A Poisson process on $[0, \infty)$ can be represented by the counting function

$$N(t) = N([0, t]) = \sum_{i=1}^{\infty} 1\{T_i \leq t\}, \quad 0 \leq t < \infty.$$

The homogeneous Poisson process on $[0, \infty)$ is defined by these properties:

1. $N(0) = 0$;
2. for $0 \leq s < t$, $N(t) - N(s) \sim \text{Pois}(\lambda(t - s))$;
3. for $0 = t_0 < t_1 < \cdots < t_m$, $N(t_i) - N(t_{i-1})$ are independent.

Simulation methods

It can be shown that

$$T_i - T_{i-1} \sim \exp(\lambda)\,, \quad i \geq 1\,. \tag{1}$$

A heuristic argument says: under (1) and for some x,

$$\Pr(T_i - T_{i+1} > x) = e^{-\lambda x}\,,$$

but if $T_i - T_{i+1} \geq x$, then the interval $(T_{i-1}, T_{i-1} + x)$ has no events. Under the Poisson model, this probability is

$$f(0; \lambda x) = \frac{(\lambda x)^0 e^{-\lambda x}}{0!} = e^{-\lambda x}\,.$$

The exponential spacings method simulates a homogeneous Poisson process thus: setting $T_0 = 0$,

$$T_i = T_{i-1} + E_i\,, \quad E_i \overset{iid}{\sim} \exp(\lambda)\,, \quad i \geq 1\,.$$

Simulation methods

Following previous discussion, we can also simulate a homogeneous Poisson process on $[0, T]$ by

1. generating $N \sim \text{Pois}(\lambda T)$,
2. generating $S_i \overset{iid}{\sim} \text{Uni}([0, T])$, $i = 1, \ldots, N$, and
3. setting $T_i = S_{(i)}$.

Non-homogeneous Poisson process on $[0, \infty)$

The non-homogeneous Poisson process on $[0, \infty)$ has these properties:

1. $N(0) = 0$;
2. for $0 \le s < t$, $N(t) - N(s) \sim \text{Pois}\Big(\int_s^t \lambda(x)dx\Big)$;
3. for $0 = t_0 < t_1 < \cdots < t_m$, $N(t_i) - N(t_{i-1})$ are independent.

The cumulative rate function is $\Lambda(t) = \int_0^t \lambda(x)dx$. Start by assuming $\lim_{t\to\infty} \Lambda(t) = \infty$ and $\lambda(t) > 0$, $\forall t$. Define variables $Y_i = \Lambda(T_i)$ and the counting function

$$N_y(t) = \sum_{i=1}^{\infty} 1\{Y_i \le t\} = \sum_{i=1}^{\infty} 1\{T_i \le \Lambda^{-1}(t)\} = N(\Lambda_{-1}(t)).$$

Non-homogeneous Poisson process on $[0, \infty)$

Define variables $Y_i = \Lambda(T_i)$ and the counting function

$$N_y(t) = \sum_{i=1}^{\infty} 1\{Y_i \leq t\} = \sum_{i=1}^{\infty} 1\{T_i \leq \Lambda^{-1}(t)\} = N(\Lambda_{-1}(t)) \,.$$

Note that $N_y(0) = 0$ and

$$\begin{aligned} N_y(t) - N_y(s) &= N(\Lambda^{-1}(t)) - N(\Lambda^{-1}(s)) \sim \text{Pois}\left(\int_{\Lambda^{-1}(t)}^{\Lambda^{-1}(s)} \lambda(x)dx\right) \\ &= \text{Pois}\Big(\Lambda(\Lambda^{-1}(t)) - \Lambda(\Lambda^{-1}(s))\Big) = \text{Pois}(t-s) \,. \end{aligned}$$

Finally the increments of $N_y(t)$ are increments of $N(\Lambda^{-1}(t))$. Independence of the latter implies independence for the former. Therefore,

$$Y_i = \Lambda(T_i) \sim PP(1) \,.$$

More exponential spacings

We have shown $Y_i = \Lambda(T_i) \sim PP(1)$. Setting $Y_0 = T_0 = 0$, we can therefore simulate T_i thus:

$$
\begin{aligned}
Y_i &= Y_{i-1} + E_i\,, \quad E_i \overset{iid}{\sim} \exp(1)\,, \quad i \geq 1\,,\\
T_i &= \Lambda^{-1}(Y_i) = \Lambda^{-1}(\Lambda(T_{i-1}) + E_i)\,.
\end{aligned}
$$

Comments:

- If $\lim_{t\to\infty} \Lambda(t) = \Lambda_0$, then $\Lambda^{-1}(y)$ does not exists for $y > \Lambda_0$. If $\Lambda(T_i) + E_i > \Lambda_0$, then there is no T_{i+1} and the process stops.
- The algorithm is convenient when Λ and Λ^{-1} are available in closed form.
- The algorithm works even when Λ takes finite jumps or is constant on some intervals by taking

$$
\Lambda^{-1}(y) = \inf\{t \geq 0 | \Lambda(t) \geq y\}\,.
$$

Thinning (rejection sampling for point processes)

Let $\widetilde{\lambda}(t) \geq \lambda(t)$ and assume we can sample from a Poisson process on $\mathcal{T}$ with $\widetilde{\lambda}$ for intensity function. The following algorithm generates $(T_1, \ldots, T_N) \sim NHPP(\mathcal{T}, \lambda)$:

1. Generate $(\widetilde{T}_1, \ldots, \widetilde{T}_{\widetilde{N}})) \sim NHPP(\mathcal{T}, \widetilde{\lambda})$;
2. if $\widetilde{N} > 0$, then for $i \in \{1, \ldots, \widetilde{N}\}$:
   - 2.1 draw $u_i \sim \text{Uni}(0, 1)$;
   - 2.2 if $u_i < \rho(\widetilde{T}_i) = \lambda(\widetilde{T}_i)/\widetilde{\lambda}(\widetilde{T}_i)$, then $\widetilde{T}_i \in \{T_1, \ldots, T_N\}$.

Why thinning works

Let $N(A)$ be the number of points T_i in a set A and $\widetilde{N}(A)$ be the analogue for points $\widetilde{T}_i$. Note that $\widetilde{N}(A) \sim \text{Pois}(\int_A \widetilde{\lambda}(t)dt)$. Then the probability a point in $\widetilde{T}_i \in A$ is accepted is

$$\rho(A) = \frac{\int_A \rho(t)\widetilde{\lambda}(t)dt}{\int_A \widetilde{\lambda}(t)dt} = \frac{\int_A \lambda(t)dt}{\int_A \widetilde{\lambda}(t)dt} .$$

It holds that $N(A)|\widetilde{N}(A) \sim \text{binom}(\widetilde{N}(A), \rho(A))$. Marginalizing over $\widetilde{N}(A)$ gives

$$N(A) \sim \text{Pois}\left(\rho(A) \int_A \widetilde{\lambda}(t)dt\right) = \text{Pois}\left(\int_A \lambda(t)dt\right) .$$

Independence of N on non-overlapping sets is inherited from $\widetilde{N}$. Therefore $(T_1, \ldots, T_N) \sim NHPP(\mathcal{T}, \lambda)$.

The temporal Hawkes process

The temporal Hawkes process is a non-homogeneous Poisson process on $[0, \infty)$ with (conditional) intensity function given by

$$\lambda(t | T_k < t) = \lambda_0 + \sum_{T_K < t} g(t - T_k)\,,$$

where $g > 0$ is a non-increasing *triggering function*. Because the intensity increases after an observation, the Hawkes process is referred to as *self-exciting* and is useful for modeling contagion.

Exponential decay

A common choice for g is the exponential decay function:

$$\lambda(t|T_k < t) = \lambda_0 + \alpha \sum_{T_K < t} e^{-\beta(t - T_k)} .$$

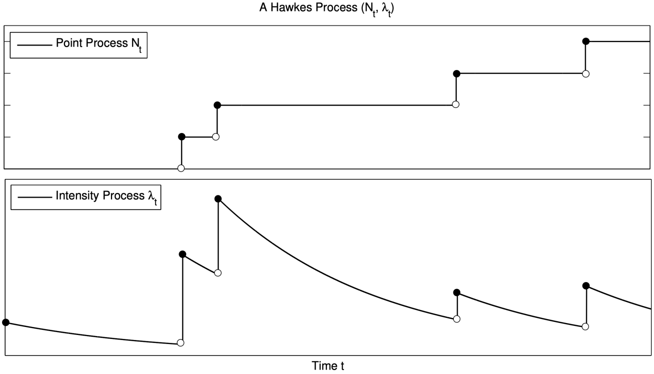


Figure 1: A Hawkes Process with Exponential Decaying Intensity (N_t, λ_t)

Dassios and Zhao, 2013

Exponential decay

A common choice for g is the exponential decay function:

$$\lambda(t|T_k < t) = \lambda_0 + \alpha \sum_{T_K < t} e^{-\beta(t - T_k)} .$$

Exponential decay has pros and cons:

- Pros: exponential decay has computational benefits. Process simulation and likelihood computations are both $O(N)$.
- Cons: the exponential rate of decay may be to fast for certain applications, precluding long-term dependencies.

We will return to linear-time computing for the exponential triggering kernel later.

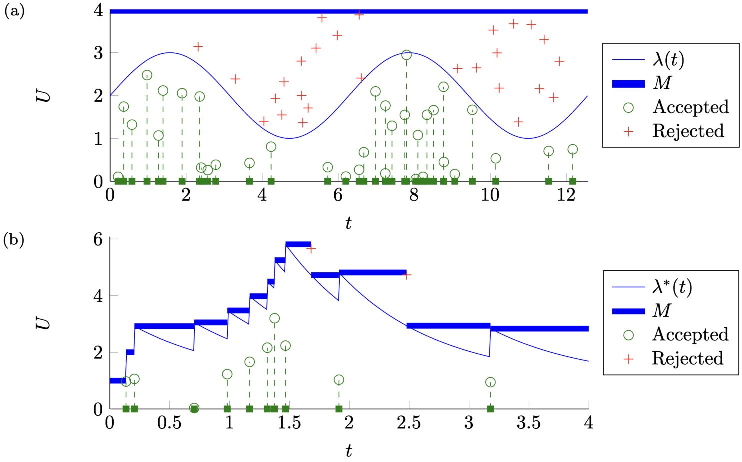
Ogata's modified thinning algorithm

Algorithm 2 Generate a Hawkes process by thinning.

procedure HawkesByThinning(T, $\lambda^*(\cdot)$)
    require: $\lambda^*(\cdot)$ non-increasing in periods of no arrivals.
    $\varepsilon \leftarrow 10^{-10}$ (some tiny value > 0).
    $P \leftarrow []$, $t \leftarrow 0$.
    while $t < T$ **do**
        Find new upper bound:
            $M \leftarrow \lambda^*(t + \varepsilon)$.
        Generate next candidate point:
            $E \leftarrow \text{Exp}(M)$, $t \leftarrow t + E$.
        Keep it with some probability:
            $U \leftarrow \text{Unif}(0, M)$.
        if $t < T$ **and** $U \leq \lambda^*(t)$ **then**
            $P \leftarrow [P, t]$.
        end if
    end while
    return P
end procedure

Laub et al., 2015

Ogata's modified thinning algorithm



Laub et al., 2015

Conditional intensity and stochastic calculus

For the counting process $N(t)$ with histories $\mathcal{H}(t)$, we can define our conditional intensity

$$\lambda(t) = \lim_{h \to 0} \frac{E(N(t+h) - N(t) | \mathcal{H}(t))}{h} = \frac{E(dN(t)|\mathcal{H}(t))}{dt}$$

and for a general Hawkes process we have

$$\lambda(t) = \lambda_0 + \int_0^t g(t-u) dN(u) \,.$$

When g specifies exponential decay, this becomes

$$\lambda(t) = \lambda_0 + \alpha \int_0^t e^{-\beta(t-u)} dN(u) \,,$$

or

$$d\lambda(t) = \beta(\lambda_0 - \lambda(t))dt + \alpha dN(t) \,.$$

Asymptotic normality of Hawkes process

Theorem 2
Assume $0 < n := \int_0^\infty g(s)ds < 1$ *and* $\int_0^\infty sg(s)ds < \infty$, *then the number of HP arrivals in* $(0, t]$ *is asymptotically normally distributed as* $t \to \infty$, *i.e.,*

$$Pr\left(\frac{N(0,t] - \lambda_0 t(1-n)^{-1}}{\sqrt{\lambda_0 t(1-n)^{-3}}} < y\right) \longrightarrow \Phi(y)\,.$$

Note that for the exponential Hawkes model we only have a CLT when $\alpha/\beta < 1$:

$$n = \int_0^\infty g(s)ds = \alpha \int_0^\infty e^{-\beta s} ds = \frac{\alpha}{\beta}\,.$$

Stationarity and explosiveness

Again, let $n := \int_0^\infty g(s)ds$ and define

$$m(t) = E(\lambda(t)) = E\left(\lambda_0 + \int_0^t g(t-u)dN(u)\right)$$
$$= \lambda_0 + \int_0^t g(t-u)\, E(dN(u))\,.$$

But

$$\lambda(t) = \lim_{h\to 0} \frac{E(N(t+h) - N(t)|\mathcal{H}(t))}{h} = \frac{E(dN(t)|\mathcal{H}(t))}{dt}\,,$$

so by iterated expectation

$$m(t) = E(\lambda(t)) = E\left(\frac{E(dN(t)|\mathcal{H}(t))}{dt}\right) = \frac{E(dN(t))}{dt}$$

and $m(t)dt = E(dN(t))$. Thus we obtain the recursion

$$m(t) = \lambda + \int_0^t g(t-s)m(s)ds = \lambda + \int_0^t m(t-s)g(s)ds\,.$$

When $n < 1$, $g(t) \to \lambda/(1-n)$, but when $n > 1$, $g(t)$ diverges to infinity.

More exponential Hawkes process

Definition 1 (Intensity-based)

A Hawkes process with exponentially decaying intensity is a Poisson process $N_t = \{T_k\}_{k=1,2,\dots}$ on $\mathbb{R}_+$ with non-negative $\mathcal{F}_t$-stochastic intensity

$$\lambda_t = a + (\lambda_0 - a)e^{-\delta t} + \sum_{0 \le T_k < t} Y_k e^{-\delta(t - T_k)}, \quad t \ge 0,$$

where:

- $\{\mathcal{F}_t\}_{t\ge0}$ *is a history of the process w.r.t.w.* $\{\lambda_t\}_{t\ge0}$ *is adapted;*
- $a \ge 0$ *is the constant reversion level;*
- $\lambda_0 > 0$ *is the initial intensity at time* $t = 0$*;*
- $\delta > 0$ *is the rate of exponential decay;*
- $\{Y_k\}_{k=1,2,\dots}$ *are positive random variables that are i.i.d. with distribution function* $G(y)$*;*
- $\{T_k\}_{k=1,2,\dots}$ *and* $\{Y_k\}_{k=1,2,\dots}$ *are mutually independent.*

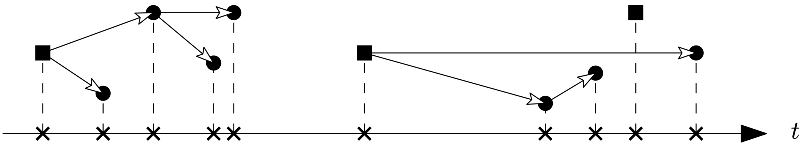
More exponential Hawkes process

Definition 2 (Cluster-based)

A Hawkes process with exponentially decaying intensity is a marked Poisson cluster process $C = \{T_i, Y_i\}_{i=1,2,\ldots}$ with times $T_i \in \mathbb{R}_+$ and marks Y_i: the number of points in $(0, t]$ is defined $N_t = N_{C(0,t]}$; the cluster centers of C are 'immigrants', the rest are 'offspring', and they share the following structure:

- *the immigrants $I = \{T_m\}_{m=1,2,\ldots}$ are distributed as an NHPP with rate $a + (\lambda_0 - a)e^{-\delta t}$;*
- *the marks $\{Y_m\}_{m=1,2,\ldots}$ associated to immigrants I are i.i.d. $Y_m \sim G(y)$ and are independent of the immigrants;*
- *each immigrant T_m generates one cluster C_m, and these clusters are independent;*
- *each cluster C_m is a random set formed by marked points of generations of order $n = 0, 1, \ldots$ with the following branching structure*
  - *the immigrant and its mark (T_m, Y_m) are generation 0;*
  - *given generations $0, 1, \ldots, n$ in C_m each $(T_j, Y_j) \in C_m$ of generation n generates are a Poisson process on (T_j, ∞) with intensity $Y_j \exp(-\delta(t - T_j))$, with $Y_j \sim G$ independent.*
- *C consists of the union of all clusters, i.e., $C = \bigcup_{m=1,2,\ldots} C_m$.*

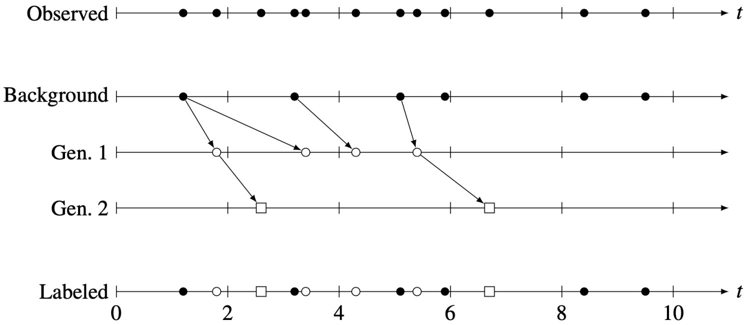
Cluster-based representation



Laub et al., 2015

- The cluster-based or *branching process* definition is equivalent to the intensity-based definition.
- The cluster representation can be used to simulate a Hawkes process just as we have used the intensity representation to simulate.
- The equivalent representations hold beyond exponential HP.

Cluster-based representation



Reinhart, 2018

Explosion and cluster-based representation

Conditioned on knowing the number of children in a generation (say, D_1), the generation's arrival times are distributed i.i.d. with density $g(t - T_i)/n$. Note that for the exponential Hawkes process, we have

$$g(t - T_i)/n = \beta e^{-\beta(t - T_i)} .$$

The expected total number of children in generation i is $E(D_i) = n^i$, so the expected total number of children for one individual is

$$E\left(\sum_{i=1}^{\infty} D_i\right) = \sum_{i=1}^{\infty} E(D_i) = \sum_{i=1}^{\infty} n^i = \begin{cases} \frac{n}{1-n}, & n < 1 \\ \infty, & n \geq 1 \end{cases} .$$

When $n < 1$, it is the ration between the number of descendants for one parent and the size of the entire family (including the parent):

$$\frac{E\left(\sum_{i=1}^{\infty} D_i\right)}{1 + E\left(\sum_{i=1}^{\infty} D_i\right)} = \frac{\frac{n}{1-n}}{1 + \frac{n}{1-n}} = \frac{\frac{n}{1-n}}{\frac{1}{1-n}} = n ,$$

therefore any HP event chosen at random is generated *exogenously* (an immigrant) with probability $1 - n$ or *endogenously* (a child) with probability n.

An incorrect cluster-based simulation algorithm

Algorithm 3 Generate a Hawkes process by clusters.

procedure HawkesByClusters(T, λ, α, β)
  $P \leftarrow \{\}$.
  Immigrants:
    $k \leftarrow \mathrm{Poi}(\lambda T)$
    $C_1, C_2, \dots, C_k \overset{\text{i.i.d.}}{\longleftarrow} \mathrm{Unif}(0, T)$.
  Descendants:
    $D_1, D_2, \dots, D_k \overset{\text{i.i.d.}}{\longleftarrow} \mathrm{Poi}(\alpha/\beta)$.
  for $i \leftarrow 1$ **to** k **do**
    if $D_i > 0$ **then**
      $E_1, E_2, \dots, E_{D_i} \overset{\text{i.i.d.}}{\longleftarrow} \mathrm{Exp}(\beta)$.
      $P \leftarrow P \cup \{C_i + E_1, \dots, C_i + E_{D_i}\}$.
    end if
  end for
  Remove descendants outside $[0, T]$:
    $P \leftarrow \{P_i : P_i \in P, P_i \le T\}$.
  Add in immigrants and sort:
    $P \leftarrow \mathrm{Sort}(P \cup \{C_1, C_2, \dots, C_k\})$.
  return P
end procedure

Laub et al., 2015

More exponential Hawkes process

Definition 3 (Intensity-based)

A Hawkes process with exponentially decaying intensity is a Poisson process $N_t = \{T_k\}_{k=1,2,\ldots}$ on $\mathbb{R}_+$ with non-negative $\mathcal{F}_t$-stochastic intensity

$$\lambda_t = a + (\lambda_0 - a)e^{-\delta t} + \sum_{0 \leq T_k < t} Y_k e^{-\delta(t - T_k)}, \quad t \geq 0,$$

where:

- $\{\mathcal{F}_t\}_{t \geq 0}$ *is a history of the process w.r.t.w.* $\{\lambda_t\}_{t \geq 0}$ *is adapted;*
- $a \geq 0$ *is the constant reversion level;*
- $\lambda_0 > 0$ *is the initial intensity at time* $t = 0$*;*
- $\delta > 0$ *is the rate of exponential decay;*
- $\{Y_k\}_{k=1,2,\ldots}$ *are positive random variables that are i.i.d. with distribution function* $G(y)$*;*
- $\{T_k\}_{k=1,2,\ldots}$ *and* $\{Y_k\}_{k=1,2,\ldots}$ *are mutually independent.*

A fast simulation method

A simulation algorithm for one sample path $\{N_t, \lambda_t\}_{t=0,1,\ldots}$ of a 1D exponential Hawkes process conditional on λ_0 and $N_0 = 0$, with jump distribution G and K jump times $\{T_1, \ldots, T_K\}$:

1. set the initial conditions $T_0 = 0$, $\lambda_{T_0^{\pm}} = \lambda_0 > a$, $N_0 = 0$ and $k \in \{0, 1, \ldots, K-1\}$;
2. simulate the $(k+1)$th inter-arrival time S_{k+1} by

$$S_{k+1} = \begin{cases} S^{(1)}_{k+1} \wedge S^{(2)}_{k+1}\,, & D_{k+1} > 0 \\ S^{(2)}_{k+1}\,, & D_{k+1} < 0 \end{cases},$$

where

$$D_{k+1} = 1 + \frac{\delta \ln U_1}{\lambda_{T_k^+} - a} \quad U_1 \sim U(0,1)\,,$$

and

$$S^{(1)}_{k+1} = -\frac{1}{\delta} \ln D_{k+1}\,, \quad S^{(2)}_{k+1} = -\frac{1}{a} \ln U_2\,, \quad U_2 \sim U(0,1)\,;$$

A fast simulation method

3. record the $(k+1)$th jump time

$$T_{k+1} = T_k + S_{k+1}\,;$$

4. record the change in the intensity at time T_{k+1} by

$$\lambda_{T_{k+1}^+} = \lambda_{T_{k+1}^-} + Y_{k+1}\,, \quad Y_{k+1} \sim G\,,$$

   where

$$\lambda_{T_{k+1}^-} = \left(\lambda_{T_k^+} - a\right) e^{-\delta(T_{k+1} - T_k)} + a\,;$$

5. record the change in the in process N_t by

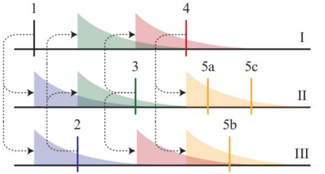
$$N_{T_{k+1}^+} = N_{T_{k+1}^-} + 1\,.$$

Proof.
On the whiteboard... □

Multivariate Hawkes process

Consider the D-dimensional point process $\{N_t^{[d]}\}_{d=1}^{D}$, where $N_t^{[d]} \equiv \{T_k^{[d]}\}_{k=1,2,\ldots}$ with the underlying intensity process

$$\lambda_t^{[d]} = a^{[d]} + \left(\lambda_0^{[d]} - a^{[d]}\right) e^{-\delta^{[d]} t} + \sum_{\ell=1}^{D} \sum_{T_k^{[\ell]} < t} Y_k^{[d,\ell]} e^{-\delta^{[d,\ell]}(t - T_k^{[\ell]})} .$$



Lemonnier et al., 2016

We call this process a *multivariate Hawkes process* or *mutually exciting Hawkes processes*. These can also be simulated using a similar algorithm but taking the next inter-arrival time as

$$S_{k+1} = \min\{S_{k+1}^{[1]}, \ldots, S_{k+1}^{[D]}\} .$$

Conditional intensity as hazard function

Given the history up until the last arrival u, $\mathcal{H}(u)$, define the conditional c.d.f. and p.d.f. of the next arrival time T_{k+1}

$$F(t|\mathcal{H}(u)) = \int_u^t \Pr(T_{k+1} \in [s, s+ds]|\mathcal{H}(u))ds = \int_u^t f(s|\mathcal{H}(u))ds\,.$$

The joint p.d.f. of a realization $\{t_1, t_2, \ldots, t_k\}$ is

$$f(t_1, t_2, \ldots, t_k) = \prod_{i=1}^{k} f(t_i|\mathcal{H}(t_{i-1}))\,.$$

The shorthand notations $f^*(t) = f(t|\mathcal{H}(u))$ and $F^*(t) = F(t|\mathcal{H}(u))$ are common. The conditional intensity can be characterized as the hazard function

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

Hawkes process likelihood

Theorem 3

Let $N(\cdot)$ be a regular point process on $[0, T]$ for some positive $T < \infty$ and let $t_1, t_2, \ldots, t_k$ denote a realization of $N(\cdot)$ over $[0, T]$. Then the likelihood L of $N(\cdot)$ is expressible as

$$L = \left(\prod_{i=1}^{k} \lambda(t_i) \right) \exp \left(- \int_0^T \lambda(t) dt \right) = \left(\prod_{i=1}^{k} \lambda(t_i) \right) e^{-\Lambda(T)} .$$

Hawkes process likelihood

Assume process observed to time of kth arrival. The joint density is

$$L = f(t_1, \ldots, t_k) = \prod_{i=1}^{k} f^*(t_i)\,.$$

Rearrange hazard function definition of $\lambda(t)$

$$\lambda(t) = \frac{f^*(t)}{1 - F^*(t)} = \frac{\frac{dF^*(t)}{dt}}{1 - F^*(t)} = -\frac{d\log(1 - F^*(t))}{dt}$$

and integrate both sides over the interval (t_k, t)

$$-\int_{t_k}^{t} \lambda(u)du = \log(1 - F^*(t)) - \log(1 - F^*(t_k))\,.$$

But $F^*(t_k) = 0$ because $T_{k+1} > t_k$, so

$$F^*(t) = 1 - \exp\left(-\int_{t_k}^{t} \lambda(u)du\right) \quad \text{and} \quad f^*(t) = \lambda(t)\exp\left(-\int_{t_k}^{t} \lambda(u)du\right)\,.$$

Hawkes process likelihood

So far, we have

$$L = f(t_1, \ldots, t_k) = \prod_{i=1}^{k} f^*(t_i)$$

and

$$F^*(t) = 1 - \exp\left(-\int_{t_k}^{t} \lambda(u)du\right) \quad \text{and} \quad f^*(t) = \lambda(t)\exp\left(-\int_{t_k}^{t} \lambda(u)du\right).$$

Now suppose the process is observed to some time $T > t_k$. Then the likelihood includes the probability of not observing anything on the interval $(t_k, T]$:

$$\begin{aligned} L &= (1 - F^*(T))\prod_{i=1}^{k} f^*(t_i) = \exp\left(-\int_{t_k}^{T} \lambda(u)du\right)\prod_{i=1}^{k} f^*(t_i) \\ &= \left(\prod_{i=1}^{k} \lambda(t_i)\exp\left(-\int_{t_{i-1}}^{t_i} \lambda(u)du\right)\right)\exp\left(-\int_{t_k}^{T} \lambda(u)du\right) \\ &= \left(\prod_{i=1}^{k} \lambda(t_i)\right) e^{-\Lambda(T)}. \end{aligned}$$

Exponential Hawkes process likelihood

The log-likelihood for the interval $[0, t_k]$ can be written

$$\ell = -\Lambda(t_k) + \sum_{i=1}^{k} \log(\lambda(t_i)) \,,$$

and $\Lambda(t_k)$ can be written

$$\Lambda(t_k) = \int_0^{t_1} \lambda(u) du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda(u) du \,.$$

For the exponential Hawkes process, this becomes

$$\begin{aligned}
\Lambda(t_k) &= \int_0^{t_1} \lambda du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda + \sum_{t_j < u} \alpha e^{-\beta(u-t_j)} du \\
&= \lambda t_k + \alpha \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^{i} e^{-\beta(u-t_j)} du \\
&= \lambda t_k + \alpha \sum_{i=1}^{k-1} \sum_{j=1}^{i} \int_{t_i}^{t_{i+1}} e^{-\beta(u-t_j)} du
\end{aligned}$$

Exponential Hawkes process likelihood

Continuing, we have

$$\begin{aligned}
\Lambda(t_k) &= \lambda t_k + \alpha \sum_{i=1}^{k-1} \sum_{j=1}^{i} \int_{t_i}^{t_{i+1}} e^{-\beta(u-t_j)} du \\
&= \lambda t_k - \frac{\alpha}{\beta} \sum_{i=1}^{k-1} \sum_{j=1}^{i} \left(e^{-\beta(t_{i+1}-t_j)} - e^{-\beta(t_i-t_j)} \right) \\
&= \lambda t_k - \frac{\alpha}{\beta} \sum_{i=1}^{k-1} \left(e^{-\beta(t_k-t_i)} - e^{-\beta(t_i-t_i)} \right) \\
&= \lambda t_k - \frac{\alpha}{\beta} \sum_{i=1}^{k-1} \left(e^{-\beta(t_k-t_i)} - 1 \right) .
\end{aligned}$$

Thus, the log-likelihood can be written

$$\ell = -\lambda t_k + \sum_{i=1}^{k} \left(\log \left(\lambda + \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)} \right) + \frac{\alpha}{\beta} \left(e^{-\beta(t_k-t_i)} - 1 \right) \right) .$$

Exponential Hawkes process likelihood

The log-likelihood

$$\ell = -\lambda t_k + \sum_{i=1}^{k} \left(\log \left(\lambda + \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i - t_j)} \right) + \frac{\alpha}{\beta} \left(e^{-\beta(t_k - t_i)} - 1 \right) \right) .$$

has quadratic computational complexity $O(k^2)$, but a recursion turns it into linear complexity $O(k)$. For $i = 2, \ldots, k$, let $A(i) = \sum_{j=1}^{i-1} e^{-\beta(t_i - t_j)}$. Then

$$\begin{aligned} A(i) &= e^{-\beta t_i + \beta t_{i-1}} \sum_{j=1}^{i-1} e^{-\beta t_{i-1} + \beta t_j} = e^{-\beta(t_i - t_{i-1})} \left(1 + \sum_{j=1}^{i-2} e^{-\beta(t_{i-1} - \beta t_j)} \right) \\ &= e^{-\beta(t_i - t_{i-1})} \left(1 + A(i-1)\right) . \end{aligned}$$

Letting $A(1) = 0$, the log-likelihood can be written

$$\ell = -\lambda t_k + \sum_{i=1}^{k} \left(\log\left(\lambda + \alpha A(i)\right) + \frac{\alpha}{\beta} \left(e^{-\beta(t_k - t_i)} - 1 \right) \right) .$$

Similar recursions are also available for the log-likelihood gradient.

Nonlinear Hawkes process

Letting $\Psi : \mathbb{R} \to \mathbb{R}_+$ be an arbitrary function, a nonlinear Hawkes process has conditional intensity function

$$\lambda(t) = \Psi\left(\int_0^t g(t-u)dN(u)\right) .$$

The linear Hawkes process is a special case. One may incorporate inhibition by specifying, say,

$$\lambda(t) = \exp\left(\mu t - \int_0^t g(t-u)dN(u)\right) .$$

Recently, autoregressive neural networks (RNN, transformer NN) have been used to model the function Ψ.

Spatiotemporal Hawkes processes

- Spatiotemporal Hawkes processes are useful for modeling earthquakes, wildfires, viral contagion and gun violence.
- The conditional intensity models the rate of events at locations $s \in X \subset \mathbb{R}^D$:

$$\lambda(s, t|\mathcal{H}(t)) = \lim_{\Delta s, \Delta t \to 0} \frac{E\left(N\left(B(s, \Delta s) \times [t, t + \Delta t)\right)|\mathcal{H}(t)\right)}{|B(s, \Delta s)|\Delta t} .$$

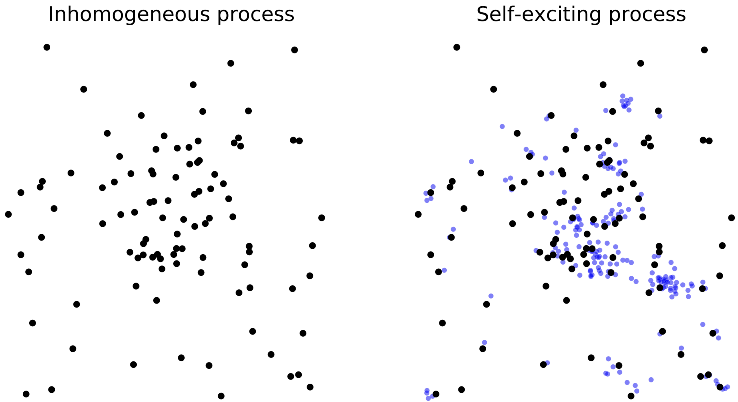
- For $\{(s_1, t_1), \ldots, (s_k, \ldots, t_k)$ observed, a spatiotemporal Hawkes process has conditional intensity

$$\lambda(s, t) = \mu(s) + \sum_{t_i < t} g(s - s_i, t - t_i) ,$$

  where the triggering function g is non-negative.
- The spatiotemporal HP also has a cluster process representation with mean number of children for each event

$$n = \int_X \int_0^T g(s, t) ds dt .$$

Spatiotemporal Hawkes processes



Reinhart, 2018

Spatiotemporal Hawkes process

The log-likelihood for the spatiotemporal Hawkes process takes a similar form to that of the temporal Hawkes process:

$$\ell = \sum_{i=1}^{k} \log\left(\lambda(s_i, t_i)\right) - \int_X \int_0^T \lambda(s,t) ds dt$$
$$= \sum_{i=1}^{k} \log\left(\mu(s) + \sum_{t_i < t} g(s_i - s_j, t_i - t_j)\right)$$
$$- \int_X \int_0^T \mu(s) + \sum_{t_j < t} g(s - s_j, t - t_j) ds dt \, .$$

The double summation in the first term implies an $O(k^2)$ computational complexity for k observations. No linear-time recursion is known. In EM/MCMC, it can be helpful to use the (still $O(k^2)$) 'complete data likelihood'

$$\ell_c = \sum_{i=1}^{k} 1\{u_i = 0\} \log\left(\mu(s_i)\right) + \sum_{i=1}^{k} \sum_{j=1}^{k} 1\{u_i = j\} \log(g(s_i - s_j, t_i - t_j))$$
$$- \int_X \int_0^T \lambda(s,t) ds dt \, .$$

Simulating the spatiotemporal Hawkes process

- Ogata (1998) proposes a two-stage algorithm that (1) simulates the temporal HP obtained by integrating over space and (2) simulates from the spatial inhomogeneous PP for each timepoint.

- Zhuang, Ogata and Vere-Jones (2004) propose a cluster based algorithm.

Simulating the spatiotemporal Hawkes process

Again, define the branching number $n = \int_X \int_0^T g(s,t)dsdt$. The following algorithm simulates a sample path of a spatiotemporal Hawkes process.

1. Generate events from the background process with intensity $\mu(s)$ using a method for simulating NHPP (e.g., thinning). Call events $G^{(0)}$;
2. let $l = 0$;
3. For each event $i \in G^{(l)}$, simulate the number of offspring $O_i^{(l)}$ $N^{(i)} \sim Pois(n)$ and the position/time of each with the normalized g as distribution.
4. Let $G^{(l+1)} = \bigcup_{i \in G^{(l)}} O_i^{(l)}$
5. If $G^{(l+1)}$ not empty, let $l = l + 1$ and return to step 3. Otherwise, return $\bigcup_{j=0}^{l} G^{(j)}$.