The Hawkes Process

Qualifying Examination Presentation

Ryan McCorvie

Updated: 2018/02/09

Berkeley
UNIVERSITY OF CALIFORNIA

Overview

1. Orientation
2. Point Processes
3. Hazard Rate Techniques
4. Branching Process Techniques
5. Martingale Techniques
6. Simulation
7. Extensions

Orientation

Point Process

A *point process* is just a random collection of points of some reference space S.

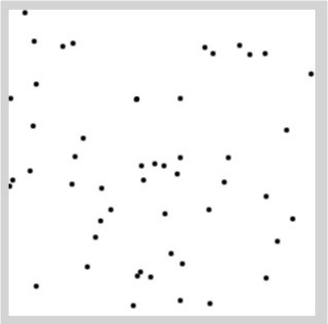
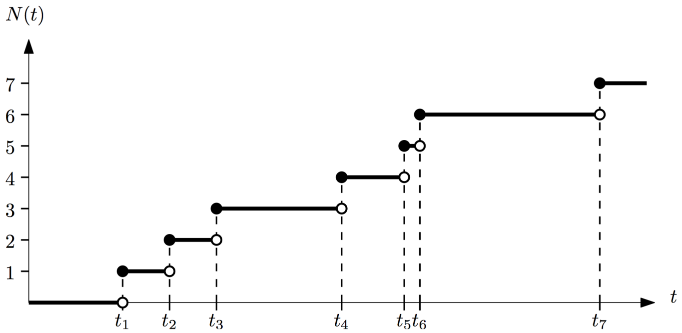


Figure: Realization of a Poisson point process on $[0, 1]^2$

Point Process on $\mathbb{R}_+$

On $\mathbb{R}_+$ we can characterize the process by:

- For measurable $A \subset \mathbb{R}_+$ the *counting measure* given by $N(A) = \#$ points in A
- The *counting function* $N_t = N([0, t])$
- The sequence of points $t_n = \inf\{t \in \mathbb{R}_+ : N_t \geq n\}$
- The interarrival times $\tau_n = t_n - t_{n-1}$ (taking $t_0 = 0$)



Hazard Rates

Informally, the *conditional hazard rate* or *conditional intensity* at time t is the probability per unit time that the next point is occurs at t, conditional surving until t. Let t^* be the time the next point occurs

$$\lambda(t)\Delta t \approx \mathbb{P}(t = t^* \mid N_s \text{ for } s < t \text{ and } t^* \geq t)$$

Conditional intensity

$$\begin{aligned}\lambda(t) &= \frac{\frac{d}{dt}\mathbb{P}(t^* \geq t \mid N_s \text{ for } s < t)}{\mathbb{P}(t^* \geq t \mid N_s \text{ for } s < t)} \\ &= -\frac{d}{dt}\log \mathbb{P}(t^* \geq t \mid N_s \text{ for } s < t)\end{aligned}$$

Hazard Rates

Examples

- The Poisson process has constant intensity λ.
- A one-point exponential process has constant intensity until the point occurs, when the intensity drops to 0.
- A renewal process has intensity $\lambda(t) = f(t - t_{n-1})$ where t_{n-1} is the last point before t, and f is a fixed function.

For a *regular* process satisfying $\mathbb{P}(\Delta N_t > 1) = o(\Delta t)$,

$$\lambda(t) = \frac{d}{dt}\,\mathbb{E}(N_t \mid N_t \text{ history})$$

Self-exciting Process

Given a *base rate* $\nu > 0$ and *excitation kernel* $\phi \in L^1(\mathbb{R})$ which is *non-negative*, $\phi \geq 0$, and *causal*, $\phi(u) = 0$ for $u < 0$.

Hawkes Process

Given base rate ν and excitation kernel ϕ, a *Hawkes process* is a point process with conditional intensity:

$$\begin{aligned}\lambda(t) &= \nu + \sum_{t_i \leq t} \phi(t - t_i) \\ &= \nu + \int_{-\infty}^{t} \phi(t - t')\, dN_{t'} \\ &= \nu + \phi * dN(t)\end{aligned}$$

Hawkes Intensity Sample

Exponential kernel

For some $\alpha, \beta > 0$ with $\alpha/\beta < 1$ let

$$\phi(u) = \alpha e^{-\beta u} \qquad \text{when } u \geq 0$$

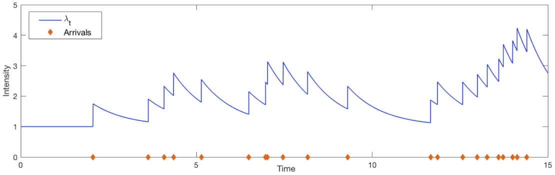


Figure: Sample path with exponential intensity

Clustering and Dispersion

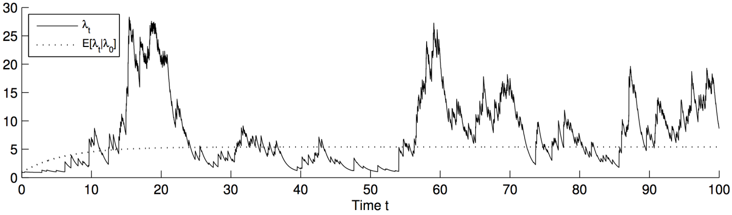


Figure: Sample path of conditional intensity

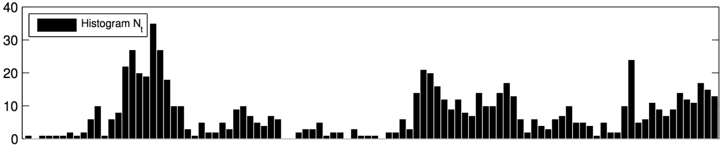


Figure: Points per interval of time

Applications

Table: Some applications of the Hawkes process model

Application	Authors	Date
Earthquakes	Ogata	1988
Neuron activity	Johnson	1996
Stock trading	Bowsher	2002
Corporate defaults	Errais, Giesecke, Goldberg	2010
Burglaries	Mohler et. al.	2011
Civilian deaths in Iraq	Lewis et. al.	2012
Online ad clickthrough	Xu, Duan, Whinston	2014

Point Processes

Formal Definition of a Point Process

Random Measure

Let $\mathcal{B}$ be the standard Borel σ-algebra on $\mathbb{R}$. A *random measure* ξ on $(\mathbb{R}, \mathcal{B})$ is a kernel from the basic probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to $\mathbb{R}$.

A *point process* N is a random measure whose values are in $\mathbb{N}$.

The measure must be finite on bounded sets almost surely.

Recall a kernel is a function $\xi : \Omega \times \mathcal{B} \to \mathbb{R}$ where $\xi(\omega, \cdot)$ is a *measure* on $\mathbb{R}$ for all $\omega \in \Omega$ and $\xi(\cdot, B)$ is *measurable* in Ω for all $B \in \mathcal{B}$.

Random Measures

Let $\mathscr{M}(\mathbb{R})$ be the set of all measures on $(\mathbb{R}, \mathscr{B})$.

Theorem: Random element of $\mathscr{M}(\mathbb{R})$

A random measure is a random element of $\mathscr{M}(\mathbb{R})$ endowed with the σ-field generated by the projections $\pi_B : \mu \mapsto \mu B$ for arbitrary $B \in \mathcal{B}$.

Lemma: Finite Dimensional Distributions

Let ξ, η be random measures, $\xi \sim \eta$ iff

$$(\xi B_1, \ldots, \xi B_n) \sim (\eta B_1, \ldots, \eta B_n)$$

for arbitrary $n \in \mathbb{N}$ and $B_1, \ldots, B_n \in \mathcal{B}$.

Factorial Process

Definition

Let $N = \sum_{k=1}^{\kappa} \delta_{x_k}$ be a point process on $(S, \mathcal{S})$. The nth *factorial process* $N^{(n)}$ on S^n is given by

$$N^{(n)} = \sum_{\substack{1 \le k_1, \ldots, k_n \le \kappa \\ k_i \text{distinct}}} \delta_{(s_{k_1}, \ldots, s_{k_n})}$$

When $A_1, \ldots, A_n$ are disjoint,

$$N^{(n)}(A_1 \times \cdots \times A_n) = N(A_1) \cdots N(A_n)$$

$$N^{(n)}(A^n) = N(A)\,(N(A) - 1) \cdots (N(A) - n + 1)$$

Moment Measures

Definitions

Let N be a point process on S and let $A \subset S^n$.

The nth *factorial moment measure* is $M_n = \mathbb{E}(N^{(n)})$

The nth *Janossy measure* is $J_n = \mathbb{E}(\mathbb{1}_{N(S)=n} N^{(n)})$

Examples

The expected number of points M_1 is sometimes called the *intensity*. The *pair correlation function* can be expressed

$$c(s, t) = M_2(ds, dt) - M_1(ds)M_1(dt)$$

When N is stationary, $M_1(ds) = m\,ds$ for some $m > 0$ and $c(x, y) = c(x - y)$ depends only on the separation.

Janossy Measures as Probability Measures

Lemma

For a partition $A_1, \ldots, A_n$ of S,

$$\mathbb{P}(NA_1 = n_1, \ldots, NA_k = n_k) = n_1! \cdots n_k! J_n(A_1^{n_1} \times \cdots \times A_k^{n_k})$$

Proof

Take expectations of both sides of the identity

$$\int_{A_1^{n_1} \times \cdots \times A_k^{n_k}} \mathbb{1}_{\{NS=n\}} dN^{(n)} = n_1! \cdots n_k! \mathbb{1}_{\{N(A_1)=n_1,\ldots,N(A_k)=n_k\}}$$

If J_n is absolutely continuous, its density is the *likelihood function*

$$\begin{aligned} J_n(ds_1, \ldots, ds_n) &= L(s_1, \ldots, s_n)\, ds_1 \cdots ds_n \\ &\approx \mathbb{P}\{N \text{ is exactly the points } s_k\ \} \end{aligned}$$

Hazard Rate Techniques

Likelihood and Hazard Rates

Theorem

Let N be a point process with conditional intensity λ. The likelihood of N on $[0,T]$ is given by

$$\log L(t_1,\ldots,t_n) = \sum_{i=1}^{n} \lambda(t_i) - \int_0^T \lambda(u)\,du$$

Proof

Let $p_n(t) = \mathbb{P}(t_n = t \mid t_1,\ldots,t_{n-1})$. Then $\lambda(t) = \frac{p_n(t)}{1-\int_{t_{n-1}}^{t} p_n(u)\,du}$ so $p_n(t) = \lambda(t)\exp(-\int_{t_{n-1}}^{t} \lambda(u)\,du)$. Finally then, $L = p_1(t_1)\cdots p_n(t_n)(1-\int_{t_n}^{T} p_{n+1}(u)\,du)$

For a Poisson process, the likelihood is $L_0(t_1,\ldots,t_n) = m^n e^{-mT}$

Maximum Likelihood Estimation

One can fit ϕ to realized data using the maximum likelihood estimate on L. Popular parameterizations are exponential-polynomial $\phi(t) = \sum_{k=1}^{n} a_k t^k e^{-\alpha t}$ and power-law $\phi(t) = \frac{K}{(c+t)^p}$

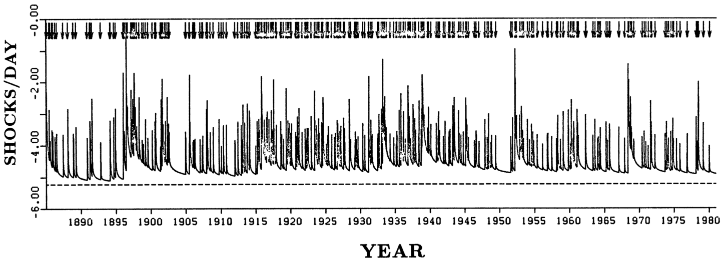


Figure: Fitted intensity for earthquakes near Tohoko Japan, Ogata 1988

Existence and Uniqueness

Theorem

There exists a unique point process N satisfying the Hawkes process condition

$$\lambda(t) = \nu + \int_{-\infty}^{t} \phi(t-u)\, dN_t$$

Proof

Existence: start with the Poisson process and introduce a new probability measure on $\mathscr{M}([0,T])$ using the likelihood ratios L/L_0 as the Radon-Nikodym derivative.

Uniqueness: the conditional intensity determines the likelihood which determines the Janossy measures which determines the finite dimensional distributions.

Unconditional Intensity

Let N be a stationary Hawkes process. Then $M_1(dt) = m\,dt$, and $m = \mathbb{E}(dN_t) = \mathbb{E}(\mathbb{E}(dN_t \mid \text{history})) = \mathbb{E}(\lambda(t))$ is the unconditional intensity.

$$\mathbb{E}(\lambda(t)) = \nu + \int_{-\infty}^{t} \phi(t-u)\ \mathbb{E}(dN_u)$$
$$m = \nu + \int_{-\infty}^{t} \phi(t-u)\, m\, du$$
$$m = \frac{\nu}{1 - \int_0^\infty \phi(u)\, du}$$

Necessary conditions: $\nu > 0$ and $r := \int_0^\infty \phi(u)\, du < 1$. Call r the *branching ratio*.

Correlation function

For a stationary, regular process

$$
\begin{aligned}
M_2(ds, dt) &= \mathbb{E}(dN_s\, dN_t) \\
&= m\delta_{t-s}\, ds + (c(t-s) + m^2) ds\, dt
\end{aligned}
$$

Conditioning with $s < t$ we find

$$
\mathbb{E}(dN_s\, dN_t) = \mathbb{E}\big(dN_s\ \mathbb{E}(dN_t \mid N_t \text{ history})\big) = \mathbb{E}(dN_s\, \lambda(t)\, dt)
$$

For $h > 0$ let $t = s + h$

$$
\begin{aligned}
c(h) &= \nu\, \mathbb{E}(dN_s) + \int_{-\infty}^{s+h} \phi(s+h-u)\, \mathbb{E}(dN_s\, dN_u) - m^2 \\
&= m\phi(h) + \int_{\infty}^{h} \phi(h-u) c(u)\, du
\end{aligned}
$$

This shows c satisfies a Wiener-Hopf equation, and so m and c uniquely specify ν and ϕ.

Exponential Kernel

Take $\phi(t) = \alpha e^{-\beta t}$ and let $\widehat{c}$ be the Laplace transform of c.

Taking transforms of the last equation

$$\widehat{c}(s) = m\widehat{\phi}(s) + \widehat{\phi}(s)(\widehat{c}(s) + \widehat{c}(\beta))$$

Solving for $\widehat{c}$

$$\widehat{c}(s) = \frac{\alpha m(2\beta - \alpha)}{2(\beta - \alpha)(s + \beta - \alpha)}$$

We recognize this as the transform of

$$c(u) = Ke^{-(\beta-\alpha)|u|} \qquad \text{with} \qquad K = \frac{\alpha m(2\beta - \alpha)}{2(\beta - \alpha)}$$

Exponential Kernel is Markov

For an exponential kernel, rewrite the intensity equation

$$\lambda(t+h) = \nu + e^{-\beta h}(\lambda(t) - \nu) + \int_t^{t+h} \alpha e^{-\beta(t+h-u)}\, dN_u$$

In the limit $h \to 0$

$$d\lambda = -\beta(\lambda - \nu)\, dt + \alpha\, dN_t$$

To go the other way, use Ito's formula.

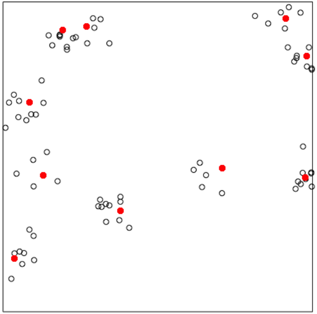
This shows the distribution of $(d\lambda(t), dN_t)$ is determined entirely by $(\lambda(t), N_t)$ so (λ, N) is Markov.

Branching Process Techniques

Clusters

Cluster Process

Let the *center* be a point process Z and let the *clusters* be a family of independent point processes $C_{(u)}$ indexed by $\mathbb{R}$. For a realization $Z = \sum_{i=1}^{\kappa} \delta_{t_i}$ a realization of the *cluster process* N is $N = Z \circ C_{(u)} = \sum_{i=1}^{\kappa} C_{(t_i)}$



Branching Process

Branching Process

Given an *immigrant* process Z and *branching processes* $C_{(u)}$, inductively define

$$N_0 = Z, \quad N_1 = N_0 \circ C_{(u)}, \quad N_2 = N_1 \circ C_{(u)}, \quad \ldots$$

A *branching process* is given by

$$\begin{aligned} N &= N_0 + N_1 + N_2 + \ldots \\ &= Z + Z \circ C_{(u)} + Z \circ C_{(u)} \circ C_{(u)} + \ldots \end{aligned}$$

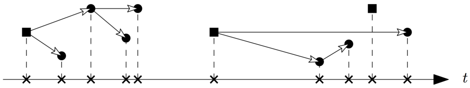


Figure: Branching structure and realized points

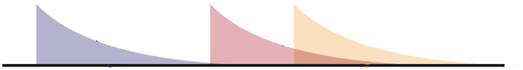
Branching Representation of Hawkes Process

Theorem

A Hawkes process is a branching process whose immigrants Z are a Poisson process with constant intensity ν and whose branching process $C_{(u)}$ is a Poisson process with intensity $\phi(\cdot - u)$.

Proof

The sum of independent Poisson processes is Poisson with a summed intensity. Since ϕ is causal, a point in $(t, t + dt)$ is an immigrant or the descendent of points prior to t. The summed intensities are $\lambda(t) = \nu + \sum_{t_t \leq t} \phi(t - t_i)$ as desired.



Stability of Hawkes

Theorem

Let N be a Hawkes process. $N([a,b]) < \infty$ almost surely for all $[a,b]$

Proof

Let $D_{(u)}$ be all decendents from a point at u. The number of points is a Galton-Watson, so $\mathbb{E}(D_{(u)}(\mathbb{R})) = \frac{1}{1-r} < \infty$. Let $p_t(a,b) = \mathbb{P}(D_{(t)}[a,b] > 1)$ and let $S = \sum_{t_i \in Z} p_{t_i}(a,b)$.

$$\begin{aligned} \mathbb{E}\, S &= \int_{\mathbb{R}} \nu p_t(a,b)\, dt \leq \int_{\mathbb{R}} \nu\, \mathbb{E}\, D_{(t)}([a,b]) \\ &= \nu \int_a^b \mathbb{E}\, D_{(t)}(\mathbb{R}) = \frac{\nu(b-a)}{1-r} < \infty \end{aligned}$$

If $S < \infty$, then by Borel-Cantelli converse, only finitely many t_i have descendents in $[a,b]$. Since $N = Z \circ D_{(u)}$, we're done.

EM Estimation

The idea is to use the EM algorithm to determine the hidden branching structure. Let p_{0i} be the probability event i is an immigrant and p_{ij} the probability that j is a direct descendent of i. Suppose ϕ and ν are parameterized by $\hat{\theta}$.

- **E-step** Estimate the p_{ij} by

$$p_{0j} = \frac{\nu}{\nu + \sum_{i<j} \phi(t_j - t_i)} \qquad p_{ij} = \frac{\phi(t_j - t_i)}{\nu + \sum_{i<j} \phi(t_j - t_i)}$$

- **M-step** Choose the parameters $\hat{\theta}$ to maximize

$$\sum_{i=0}^{n} \sum_{j=i+1}^{n} p_{ij} \log L(i \to j \mid \hat{\theta})$$

Functionals

Probability Generating Functional (PGFL)

Let $N = \sum_{k=1}^{\kappa} \delta_{t_i}$ and measurable $f : \mathbb{R} \to (0, 1]$. The *probability generating functional* is

$$G_N[f] = \mathbb{E} \prod_{i=1}^{\kappa} f(t_i) = \mathbb{E}\left(\exp \int_{\mathbb{R}} \log f \, dN\right)$$

For a Poisson process intensity measure μ the pgfl is $\exp \int_{\mathbb{R}} (f - 1) \, d\mu$. The PGFL uniquely determines the law of N.

Branching and Cluster Process

Let N be a cluster process and let $G_C[f \mid s]$ the PGFL for $C_{(s)}$.
Let $C_{(s)} = \sum_{i=1}^{\kappa_s} \delta_{t_i^{(s)}}$ and let $N = \sum_{i=1}^{\kappa} s_i$

$$
\begin{aligned}
G_N[f] &= \mathbb{E}\,\mathbb{E}\left(\prod_{i=1}^{\kappa} \prod_{j=1}^{\kappa_{s_i}} f(t_j^{(s_i)}) \,\middle|\, N \right) \\
&= \mathbb{E} \prod_{i=1}^{\kappa} G_C[f \mid s_i] \\
&= G_Z[G_C[f \mid \cdot]]
\end{aligned}
$$

Let $H[f \mid s]$ be the PGFL for the descendents starting at s, and $H_n[f \mid s]$ for the first n generations

$$
H_n[f \mid s] = f(s) G_C[H_{n-1}[f \mid \cdot] \mid s]
$$
$$
H[f \mid s] = f(s) G_C[H[f \mid \cdot] \mid s]
$$

Functional Equations

The PGFL for the Hawkes process is $G[f]$ where

$$
\begin{aligned}
G[f] &= \exp \int_{\mathbb{R}} (H[f(\cdot - t)] - 1)\nu \, dt \\
H_{n+1}[f] &= f(0) \exp \int_0^\infty (H_n[f(\cdot - t)] - 1)\phi(t) \, dt \\
H[f] &= \lim_{n\to\infty} H_n[f]
\end{aligned}
$$

Judiciously choosing f we get functional equations for interesting quantities. For example let $D(s)$ be distribution for the cluster length

$$
D(x) = \begin{cases} \exp(-r + \int_0^x D(x-u)\phi(u) \, du) & x \geq 0 \\ 0 & x < 0 \end{cases}
$$

Expansions of the PGFL

We can expand the PGFL in terms of moment measures

$$G[f] = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^n} \prod_{i=1}^{n} f(t_i) J_n(dt_1, \ldots, dt_n)$$

$$G[1-g] = 1 + \sum_{n=0}^{\infty} (-1)^n \frac{1}{n!} \int_{\mathbb{R}^n} \prod g(s_i) M_n(ds_1, \ldots, ds_n)$$

From this and the functional equation we get functional equations for the moments

$$M_1(A) = \nu \delta_0(A) + \nu \int M_1(A-s)\, ds$$

$$M_2(A \times B) = M_1(A) M_1(B) + \int_{\mathbb{R}} M_2(A-s, B-s) \phi(s)\, ds - \delta_0(A)\delta_0(B)$$

Martingale Techniques

Compensator

Martingale representation

Let $\mathcal{H}_t$ be a filtration representing the history of a point process N_t. There is a previsible, monotonic process A_t adapted to $\mathcal{H}_t$ called the *compensator* such that $Y_t = N_t - A_t$ is a martingale. For a Hawkes process, $Y_t = N_t - \int_0^t \lambda(u)\,du$

Proof

Let t_n be the stopping time for the nth point, and let

$$A_t = \sum_{n=1}^{\infty} \int_0^{(t_n - t_{n-1}) \vee (t - t_{n-1})} \frac{G_n(du \mid \mathcal{H}_{n-1})}{1 - G_n(u- \mid \mathcal{H}_{n-1})} \mathbb{1}_{t \geq t_{n-1}}$$

At this point its a calculation to verify $N_t - A_t$ is a martingale, and previsible, and that the nth term is $\int_{t \vee t_{n-1}}^{t \vee t_n} \lambda(t)\,dt$

Prediction

Every $\mathcal{H}_t$ adapated martingale can be written $\int_0^t h(u)\,dY_u$ for some previsible h. We can sometimes use this to build "filters" for interesting quantities.

Let $\psi(t)$ satisfy $\psi(t) = \phi(t) + \phi * \psi(t)$. Then it can be shown

$$\lambda(t) = \nu + \int_0^t \psi(t-u)\nu\,du + \int_0^t \psi(t-u)\,dY_u$$

So, for example

$$\mathbb{E}(\lambda(t) \mid \mathcal{H}_s) = \nu + \int_0^t \psi(t-u)\nu\,du + \int_0^s \psi(t-u)\,dY_u$$

Simulation

Random Time Change

Theorem: Simple Process with Continuous Compensator

Let N be a $\mathcal{H}_t$ adapted point process with compensator $\Lambda(t) = \int_0^t \lambda(u)\,du$. Under random time change $t \mapsto \Lambda(t)$

$$\tilde{N}(t) = N(\Lambda^{-1}(t))$$

is Poisson with unit rate.

Conversely if A is a.s. finite, continuous, monotonically increasing $\mathcal{H}_t$ adapted random process, and $\tilde{N}$ is a unit Poisson process, then $N(t) = \tilde{N}(A(t))$ has compensator $M(t)$.

To generate point t_n given the history $\mathcal{H}_t$, generate $E \leftarrow \mathrm{Exp}(1)$, and let t_n solve $E = \int_{t_{n-1}}^{t_n} \lambda(u)\,du$

Thinning

Key observation: a 1d Poisson process with variable intensity can be simulated with a 2d Poisson process with constant intensity, rejecting points above the graph of the intensity

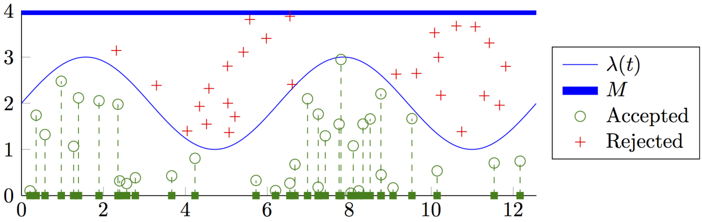


Figure: Thinning a Poisson process

Ogata's modified algorithm

Set $P = \{\}$, $t \leftarrow 0$

while $t < T$ **do**

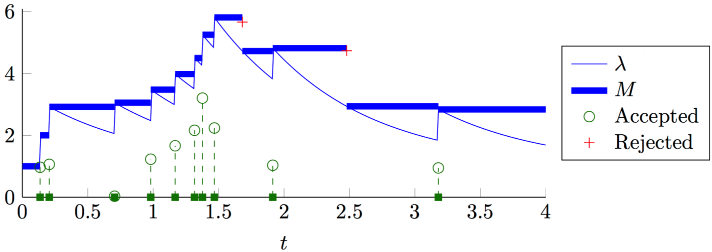
$M \leftarrow \lambda(t + \epsilon)$

Generate $E \leftarrow \mathrm{Exp}(M)$, $U \leftarrow \mathrm{Unif}(0, M)$

$t \leftarrow t + E$

If $U \leq \lambda(t)$ then $P \leftarrow P \cup \{t\}$

end while



Extensions

Multivariate Hawkes

The *multivariate Hawkes process* is a point process on $\mathbb{R} \times \{1, \ldots, n\}$ with base rate $\vec{\nu}$ and excitation kernel matrix $\Phi(t) = (\phi_{ij}(t))$. The ϕ_{ij} must be non-negative and causal. The intensity dynamics are

$$\lambda_i(t) = \nu_i + \sum_{j=1}^{n} \int_{-\infty}^{t} \phi_{ij}(t-u)\, dN_j(t)$$

- The process is stable when $\left\| \int_0^\infty \Phi(t) \right\| < 1$
- Can be used to model dependence in a network

Other extensions

- **Marked point process**: Each point t_i is associated with an i.i.d. random variable $Y_i \geq 0$ and the intensity formula is

$$\lambda(t) = \nu + \sum_{t_i \leq t} Y_i \phi(t - t_i)$$

The Y_i are *marks*, and represent magnitude or impact of a point.

- **Non-linear Hawkes**: The intensity dynamics are given by

$$\lambda(t) = h\left(\nu + \sum_{t_i \leq t} \phi(t - t_i)\right)$$

Common choices for h are $h(x) = \max(0, x)$ and $h(x) = \exp(x)$. Here ϕ can be negative, so events can display inhibitory behavior.

References

Alan G. Hawkes
Spectra of Some Self-exciting and Mutually Exciting Point Processes
Biometrika, 1971

Daryl Daley and David Vere-Jones
An Introduction to the Theory of Point Processes
Springer, 2011

THE

END