Handling Missing Data in Self-Exciting Temporal Point Processes

J. Derek Tucker, Jonathan Lane, John Lewis, and Kathy Simonson

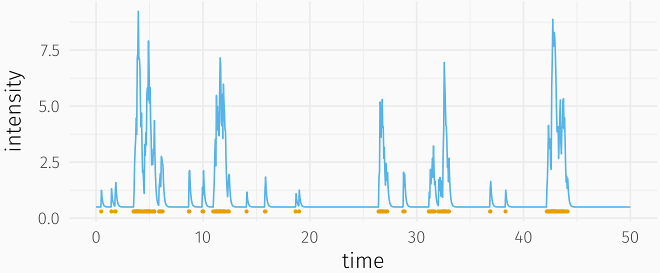
July 31, 2017

Department of Statistical Sciences Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. SAND NO. SAND2017-4379 C

Introduction

- For many critical intelligence and surveillance applications a high degree of human interaction is required to interpret events
- Thorough exploitation remains elusive due to ever-increasing volume of data and points in time where data is missing
- Our question is can we model and correlate events that happen in a self-exciting process where the history has missing data
- Our method is to develop missing data estimation for the Hawkes process using Bayesian methods



Temporal Hawkes Process

- A temporal point process $N(t)$ is characterized by its conditional intensity $\lambda(t)$

$$\lambda(t) = \lim_{\Delta t \downarrow 0} (E[N\{(t, t + \Delta t)\}|\mathcal{H}_t]/(\Delta t))$$

- Conditional intensity often takes the form

$$\lambda(t) = \mu + \alpha \sum_{k: t_k < t} g(t - t_k)$$

- where $g(t) = \beta \exp(-\beta(t))$
- Assume we have observed points given by $x = (t_1, \ldots, t_n)$ on $[0, T)$ for some fixed time $T > 0$, the likelihood is

$$p(x|\phi) = \left(\prod_{i=1}^{n} \lambda(t_i|\mathcal{H}_{t_i}) \right) \exp(-\Lambda^*(T)),$$

- where
$\Lambda^*(t) = \int_0^t \lambda^*(s|\mathcal{H}_s)\, ds = M(t) + \alpha \sum_{k: t_k < t} G(t - t_k), \quad M(t) = \int_0^t \mu(s)\, ds$

Branching Structure

- Another view of the Hawkes process is to define it as a Poisson cluster process:

Branching Structure

1. The parents I follow a Poisson process with intensity μ
2. Each parent $t_i \in I$ generates a cluster C_i, where the clusters are assumed to be independent
3. A cluster C_i consists of points of offspring with the following structure: Generation 0 consists of the parents. Recursively, each generation l generates points of generation $l+1$, where the offspring are generated as a Poisson process O_j with intensity function $g(t-t_j)$.
4. The process, X is the union of all the clusters.

- The relationship of all points and their parents is known as the branching structure
- Denote by $Y = \{y_i\}$, where $y_i = 0$ means t_i is a parent point and $y_i = j$ means t_i was an offspring of the point t_j

Bayesian Parameter Estimation

- We will estimate using a Bayesian approach similar to (Rasmussen 2013) using a Metropolis-with-in-Gibbs approach
- The branching structure will be used to facilitate more computationally efficient sampling and is estimated along with parameters (Ross 2016)
- For example, to sample a new value of $\alpha^{(k+1)}$ from $p(\alpha|x, \phi, Y)$ where

$$p(\alpha|x, \phi, Y) \propto \pi(\alpha) \prod_{i=1}^{n} \exp(-\alpha G(T - t_i|\beta))\alpha^{|S_i|}.$$

- The branching structure is sampled using the context of stochastic declustering (Zhuang, Ogata, and Vere-Jones 2002) where the probabilities of a point being a parent or offspring of the process are

$$p(Y_i^{(k+1)} = j|x, \phi) = \begin{cases} \frac{\mu}{\lambda(t)} & \text{if } j = 0 \\ \frac{\alpha g(t_j)}{\lambda(t)} & \text{if } j \in 1, 2, \ldots, i-1. \end{cases}$$

Bayesian Missing Data Estimation

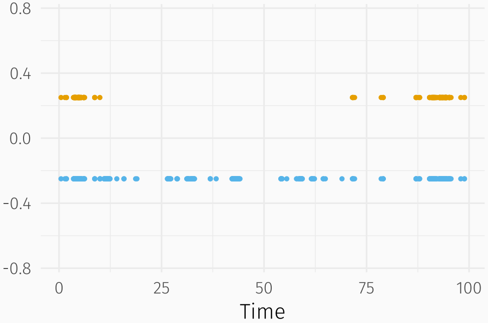
- Assume we have an interval, $[T_1, T_2]$ in which the observed points x are missing on $[0, T)$
- Estimate missing data as part of Metropolis-Hastings Algorithm
- Will propose samples using the conditional intensity, current parameters (ϕ), and history (x) up to time T_1
- Simulated using the thinning method developed by (Ogata 1981)
- The Metropolis ratio for this missing data is

$$H_t = \frac{p(\tilde{x}|\phi)}{p(x|\phi)} \frac{p(x_{T_2}|\phi)}{p(\tilde{x}_{T_2}|\phi)}$$

- where t_{miss} and $\tilde{t}_{miss}$ be the current and proposed set of missing data and $x = (t_{miss}, t_{obs})$ and $\tilde{x} = (\tilde{t}_{miss}, t_{obs})$

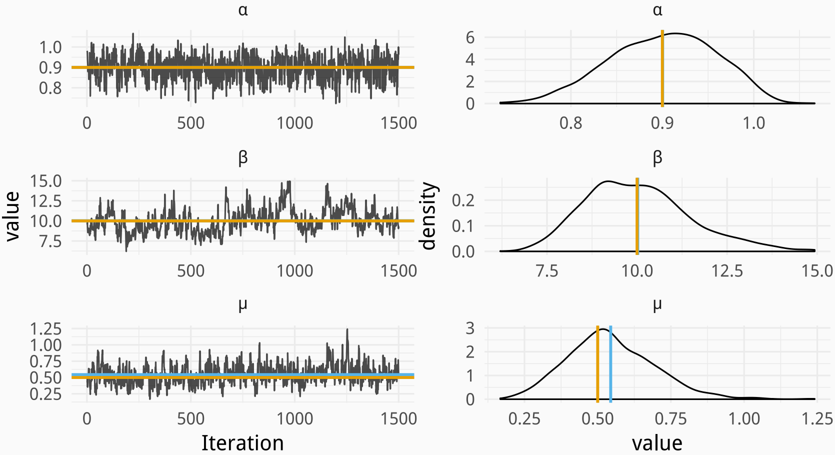
Simulations

- We generated a simulation of data from the model using the parameters $\mu = .5$, $\alpha = 0.9$, and $\beta = 10$.
  - The temporal region of interest was $T = 100$s



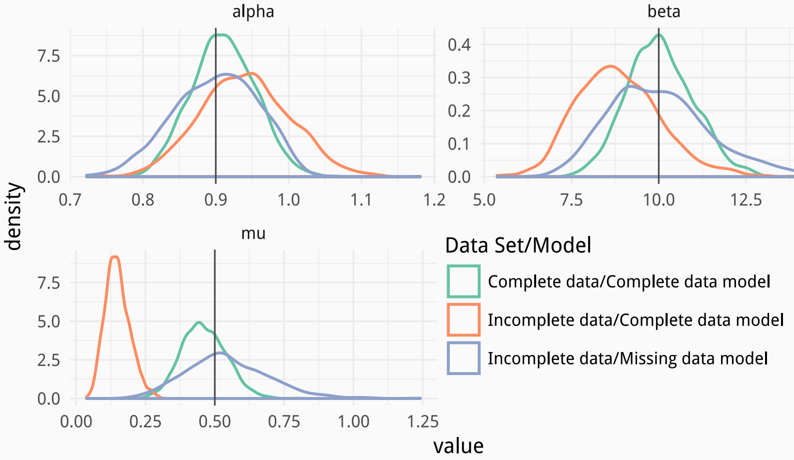
- Arrival times for full data (blue) and arrival times with missing data (orange)

Traceplots



- The true values are shown by the orange line and the estimates by the blue line
- Estimated parameters using the incomplete data are different and μ is greatly underestimated

Parameter Estimates Handling Missing Data

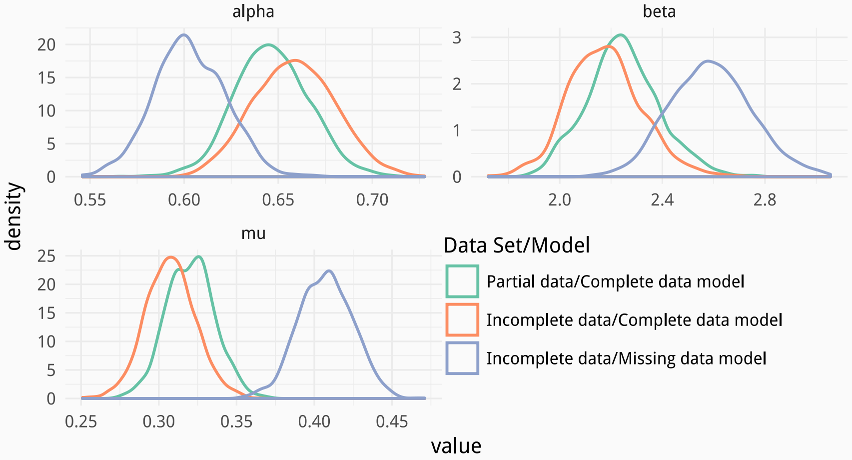


- Accounting for the missing data we are able to recover closely the true parameter
- However, the uncertainty is larger which relates to the fact we estimated the missing data

Global Terrorism Database

- The Global Terrorism Database (2017) (GTD) is an open-source database including information on terrorism events around the world from 1970-2015
- For our study, we will look at the year span between 1990-1997 in Columbia which had multiple problems with guerrillas, paramilitaries, and narcotics
- The database is missing records for the entire year of 1993

Golobal Terrorism Database



- Parameter estimates accounting for missing data increase, number of estimated events on order of those recovered in the data set
  - Not all data has been recovered
  - Working out prediction currently

Conclusions

- Developed a parameter estimation procedure for the Hawke's process that handles intervals of missing data in the history
    - Used MCMC to estimate the missing data, branching structure, and parameter estimates
    - Utilizing the branching structure accounts for more efficient estimation
- Demonstrated the capability on simulated data and the Global Terrorism Database
    - Captured true parameter values with a larger confidence region

Future Work

- Developing prediction comparison capabilities
- More efficient sampling schemes
    - Only relied on a Metropolis-with-in-Gibbs approach
- Extension to marked point processes (spatio-temporal)

Questions?

References

- Global Terrorism Database (2017). *National Consortium for the Study of Terrorism and Responses to Terrorism (START).* URL: https://www.start.umd.edu/gtd.
- Ogata, Y. (1981). "On Lewis' simulation method for point processes". In: *IEEE Transactions on Information Theory* 27.1, pp. 23–31.
- Rasmussen, J. G. (2013). "Bayesian Inference for Hawkes Processes". In: *Methodology and Computing in Applied Probability* 15.3, pp. 623–642.
- Ross, G. J. (2016). "Bayesian Estimation of the ETAS Model for Earthquake Occurrences". In: *http://www.gordonjross.co.uk/bayesianetas.pdf.*
- Zhuang, J., Y. Ogata, and D. Vere-Jones (2002). "Stochastic declustering of space-time earthquake occurrences". In: *Journal of the American Statistical Association* 97.458, pp. 369–380.