Statistics 222, Spatial Statistics.

Outline for the day:

1. Continue with day08.r.
2. Nonparametric estimation of Hawkes processes using MISD.
3. Analytic nonparametric estimates.
4. Application to earthquakes and DRC Monkeypox.

Background and motivation.

* History of numerous models for earthquake forecasting, with mostly failures.
(elastic rebound, water levels, radon levels, animal signals, quiescence, electro-magnetic signals, characteristic earthquakes, AMR, Coulomb stress change, etc.)

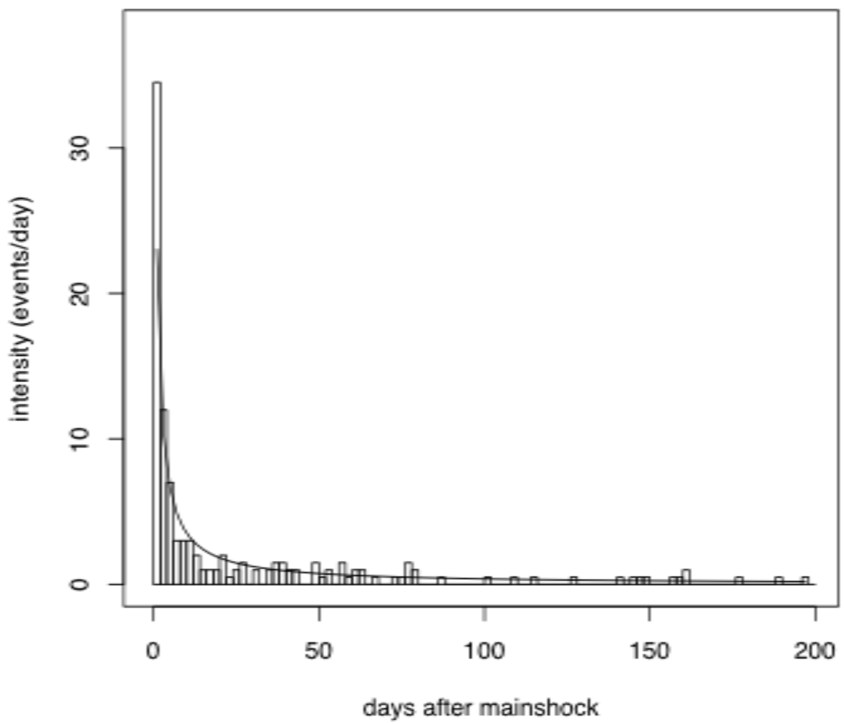
* Skepticism among many in seismological community toward *all* probabilistic forecasts.

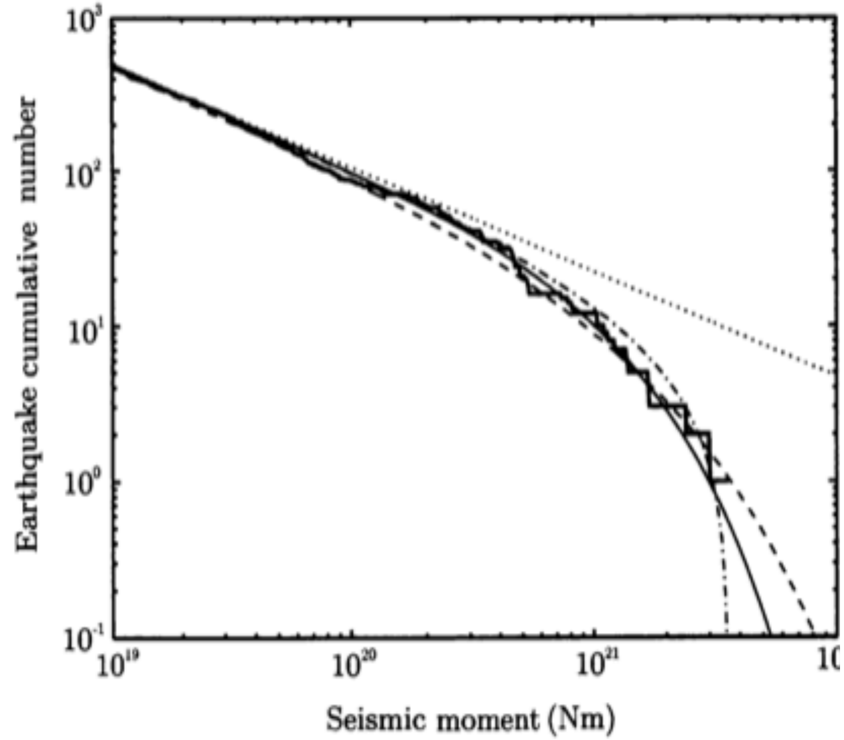
* Different models can have similar fit and very different implications for forecasts.
(e.g. Pareto vs. tapered Pareto for seismic moments. Fitting these by MLE to 3765 shallow worldwide events with M≥5.8 from 1977-2000,
the Pareto says there should be an event of M ≥ 10.0 every 102 years,
the tapered Pareto every 10^{436} years.
The fitted Pareto predicts an event with M≥12 every 10,500 years,
the tapered Pareto every 10^{43400} years.)

* Model evaluation techniques and forecasting experiments to discriminate among competing models and improve them are very important.

* We also need <u>non-parametric</u> alternatives to these models.

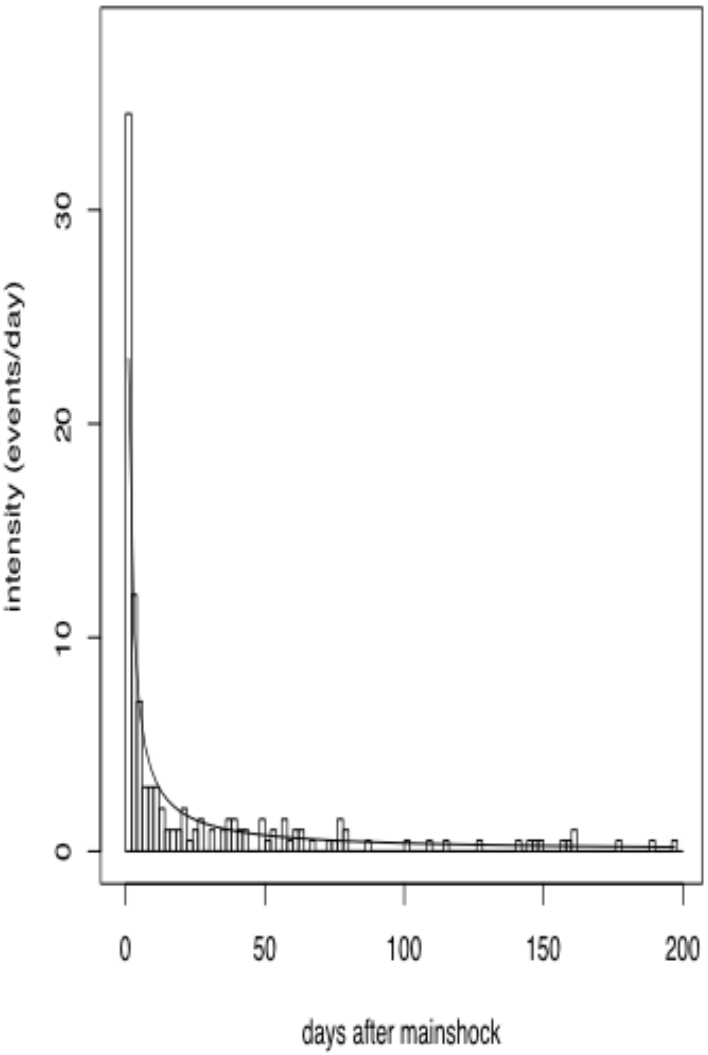
Kagan and Schoenberg (2001)



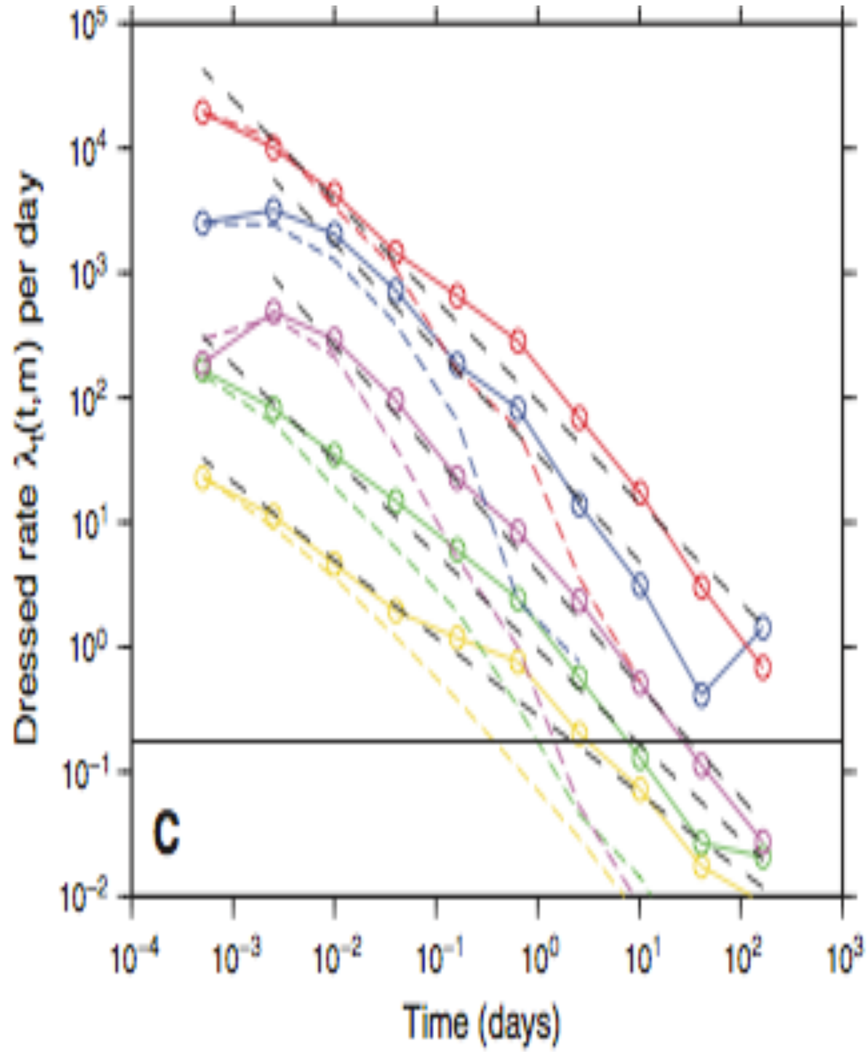


* We also need <u>non-parametric</u> alternatives to these models.

Temporal activity described by modified Omori Law: $K/(u+c)^p$



Marsan and Lengliné (2008)



Nonparametric estimation of Hawkes and ETAS processes.

Let $\mathbf{x}$ mean spatial coordinates = (x,y).
Hawkes processes have $\lambda(t,\mathbf{x}) = \mu(\mathbf{x}) + K \sum_i g(t-t_i, \mathbf{x}-\mathbf{x}_i)$.

- An ETAS model may be written

$$\lambda(t, \mathbf{x}|\mathcal{H}_t) = \mu(\mathbf{x}) + K \sum_{i:t_i<t} g(t - t_i, \mathbf{x} - \mathbf{x_i}, m_i),$$

  with triggering function

$$g(t-t_i, \mathbf{x}-\mathbf{x_i}, m_i) = \exp\{a(m_i-M_0)\}(t-t_i+c)^{-p}(||\mathbf{x}-\mathbf{x_i}||^2+d)^{-q}.$$

with e.g. $g(u, \mathbf{x}\ ; m_i) = (u+c)^{-p} \exp\{a(m_i-M_0)\}\ (||\mathbf{x}||^2 + d)^{-q}$.

These ETAS models were introduced by Ogata (1998).

Instead of estimating g parametrically, one can estimate g nonparametrically, using the method of Marsan and Lengliné (2008), which they call Model Independent Stochastic Declustering (MISD).

Extending Earthquakes' Reach Through Cascading

David Marsan* and Olivier Lengliné

Earthquakes, whatever their size, can trigger other earthquakes. Mainshocks cause aftershocks to occur, which in turn activate their own local aftershock sequences, resulting in a cascade of triggering that extends the reach of the initial mainshock. A long-lasting difficulty is to determine which earthquakes are connected, either directly or indirectly. Here we show that this causal structure can be found probabilistically, with no a priori model nor parameterization. Large regional earthquakes are found to have a short direct influence in comparison to the overall aftershock sequence duration. Relative to these large mainshocks, small earthquakes collectively have a greater effect on triggering. Hence, cascade triggering is a key component in earthquake interactions.

Earthquakes of all sizes, including aftershocks, are able to trigger their own aftershocks. The cascade of earthquake triggering causes the seismicity to develop complex, scale-invariant patterns. The causality of "mainshock A triggered aftershock B," which appears so obvious if mainshock A happens to be large, must then be modified into a more subtle "mainshock A triggered C1, which triggered C2, ..., which triggered B." This has paramount consequences: The physical mechanism that causes direct triggering (static or dynamic stress changes, fluid flow, afterslip, etc.) cannot be studied by looking at aftershocks that were not directly triggered by the mainshock. Moreover, if indirect triggering is important in the overall aftershock budget (*1–3*), then direct triggering must be confined to spatial ranges and times shorter than the size of the total

Laboratoire de Géophysique Interne et Tectonophysique, CNRS, Université de Savoie, 73376 Le Bourget du Lac, France.

*To whom correspondence should be addressed. E-mail: david.marsan@univ-savoie.fr

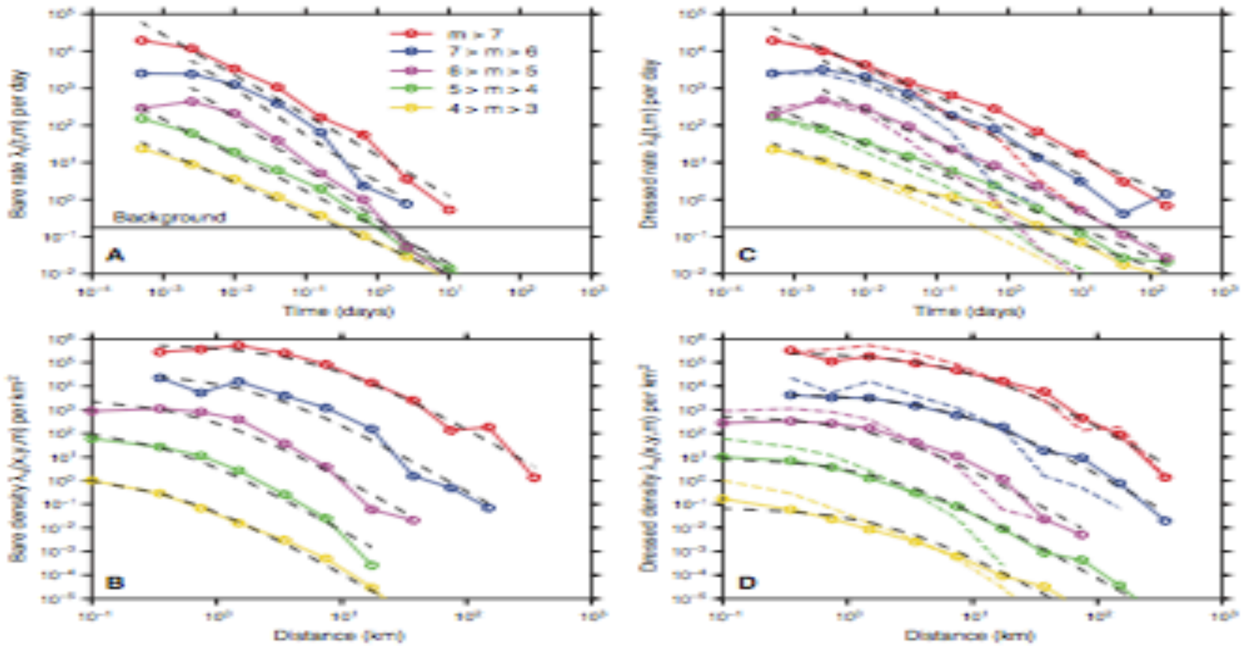


Fig. 1. Estimated rates and densities for California. (**A** and **B**) Bare kernels; (**C** and **D**) dressed kernels. The best power laws for the temporal rates $\lambda_t(t, m)$ and the best $[1 + (r/L)]^{-\delta}$ laws for the densities $\lambda_r(x, y, m)$ are shown as black dashed lines. The background temporal rate $\lambda_{0,t}$ [black horizontal line in (A) and (C)] is computed as $\sum_{i=1}^{n} w_{0,i}/T$. In (C) and (D), the dressed kernels (continuous lines) are compared to the bare ones (color dashed lines). The densities λ_r have been vertically shifted for clarity.

Model Independent Stochastic Declustering

- The method of Marsan and Lengliné (2008):

$$\lambda(t, m, x, y|\mathcal{H}_t) = \mu(x, y) + \sum_{j:t_j<t} \kappa(m_j)g(t - t_j)f(x - x_j, y - y_j),$$

- Maximizes the expectation of the complete data log-likelihood and assigns probabilities that a child event i is caused by an ancestor event j.

Expectation Step

$$p_{ij} = \frac{g(u)f(x, y)}{\mu(x, y) + \sum g(u)f(x, y)},$$

$$p_{ii} = \frac{\mu(x, y)}{\mu(x, y) + \sum g(u)f(x, y)}.$$

Nonparametric estimation of Hawkes and ETAS processes.

Gordon et al. (2017) let the triggering function, g,
depend on *magnitude*, *sub-region*, *distance*, and *angular separation* from the location (x, y) in question to the triggering event.

$$\lambda(t, m, x, y|\mathcal{H}_t) = \mu(x, y) + \sum_{j:t_j<t} \kappa(m_j)g(t - t_j)f(x - x_j, y - y_j; \phi_j, m_j),$$

Josh Gordon

Nonparametric estimation of Hawkes and ETAS processes.

Expectation Step

$$p_{ij} = \frac{g(u)f(x,y,\phi,m)}{\mu(x,y) + \sum g(u)f(x,y,\phi,m)},$$
$$p_{ii} = \frac{\mu(x,y)}{\mu(x,y) + \sum g(u)f(x,y,\phi,m)}.$$

Maximization Step

$$h(r,\theta,m)_{k,\ell,q} = \frac{\sum_{C_{k,\ell,q}} p_{ij}}{\Delta r_k \Delta\theta_\ell \Delta m_q \underbrace{\sum_{i=1}^{N}\sum_{j=1}^{i-1} p_{ij}}_{\#\text{ of Aftershocks}}},$$

- $C_{k,\ell,q} = \left\{(i,j) \middle| \delta r_k \le r_{ij} \le \delta r_{k+1},\ \delta\theta_\ell \le \theta_{ij} \le \delta\theta_{\ell+1},\ \delta m_q \le m_j \le \delta m_{q+1}, i > j\right\}$ is the set of indices of all pairs of events that fall within the bins specified by the multidimensional histogram density estimator for *magnitude*, *distance*, and *angular separation* $h(r,\theta,m)$.
- κ and g are maximized similarly

Nonparametric Marsan and Lengliné (2008) estimator.

Marsan and Lengliné (2008) assume g is a step function, and estimate steps β_k as parameters.

$$\ell(\theta) = \sum_i \log\left(\lambda(\tau_i, \mathbf{x}_i|\mathcal{H}_{\tau_i})\right) - \int_0^T \int_S \lambda(t, \mathbf{x}|\mathcal{H}_t)\, d\mathbf{x}dt$$

Setting the partial derivatives of this loglikelihood with respect to the steps β_k to zero yields

$$0 = \partial\ell(\theta)/\partial\beta_k = \sum_{(i,j):\tau_i - \tau_j \in U_k} K/\lambda(\tau_i) - Kn|U_k|,$$

where $|U_k|$ is the width of step k, for $k = 1,2,\ldots, p$. This is a system of p equations in p unknowns.

However, the equations are nonlinear. They depend on $1/\lambda(\tau_i)$.

Gradient descent methods: way too slow for large p.

Marsan and Lengliné (2008) find *approximate* maximum likelihood estimates using the E-M method for point processes. You pick initial values of the parameters, then given those, you know the probability event *i* triggered event *j*. Using these, you can weight each pair of points by its probability and re-estimate the parameters, and repeat until convergence.

This method works well but is iterative and time-consuming.

Nonparametric Marsan and Lengliné (2008) estimator.

The last step of Marsan and Lengliné uses essentially a histogram estimator.

Others have used slightly different approaches for smoothing.

- Lewis and Mohler (2011) use maximum penalized likelihood.
- Bacry et al. (2012) use the Laplace transform of the covariance function.
- Adelfio and Chiodi (2015) use a semi-parametric method where the background rate λ is estimated nonparametrically and the triggering function g parametrically.

There are also standard non-parametric methods for smoothing points generally, using splines, kernels, or wavelets. (Brillinger 1997).

Marsan and Lengliné's method is different. It estimates g, not the overall rate.

Analytic solution.

Set $p = n$. (p = *number of steps in the step function*, g, *and* n = *# of observed points.*)

Setting the derivatives of the loglikelihood to zero we have the p equations

$$0 = \partial \ell(\theta)/\partial \beta_k = \sum_{(i,j):\tau_i - \tau_j \in U_k} K/\lambda(\tau_i) - Kn|U_k|,$$

which are p linear equations in terms of $1/\lambda(\tau_i)$, for $i = 1,2, ..., n$. (!)

So, if $p=n$, then we can use these equations to solve for $1/\lambda(\tau_i)$,

and if we know $1/\lambda(\tau_i)$, then we know $\lambda(\tau_i)$,

and if we know $\lambda(\tau_i)$, then we can solve for β_i because the def. of a Hawkes process is

$$\lambda(\tau_j) = \mu + K \sum_{i<j} g(\tau_j - \tau_i),$$

which results in n linear equations in the p unknowns $\beta_1, \beta_2, ..., \beta_p$, when g is a step function.

Analytic solution.

We can write the resulting estimator in very condensed form.

Let $\boldsymbol{\lambda} = \{\lambda(\tau_1), \lambda(\tau_2), \ldots, \lambda(\tau_n)\}$.

Suppose the steps of g have equal widths, $|U_1| = |U_2|$, etc. Call this width U.

Let A_{ij} = the number points τ_k such that $\tau_j - \tau_k$ is in U_i, for i,j in $\{1,2,\ldots,p\}$.

Then the loglikelihood derivatives equalling zero can be rewritten

$$0 = KA(1/\boldsymbol{\lambda}) - Kb,$$

where $\mathbf{b} = nU\mathbf{1}$, with $\mathbf{1} = \{1,1,\ldots,1\}$.

This has solution $\underline{1/\boldsymbol{\lambda} = A^{-1} b}$, if A is invertible.

Similarly, the Hawkes equation can be rewritten $\boldsymbol{\lambda} = \mu + KA^T\boldsymbol{\beta}$, whose solution is

$$\underline{\hat{b} = (KA^T)^{-1}(\lambda - \mu)}.$$

Combining these two underlined formulas yields the estimates

$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$

This is very simple, trivial to program, and rapid to compute.

Analytic solution.

There are problems, however.

1. Estimating $n=p$ steps. High variance.

   However, if we can assume g is smooth, then we can smooth our estimates for stability.

2. Need to estimate K and μ too.

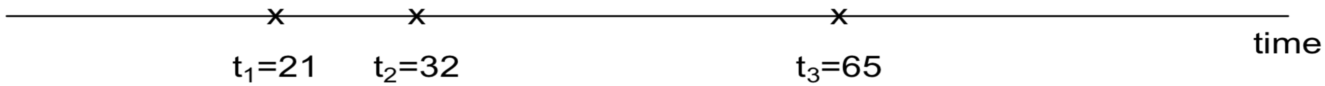
   We can use Marsan or take derivatives for these as well.

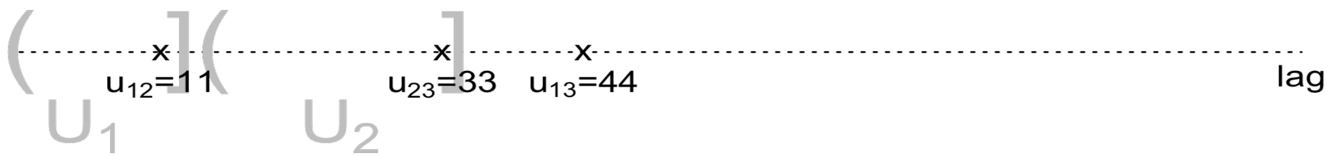
3. What about spatially-varying steps and unequally sized steps for g?

   No problem. The estimation generalizes in a completely obvious way.

4. A can be singular.
   We may need better solutions for this.
   I let $u_j = \tau_j - \tau_{j-1}$, sorted the u_j values, and then used $[u_{(1)}, u_{(2)})$, etc. as my binwidths, so each row and column of A would have at least one non-zero entry. If it still isn't singular, adding in a few random 1's into A often helps.

$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$





4. *A* can be singular.

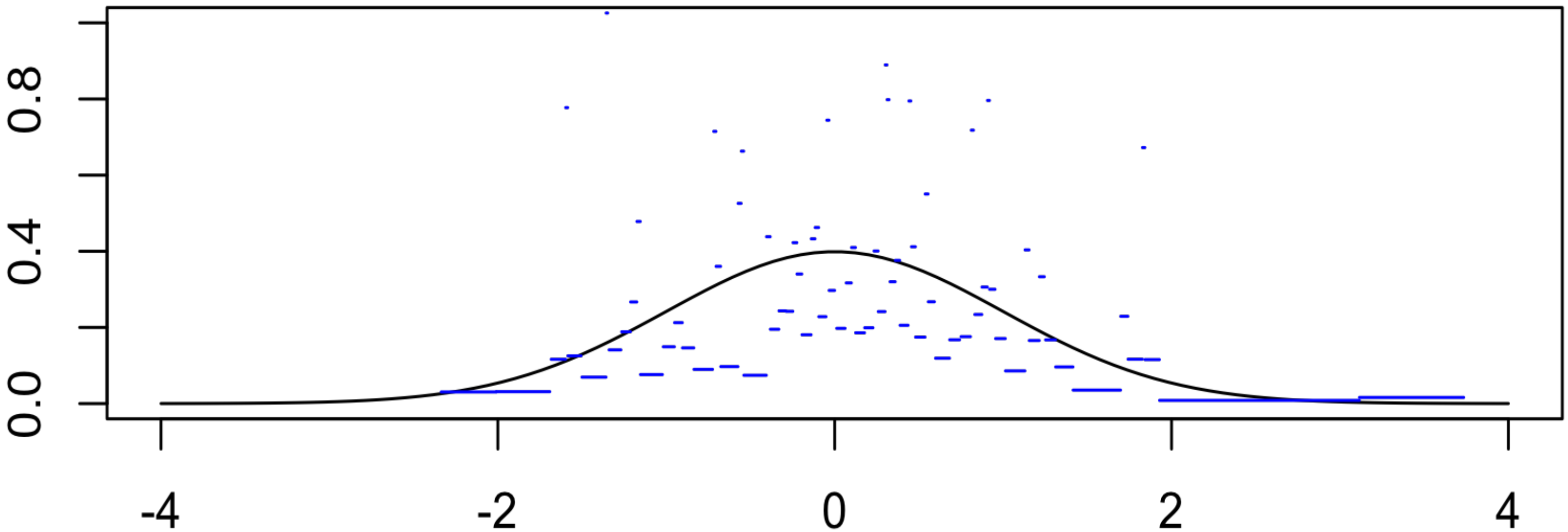
We may need better solutions for this.

I let $u_j = \tau_j - \tau_{j-1}$, sorted the u_j values, and then used $[u_{(1)}, u_{(2)})$, etc. as my binwidths, so each row and column of *A* would have at least one non-zero entry. If it still isn't singular, adding in a few random 1's into *A* often helps.

$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$

Note: take the simple case of a dataset where point i is only influenced by point i-1. This is basically a renewal process, and we are just estimating a renewal density.

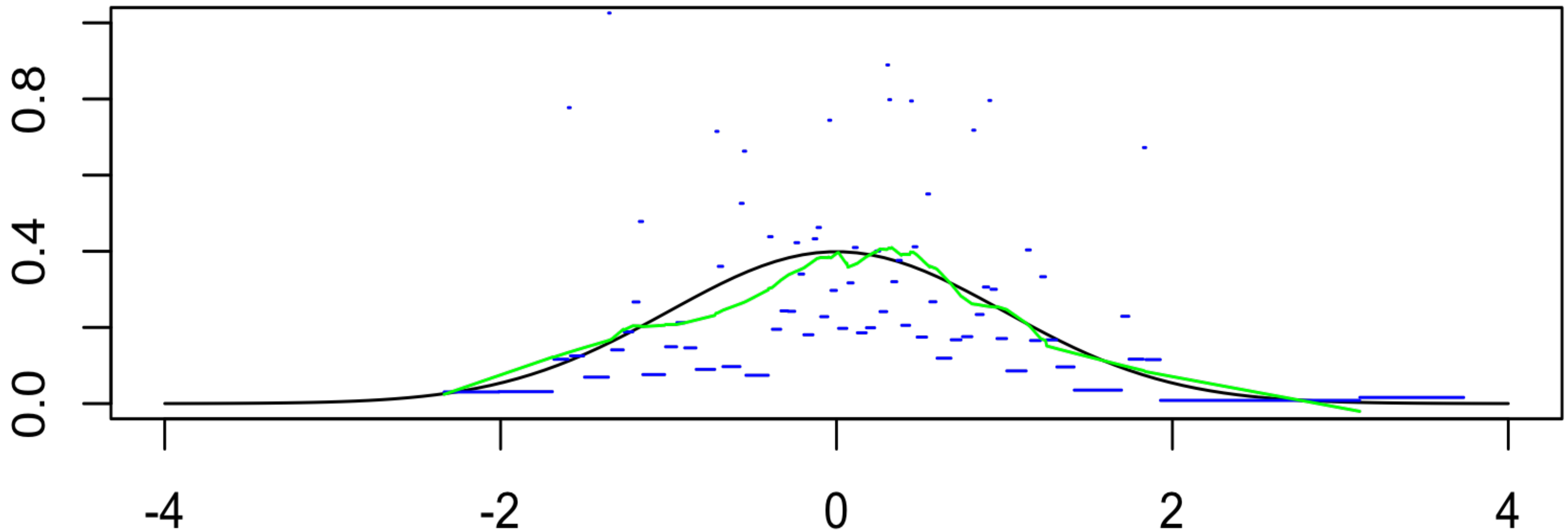
Here A = I, K = 1, and we get the density estimator $1/\{n(x_i - x_{i-1})\}$.



$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$

Note: take the simple case of a dataset where point i is only influenced by point i-1. This is basically a renewal process, and we are just estimating a renewal density.

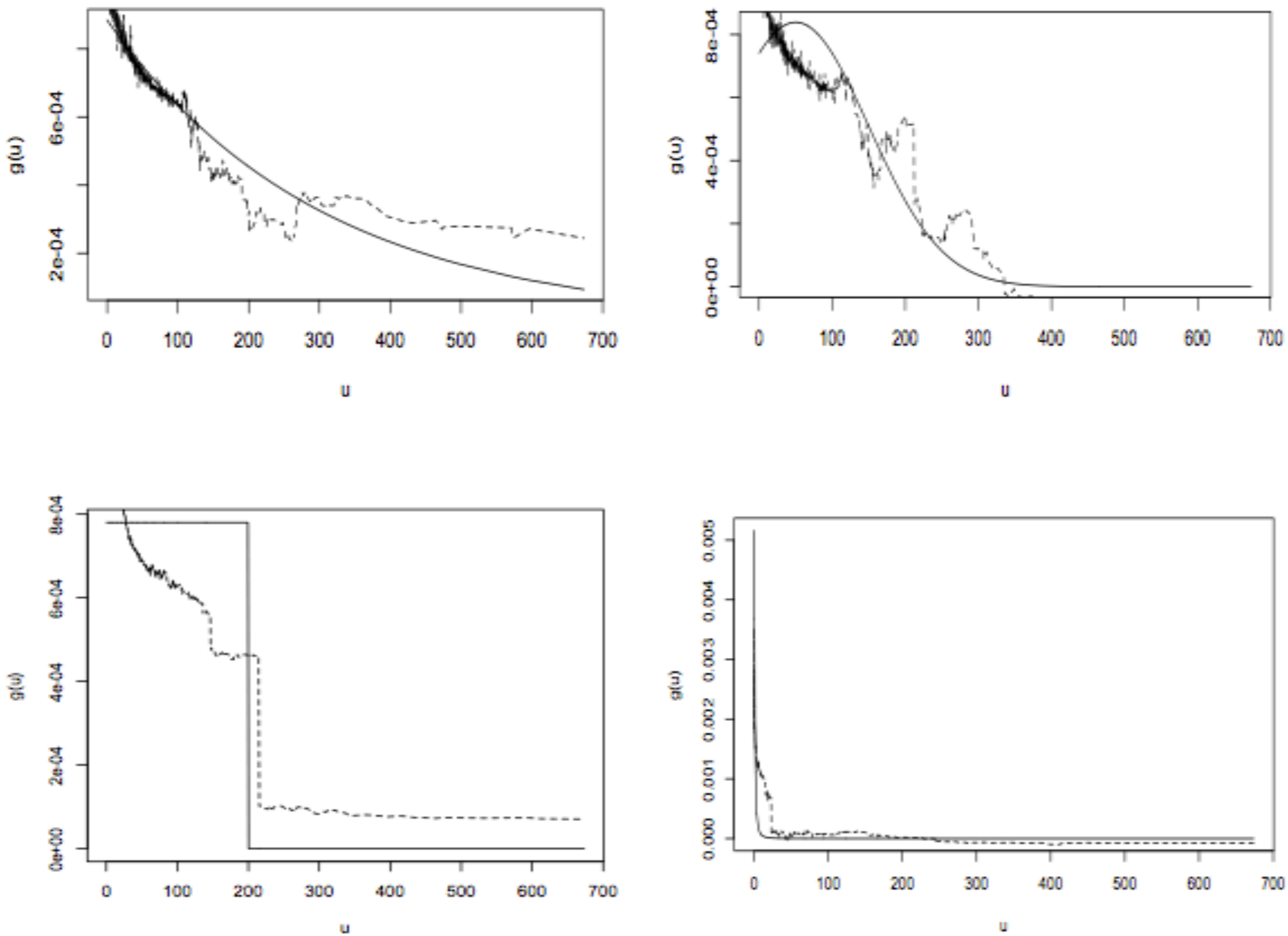
Here $A = I$, $K = 1$, and we get the density estimator $1/\{n(x_i - x_{i-1})\}$.



$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$

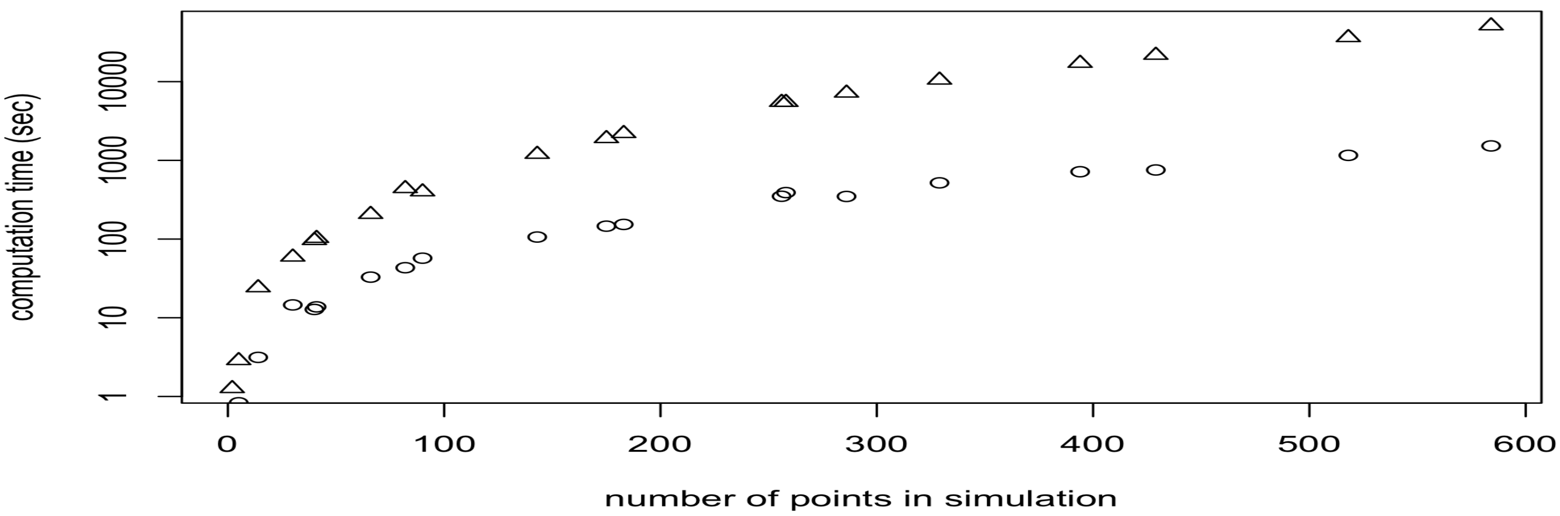
Computation time and performance comparison.

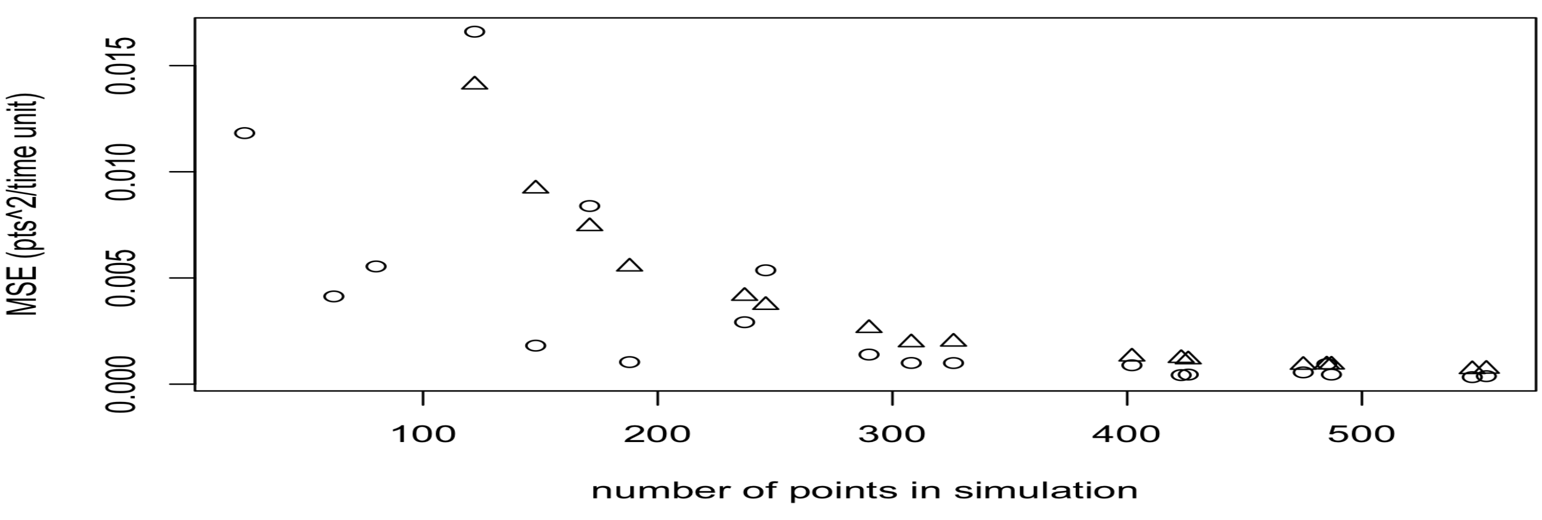
Test of concept. Examples of exponential, truncated normal, uniform, and Pareto g.



Computation time and performance comparison.

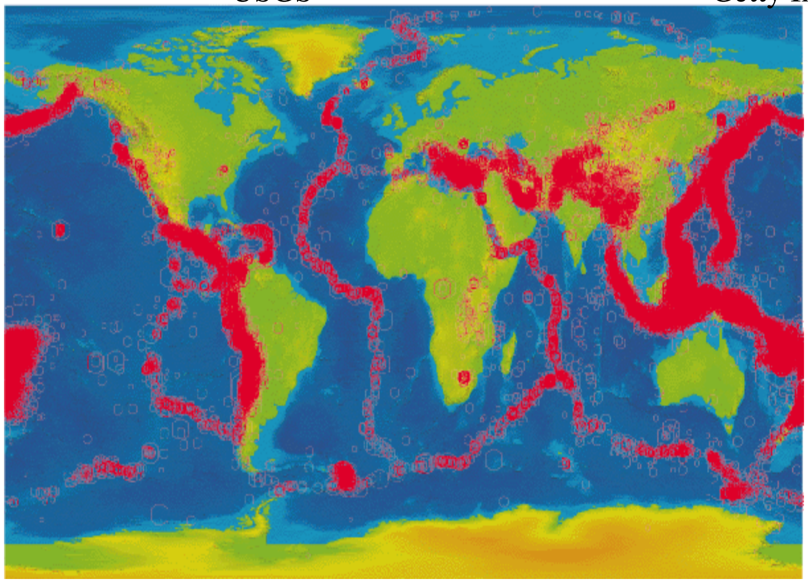
Triangles = Marsan and Lengliné (2008) method. Circles = analytic method.





Applications to earthquakes and DRC Monkeypox

USGS

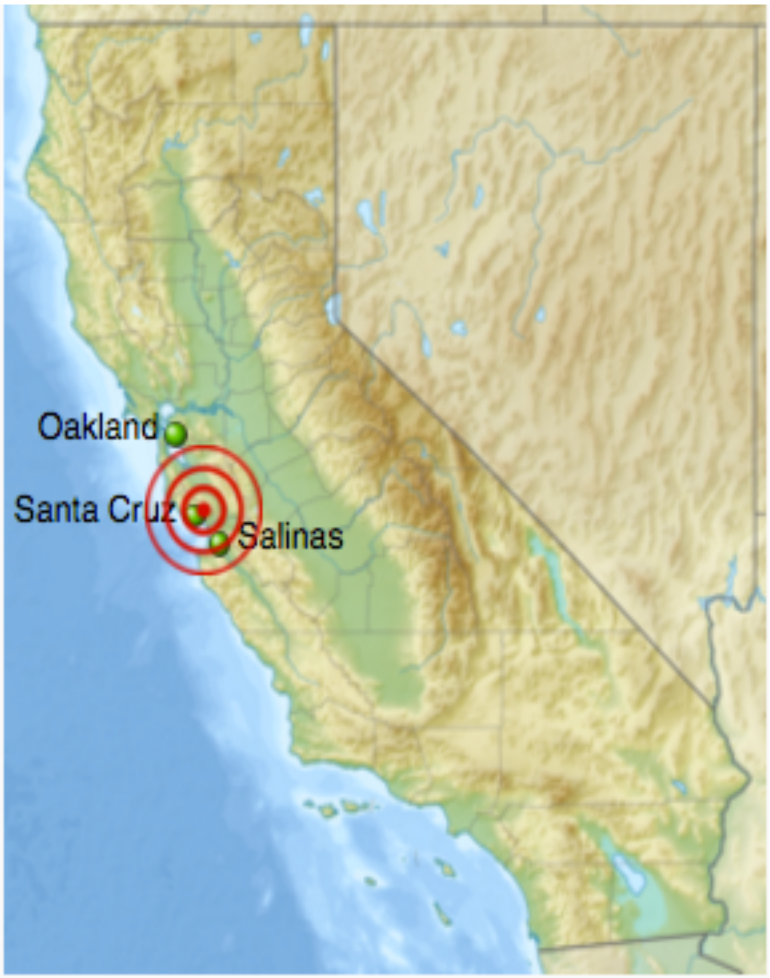


Getty Images

Application to Loma Prieta earthquake data.

Loma Prieta earthquake was Mw 6.9 on Oct 17, 1989.

As an illustration, we will estimate *g* on its 5566 aftershocks M ≥ 3 within 15 months.



(Google images)

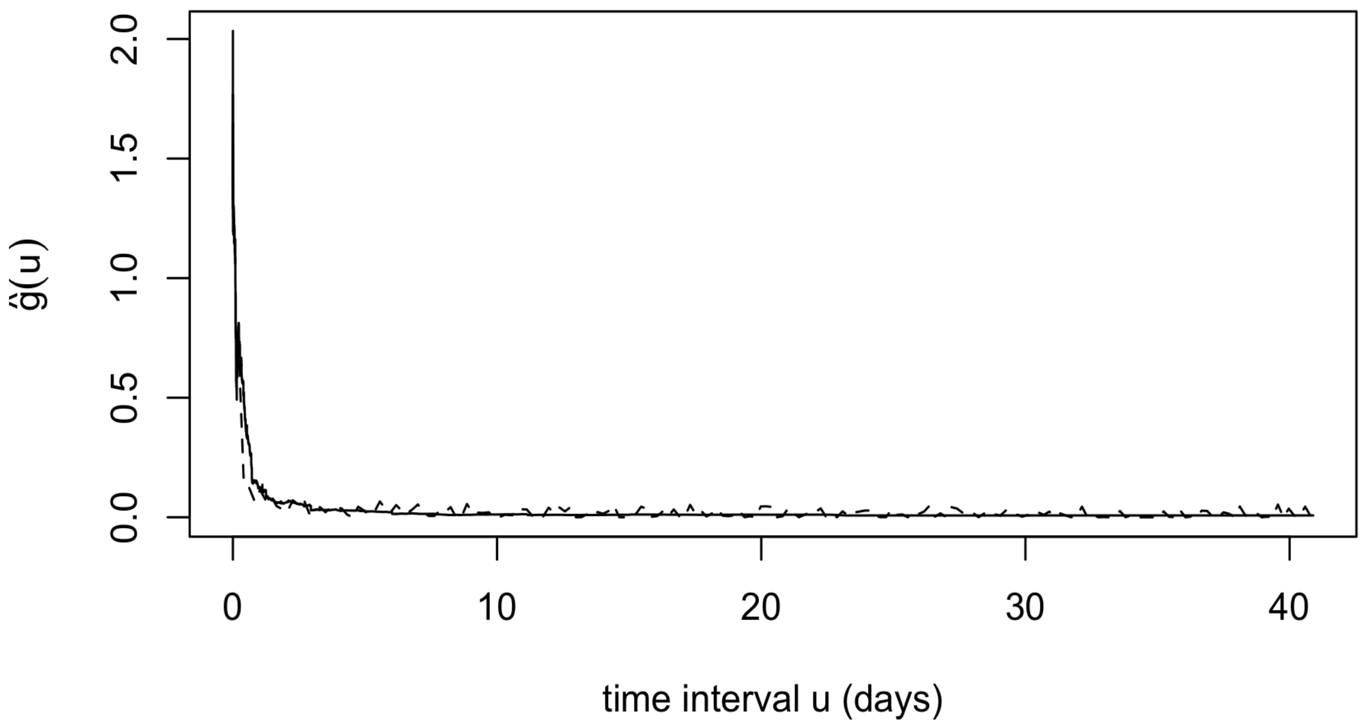
Application to Loma Prieta earthquake data. (SCEC.ORG)

Estimated triggering function for 5567 Loma Prieta M ≥ 3 events, 10/16/1989 to 1/17/1990.

Solid curve is the analytic method and dashed curve is Marsan and Lengliné (2008).

Dotted curves are estimates based on analytic method +/- 1 or 2 SEs, respectively, for light grey and dark grey.

SEs were computed using the SD of analytic estimates in 100 simulations of Hawkes processes with triggering functions sampled from the solid curve.



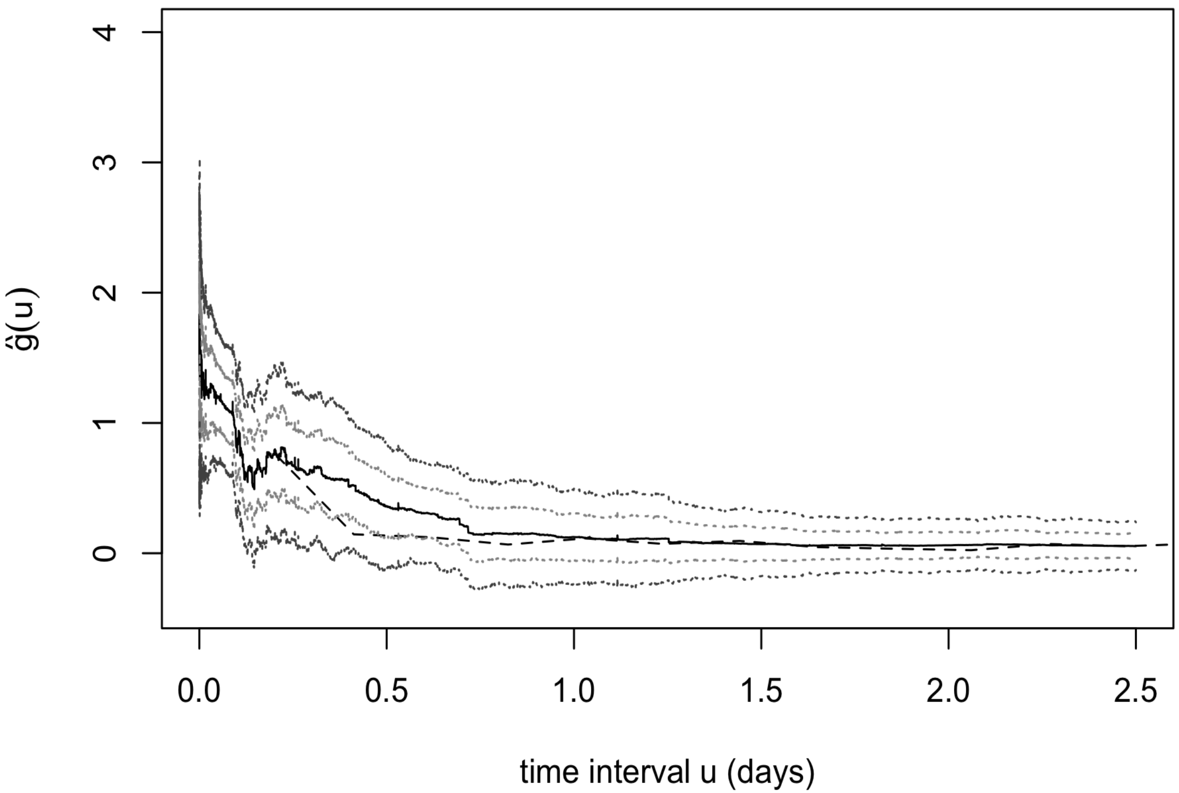
Application to Loma Prieta earthquake data.

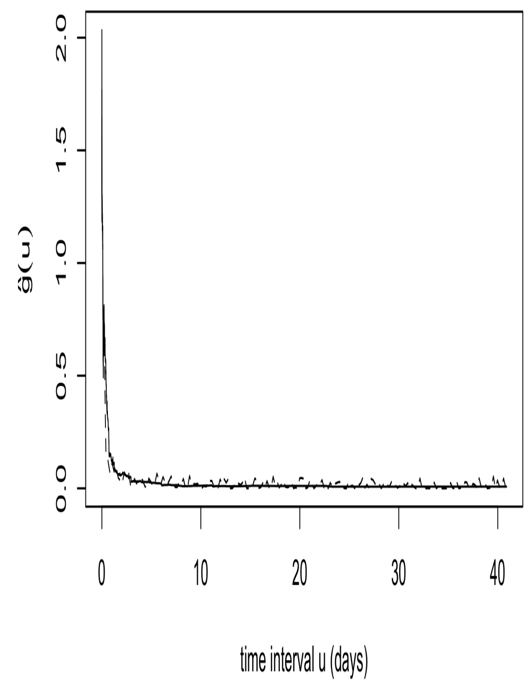
Estimated triggering function for Loma Prieta seismicity M ≥ 3, 10/16/1989 to 1/17/1990.

Solid curve is the analytic method and dashed curve is Marsan and Lengliné (2008).

Dotted curves are estimates based on analytic method +/- 1 or 2 SEs, respectively, for light grey and dark grey.

SEs were computed using the SD of analytic estimates in 100 simulations of Hawkes processes with triggering functions sampled from the solid curve.





Application to DRC Monkeypox data.

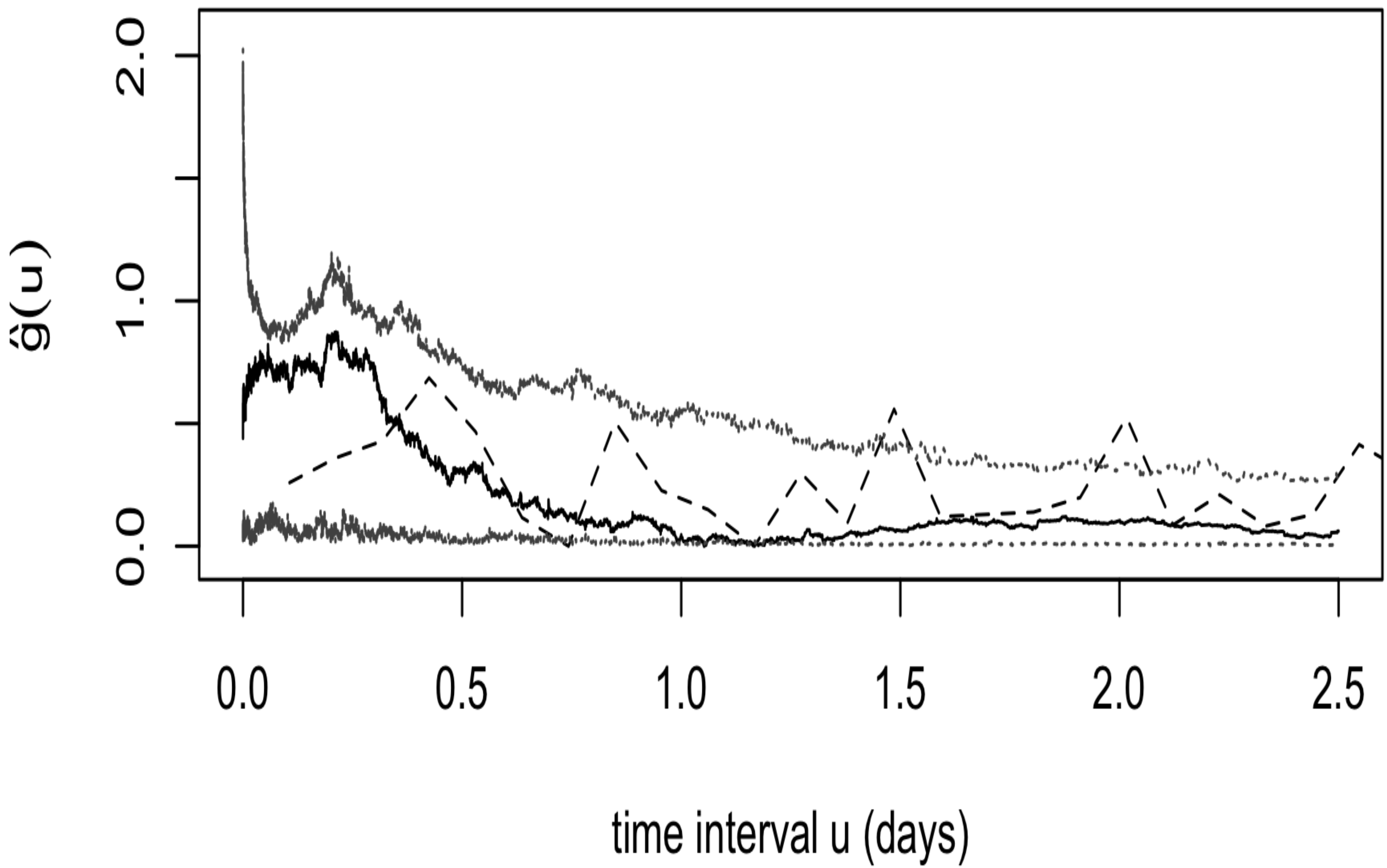
566 investigated cases reported by WHO
with incident date in 2005 or 2006.
Times within each day randomized uniformly.

(Google images)

(Getty images)

Application to DRC Monkeypox data.

Application to DRC Monkeypox data.



Concluding remarks.

The idea is to

let p=n,

let the derivatives of the log-likelihood be zero,

solve for $1/\lambda_i$ and therefore get λ_i,

and solve for β.

a. One can have major computation time savings from this method.

For datasets of only 100-300 points the savings are negligible.

However, for 5,000 points, the Marsan and Lengliné (2008) algorithm with 100 iterations takes about 7 hours, whereas the analytic method here takes 1.3 min.

This speed facilitates computations like simulation based confidence intervals.

b. How far can this go?

It extends very readily to space-time-magnitude and estimation of μ .

Would this work for other types of models too? What are the limits on this method?

c. What about when A is singular? More work is needed.

d. Meaning of CIs in this context?

Concluding remarks.

The idea is to

let p=n,

let the derivatives of the log-likelihood be zero,

solve for $1/\lambda_i$ and therefore get λ_i,

and solve for β.

a. One can have major computation time savings from this method.

For datasets of only 100-300 points the savings are negligible.

However, for 5,000 points, the Marsan and Lengliné (2008) algorithm with 100 iterations takes about 7 hours, whereas the analytic method here takes 1.3 min.

This speed facilitates computations like simulation based confidence intervals.

b. How far can this go?

It extends very readily to space-time-magnitude and estimation of μ .

Would this work for other types of models too? What are the limits on this method?

c. What about when A is singular? More work is needed.

d. Meaning of CIs in this context?