Identification of Latent Structure in Spatio-Temporal Models

Nick Clark, United States Military Academy -
Krista Watts, United States Military Academy -

The Devil’s in the dependency!

Goals of Statistical Models

- Inference - Learn about the underlying data generating mechanism
- Prediction - Use model to predict future observations
- Real world it isn't always so cut and dry
  - Often data is used both to infer and predict
  - The challenge is, how do we know our model is appropriate?

Modeling of Violence

- Motivation: 2017 National Institute of Justice competition
  - Can you predict where/when violence will occur in a city
  - Overall goal: Put cops on dots
- On the surface, two very different approaches both yielded excellent results
  - One relied on first order principles to build out a self-exciting process
  - The second relied on traditionally spatio-temporal statistical model
- Does it matter? Can we differentiate between the two processes?

Basics

Y is a Poisson process defined on a space-time set $s_i \times t$ with an intensity function for all bounded $B \in S$ given by $\lambda(s_i, t) = \int_{s_i \times t} \rho(\zeta) d\zeta$

- Informally - terrorism or criminal events arising at higher rates in regions of high intensity and lower rates in regions of lower intensity.
- How do we structure intensity field?

Hawkes Process

- Self-exciting process, intensity field is increased by previous observations
- Motivated by Partial Differential Equation similar to reaction diffusion equation

$$\lambda(s_i, t) = \mu(s_i, t) + \sum_{k:t_k<t} g(t - t_k, |s_i - s_k|)$$

$$g(t - t_k, |s_i - s_k|) = \lambda \exp(-\lambda(t - t_k)) \frac{1}{2\sigma^2} \exp(-||s_i - s_j|| \frac{1}{2\sigma^2})$$

Log Gaussian Cox Process

- Place spatio-temporal structure in an unobserved latent Gaussian process

$$Z(A) \sim \mathrm{Po}\left(\int_A \lambda(s_i, t) d(s_i) dt\right)$$

$$\log(\lambda) \sim MVN(\mu, \Sigma(\theta))$$

- Assume a sparse spatial precision matrix and a sparse temporal precision matrix

$$\Sigma_{s,t}^{-1} = \Sigma_t(\theta)^{-1} \otimes \Sigma_s(\theta)^{-1}$$

So does any of this matter?

Posterior Predictive Checks

- How do we determine if a model is appropriate? χ^2 GOF not an option…
- Can a fit model replicate key data characteristics?
- Compare the distribution of a given statistic of our data, $T(y)$ with the same statistic calculated from the posterior distribution of data simulated from our fitted model, $T(y_{rep}|y)$
- Estimating $P(T(y_{rep}|y) > T(y))$
- What statistic to use?

Ripley's K

- Quantifies the expected number of additional points within a set distance from another point over what is expected
- High values of Ripley's K mean there is substantial clustering in the data, whereas low values means the data are more dispersed than what would be expected
- Does not necessarily uniquely define a distribution
- Does it in our case?

Simulation

- Assume that underlying generative process is Hawkes
- Constant background intensity and λ, range σ

$$\lambda(s_i, t) = \mu(s_i, t) + \sum_{k:t_k<t} g(t - t_k, |s_i - s_k|)$$

$$g(t - t_k, |s_i - s_k|) = \lambda \exp(-\lambda(t - t_k)) \frac{1}{2\sigma^2} \exp(-||s_i - s_j|| \frac{1}{2\sigma^2})$$

Simulation

- Generate events from the background process over a 10x10 field of arbitrary units and 15 time periods

- Simulate using parent/child methodology

    - Parent events simulated from Poisson process with fixed mean
    - Each parent event spawns $N \sim Po(m)$ events spread spatially and temporally according to triggering function

- Data is then fit to both Hawkes Process as well as LGCP

Results

- Simulated from Hawkes Process across range of σ values, fit data to LGCP and then simulated from fitted LGCP
- As σ increased within the Hawkes process, the amount of clustering at a distance of 1 unit decreases, which is to be expected

$\sigma = 0.3$	$\sigma = 0.7$	$\sigma = 1.5$	$\sigma = 1.9$	$\sigma = 2.3$
3.75	3.74	3.72	3.69	3.64

LGCP fit to Hawkes Process

- Took Hawkes process data, fit it to LGCP, simulated from LGCP
- Posterior predictive checks indicated that the Ripley's K value from the LGCP routinely was similar to the Ripley's K from the Hawkes process

$\sigma = 0.3$	$\sigma = 0.7$	$\sigma = 1.5$	$\sigma = 1.9$	$\sigma = 2.3$
0.63	0.61	0.67	0.72	0.57

- About 60 % of the time the LGCP fit model generated Ripley's K values that were larger than the actual Ripley K value from the Hawkes process

Conclusion

- Second order process likely cannot differentiate between Hawkes and LGCP
- No clear rule to differentiate these two processes from each other
- Practitioners should use first-order principles to create testable models
- Recall goals of predictive models vs goals of inferential models
- Caution when assigning meaning to statistical model, competing model may fit data just as well with alternate meaning

Thanks!!