Bayesian Estimation of Modified ETAS Model for Invasive Species

Binhui Deng Dr. Earvin Balderama

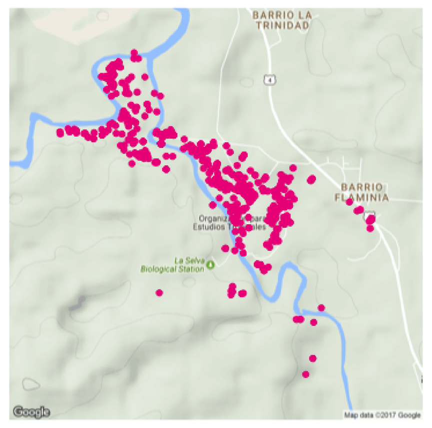
Department of Mathematics & Statistics, Loyola University Chicago, Chicago, IL, USA

Motivation

- **Invasive Plant:** Invasive alien species of plants can destroy the habitat of native species and disturb the natural evolution that takes place in the environment in which they spread.
- **ETAS Model:** Epidemic-Type Aftershock Sequence (ETAS) model, a spatial-temporal marked point process model, is the most popular stochastic model used to describe earthquake occurrence, to forecast earthquakes and to detect fluid/magma signals or induced seismicity.
- Adjusted ETAS model is applied onto the study of red banana tree to characterize its spatial-temporal spreading patterns.

Data

- **Red Banana Tree:** An invasive plant in Costa Rican rainforest have been observing in recent few years. Heights of 1008 red banana plants were observed and longitude and latitude coordinates were recorded by GPS. The age of plants was estimated according to the growth rate.
- 788 plants have complete location and origin times data, and were used for the subsequent analyses.



Epidemic-Type Aftershock Sequence (ETAS) Model

- The spatial-temporal ETAS model (Ogata, 1998) is a point process model where the rate of an earthquake of magnitude-M occurring at time t and location (x, y) depends on the histroy of occurrences H_t, characterized by the conditional intensity function:

$$\lambda(t,x,y|H_t) = \mu(x,y) + \sum_{\{i:t_i<t\}} \frac{K_0}{(t-t_i+c)^p} \cdot \frac{e^{\alpha(M-M_0)}}{((x-x_i)^2+(y-y_i)^2+d)^q}$$

$\mu(x,y)$: A non-homogeneous background rate.
$\sum$: Triggering function defined in the earthquake context.
K_0, α, c, p, d, q : Parameters to be estimated.

- The ETAS model for red banana plants by Balderama (2012) is

$$\lambda(t,x,y|H_t) = (1-\rho)\mu(x,y) + \frac{\rho\alpha\beta}{\pi} \sum_{\{i:t_i<t\}} e^{-\alpha(t-t_i)-\beta\{(x-x_i)^2+(y-y_i)^2\}}$$

$\frac{\alpha\beta}{\pi}$: A normalizing constant so that $\sum$ integrates to unity.
α : Exponential decay rate due to temporal distance apart.
β : Exponential decay rate due to spatial distance apart.
ρ : Proportion of events attributed to triggering by other events.

- The log-likelihood function for the occurrence of red banana plants is

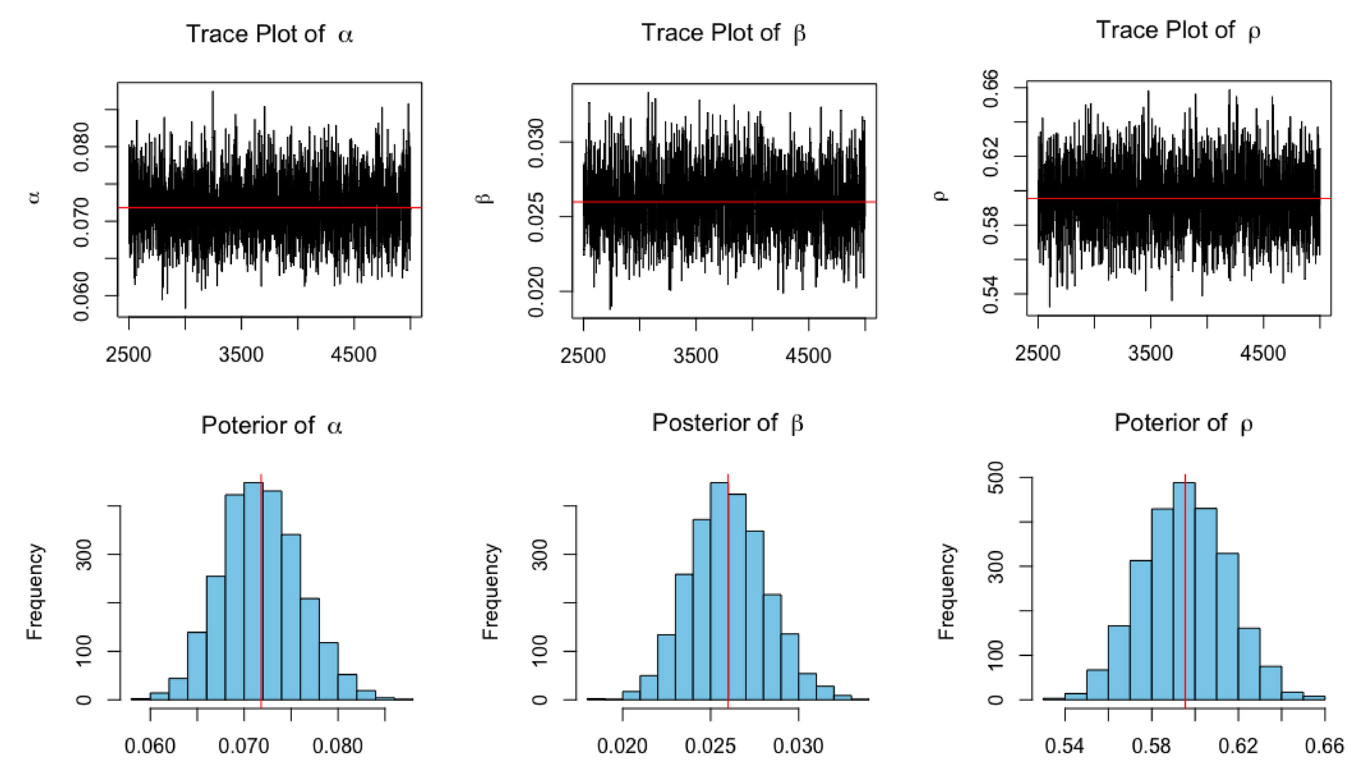
$$\begin{aligned} logL &= \sum_{i=1}^{n} log\lambda(t_i,x_i,y_i) - \iint_A \int_0^\infty \lambda(t,x,y)dtdxdy \\ &= \sum_{i=1}^{n} log\lambda(t_i,x_i,y_i) - (1-p)T\iint_A \mu(x,y)dxdy - pN \end{aligned}$$

Bayesian Estimation

- **Bayesian Analysis:** A statistical paradigm that answers research questions about unknown parameters using probability statements.
- **Prior Distribution:** A probability distribution with respect to current knowledge about parameters, written as $p(\theta)$.
- **Likelihood:** The information regarding parameters contained by new data y, which is proportional to $p(y|\theta)$.
- **Posterior Distribution:** An updated probability distribution, on which all Bayesian inference is based, produced by the combination of likelihood and the prior, written as $p(\theta|y)$.
- The posterior is proportional to the prior times the likelihood:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta}, \text{ where } \theta = \{\alpha, \beta, \rho\}$$

Results of Bayesian Estimation



Parameter	Posterior Mean	Posterior SD
α	0.0718	0.0042
β	0.0260	0.0022
ρ	0.5958	0.0195

- α: The intensity of triggering events decays at a rate of $1 - e^{-0.0718} = 6.93\%$ for every week that passes.
- β: The intensity of triggering events decays at a rate of $1 - e^{-0.026} = 2.57\%$ for every squared distance (in meters) away.
- ρ: 59.58% of events were triggered by previous events. Thus, 40.42% is the background rate, i.e., percentage of immigrant events.

Bayesian Estimate Versus MLE

Maximum Likelihood Estimates obtained by Newton-Raphson algorithm:

Parameter	MLE	Standard Error
α	0.0761	0.0045
β	0.0292	0.0022
ρ	0.5767	0.0193

- Estimates from Bayesian MCMC are very close to MLEs.
- Bayesian estimates have such small standard deviations that can be considered as excellent estimates.

Markov Chain Monte Carlo (MCMC Algorithm)

- **Goal of MCMC:** Draw samples from a posterior distribution without having to know its exact height (probability) at any point.
- **Monte Carlo Methods:** A class of algorithms relies on repeated random sampling to obtain numerical results.
- **Markov Chain:** A type of Markov Process that has either discrete state space or discrete index set (often representing time).
  - The state of the chain after a number of steps is then used as a sample of the desired distribution.
- **MCMC Algorithm:**
  1. Set up prior distributions and proposal distributions

Parameter	Prior Distribution	Proposal Distribution
α	Gamma(1,10)	Normal(α_0,0.1)
β	Gamma(1,10)	Normal(β_0,0.1)
ρ	Uniform(0,1)	Beta(ρ_0,$1-\rho_0$)

  2. Initialize θ by randomly picking a initial state from each prior.
  3. For t in (1:50000), update θ:
     - Propose a candidate α_1 (randomly pick a candidate state from the proposal distribution based on current state α_0).
     - Compute Metropolis-Hastings ratio **R**, which is $\frac{p(\alpha_1|y)}{p(\alpha_0|y)}$.
     - Accept α_1 and replace α_0 at probability **R**.
     - Proceed the same procedure above on β and ρ.
     - Record θ after each iteration.
  4. After 50,000 iterations, return all states for parameters. The 50,000 samples of each parameter make the posterior distributions.
  5. Compute the means of samples as the estimates of parameters.
- **Thinning:** Discard all but 10^{th} observations to get rid of autocorrelation to guarantee the independecy of samples.
- **Burn-in:** Throwing away first 2500 observations to minimize the effect of initial values on the posterior inference.

Future Work

- Simulation needs to be done to check if the adjusted ETAS model fits the actual red banana plants data.
- More environment covirates and variables should be considered into the spreading pattern of red banana plants, such as the distance from rivers, the amount of sunlight, elevation of plants etc.
- The adjusted ETAS model needs to be applied on other invasive speicies or point-process natural events to assess its effectiveness.

Reference

- Ogata, Y.(1988), "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes," *Journal of the American Statistical Association*, 83, 9-27.
- Balderama, E.(2012), "Spatial-Temporal Branching Point Process Models in the Study of Invasive Species," *University of California, Los Angeles*

http://math.luc.edu/~ebalderama/ ebalderama@luc.edu