Hawkes Processes with Stochastic Excitations

Young Lee*, Kar Wai Lim†, Cheng Soon Ong†

* National ICT of Australia & London School of Economics

†National ICT of Australia & Australian National University

Outline

Background

- Simple point processes:
  - $(T_i)_i$ a sequence of non-negative random variables such that $T_i < T_{i+1}$. Also known as random times.
- Counting processes:
  - Given simple point process $(T_i)_i$

$$N(t) = \sum_{i>0} 1_{T_i \leq t}$$

    is called the counting process associated with T.
- Interarrival times:
  - The process Δ defined by

$$\Delta_i = T_i - T_{i-1}$$

    is called the interarrival times associated with T.
- Intensity process: The intensity process is defined as

$$\lambda(t) = \lim_{h \to 0} \frac{1}{h} E[N(t+h) - N(t) | \mathcal{F}_t]$$

Recap: Poisson → Hawkes → Stochastic Hawkes

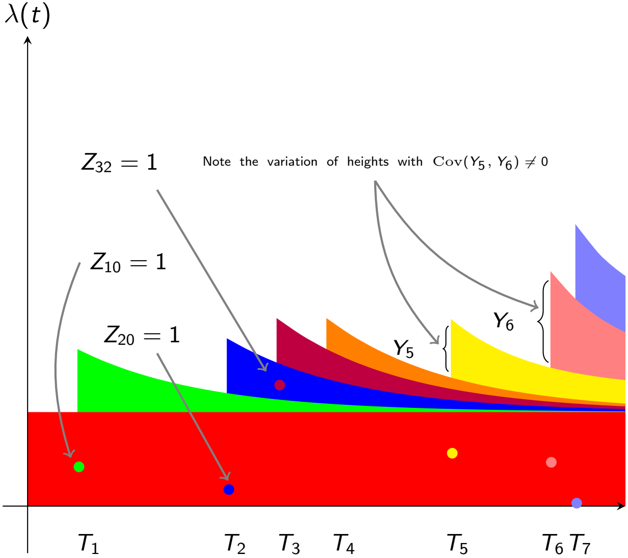
- N_t as the number of arrivals or events of the process by time t.
- $\lambda = const.(Poisson)$, does not take the history of events into account. However, if an arrival causes the intensity function to increase then the process is said to be self-exciting (Hawkes Process).
- Hawkes flavour:

$$\lambda(t) = \hat{\lambda}_0(t) + \sum_{i:t>T_i} Y(T_i)\,\nu(t - T_i), \tag{1}$$

where the function ν takes the form $\nu(z) = e^{-\delta z}$.

- $\exists$ different formulations for Y
  1. Constant, Hawkes (1971), Hawkes & Oakes (1974)
  2. Random excitations, Brémaud & Massoulié (2002), Dassios & Zhao (2013),
  3. Stochastic differential equations.

Illustration of Stochastic Hawkes



Our model

- The intensity function

$$\lambda(t) = \underbrace{\hat{\lambda}_0(t)}_{Base\ intensity} + \sum_{i:t>T_i} \underbrace{Y(T_i)}_{Contagion\ process\,/\,Levels\ of\ excitation} \nu(t - T_i)$$

where $\hat{\lambda}_0 : \mathbb{R} \mapsto \mathbb{R}_+$ is a deterministic base intensity, Y is a stochastic process and $\nu : \mathbb{R}_+ \mapsto \mathbb{R}_+$ conveys the positive influence of the past events T_i on the current value of the intensity process.

- Base intensity $\hat{\lambda}_0$
- Contagion process / Levels of excitation $(Y_i)_{i=1,2,..,N_T}$ measure the impact of clustering of the event times
- We take ν to be the exponential kernel of the form $\nu(t) = e^{-\delta t}$.

Stochastic differential equations to describe evolution of Y

- Changes in the levels of excitation Y is assumed to satisfy

$$Y_\cdot = \int_0^\cdot \hat{\mu}(t, Y_t)dt + \int_0^\cdot \hat{\sigma}(t, Y_t)dB_t$$

  where B is a standard Brownian motion and $t \in [0, T]$ where $T < \infty$.

- **Standing assumption:**

$$Y_t > 0, \quad \forall t \geq 0.$$

- Geometric Brownian Motion (GBM):
- Exponential Langevin:

Two representations for Stochastic Hawkes

- Intensity based.

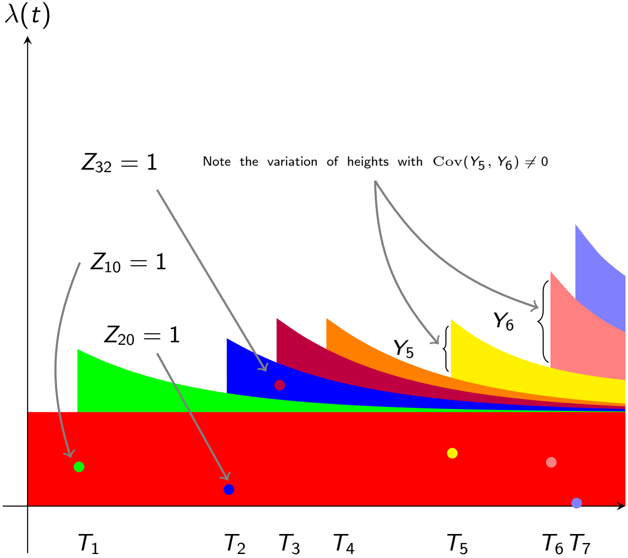
$$\lambda_t = a + (\lambda_0 - a)e^{-\delta t} + \sum_{i:\, T_i < t}^{N_t} Y_i \, e^{-\delta(t - T_i)} \tag{2}$$

- Cluster based. Immigrants and offsprings. We say an event time T_i is an
  1. *immigrant* if it is generated from the base intensity $a + (\lambda_0 - a)e^{-\delta t}$, otherwise
  2. we say T_i is an *offspring*.

It is natural to introduce a variable that describes the specific process to which each event time T_i corresponds to.

- $Z_{i0} = 1$ if event i is an immigrant,
- $Z_{ij} = 1$ if event i is an offspring of j

Quick recap - Stochastic Hawkes



Outline

Simulation & Inference

- Simulation framework of Dassios & Zhao (2011) is adopted,
- Decompose the inter-arrival event times into two independent simpler random variables: $S^{(1)}, S^{(2)}$; S_{j+1} is the inter-arrival time for the $(j+1)$-th jump:

$$S_{j+1} = T_{j+1} - T_j \, .$$

Given the intensity function, we can derive the cumulative density function for S_{j+1} as

$$F_{S_{j+1}}(s) = 1 - \exp\left(-\left(\lambda_{T_j^+} - a\right) \frac{1 - e^{-\delta s}}{\delta} - as \right) \quad .$$

Decompose S_{j+1} into $S_{j+1}^{(1)}$ and $S_{j+1}^{(2)}$:

$$\begin{aligned} \mathbb{P}(S_{j+1} > s) &= \exp\left(-\left(\lambda_{T_j^+} - a\right) \frac{1 - e^{-\delta s}}{\delta} \right) \times e^{-as} \\ &= \mathbb{P}\Big(S_{j+1}^{(1)} > s\Big) \times \mathbb{P}\Big(S_{j+1}^{(2)} > s\Big) \\ &= \mathbb{P}\Big(\min\big(S_{j+1}^{(1)}, S_{j+1}^{(2)}\big) > s\Big). \end{aligned}$$

Simulation & Inference

$$F_{S^{(1)}_{j+1}}(s) = \mathbb{P}\Big(S^{(1)}_{j+1} \le s\Big) = 1 - \exp\Big(-\big(\lambda_{T_j^+} - a\big)\,\frac{1-e^{-\delta s}}{\delta}\Big),$$
$$F_{S^{(2)}_{j+1}}(s) = \mathbb{P}\Big(S^{(2)}_{j+1} \le s\Big) = 1 - e^{-as}.$$

for $0 \le s < \infty$. To simulate S_{j+1}, we simply need to independently simulate both $S^{(1)}_{j+1}$ and $S^{(2)}_{j+1}$. Simulating $S^{(2)}_{j+1}$ is trivial since $S^{(2)}_{j+1}$ follows an exponential distribution with rate parameter a. To simulate $S^{(1)}_{j+1}$, we use the inverse CDF approach:

$$S^{*}_{j+1} = -\frac{1}{\delta}\ln\left(1 + \frac{\delta\ln(v)}{\lambda_{T_j^+} - a}\right) \qquad \text{if } \exp\left(-\frac{\lambda_{T_j^+} - a}{\delta}\right) \le v < 1,$$

we discard S^{*}_{j+1} otherwise, that is, $v < \exp\Big(-\frac{\lambda_{T_j^+}-a}{\delta}\Big)$ (this corresponds to the defective part), where v is simulated from a standard uniform distribution $V \sim U(0,1)$.

Simulation & Inference

Inference - Hybrid of MH and Gibbs

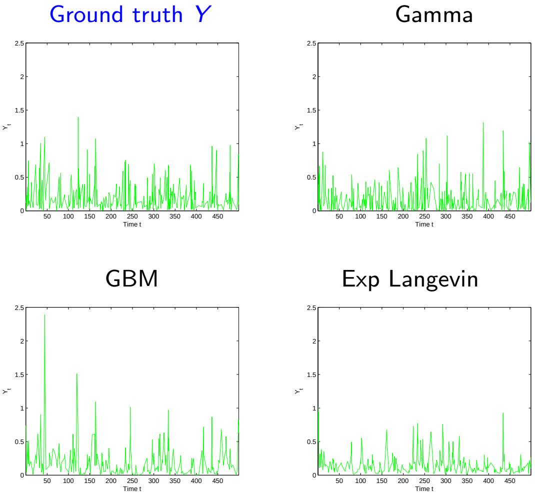
- The employment of branching representation enables the use of Gibbs sampling to learn Z,μ and σ,
- Other parameters a, λ_0, k and Y are learned with the vanilla MH algorithm.

Outline

Synthetic validation

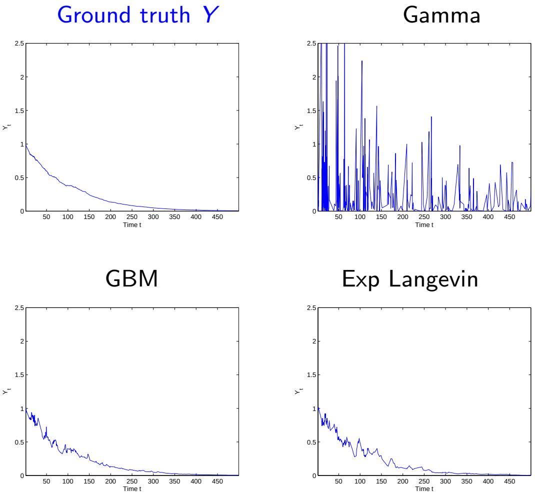
- Inference algorithm is first tested on synthetic data generated from Stochastic Hawkes
- Event times are generated assuming Y follows iid Gamma, GBM or Exponential Langevin,
- Performing experiments to recalibrate the parameters and subsequently sample the posterior Y gives the following interesting results

Inference learns Gamma ground truth



- All seems good.

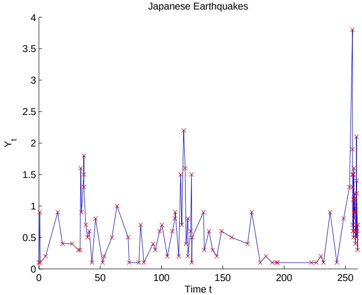
Inference learns G.B.M.

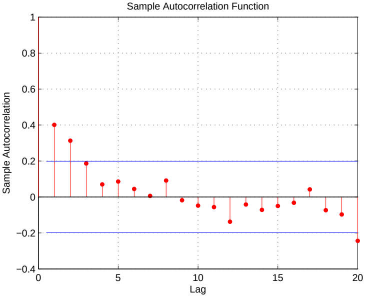


- iid Gamma fails, but a posteriori trying to capture a downward trend.
- GBM learns well. Exp Langevin too!!

Japanese Earthquakes Data (Di Giacomo et. al 2015)

- Plot of Y vs time:

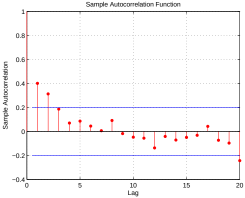




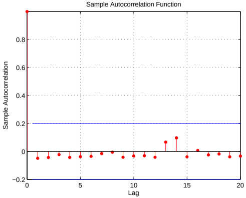
- Y might not be iid as earthquake occurrence tend to be correlated.
- Geophysical TS are frequently autocorrelated because of inertia or carryover processes in physical system.
- Autocorrelations should be near-zero for randomness, else will be significantly non-zero

Autocorrelation functions - SDEs retrieve correlated Y

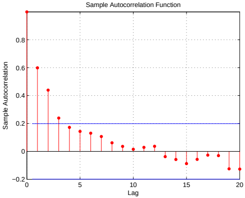
Ground truth Y



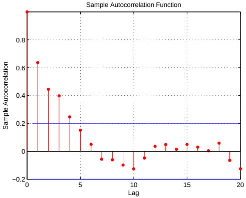
Gamma



GBM



Exp Langevin



Prediction - Stochastic Hawkes performs reasonable well

TABLE: Prediction of number of Earthquakes on Test Set. Result is averaged over 5 runs.

Model	Predicted	Observed	Diff
Poisson Process	62.80 ± 0.00	73.00	-10.20 ± 0.00
Classical Hawkes	61.13 ± 2.80	73.00	-11.87 ± 2.80
Stochastic Hawkes (GBM)	64.38 ± 6.82	73.00	-8.62 ± 6.82
Stochastic Hawkes (Langevin)	63.54 ± 4.09	73.00	-9.46 ± 4.09

Outline

Summary

- Motivation for Stochastic Hawkes

$$\lambda_t = a + (\lambda_0 - a)e^{-\delta t} + \sum_{i\colon T_i < t}^{N_t} Y_i \, e^{-\delta(t - T_i)}$$

1. Constant
2. Independent and identically distributed
3. Stochastic differential equations

- Simulation and Inference - with Z
- Experiments - Synthetic / Earthquake
- Poster #32, 3pm - 7pm later today