Statistics 222, Spatial Statistics.

Outline for the day:

1. Irregular boundaries.
2. Exponential density in the plane.
3. Review exercises.
4. Nonparametric estimation of Hawkes models.
5. Application of nonparametric estimates to earthquakes and plague.

1. Modifying F,G,J,K,L functions to deal with irregular boundaries is in the file custom_obs_window_jkl thanks to Michael Tzen. It is on the course site.

2. Exponential density in the plane.

I originally had this.

Fitting a Pseudo-Likelihood model.
I'm using the model lambda_p (z | z_1, ..., z_k) =
mu + alpha x + beta y + gamma SUM_{i = 1 to k} a1 exp{-a1 D(z_i,z)}
where z = (x,y), and where D means distance.
So, if gamma is positive, then there is clustering; otherwise inhibition.

But $g(r) = a_1 \exp(-a_1 r)$ is actually not a density.
$g(t) = a_1 \exp(-a_1 t)$ is a density, because $\int_0^\infty a_1 \exp(-a_1 t)\, dt = 1$, for $a_1 > 0$,
but not $\int\int a_1 \exp(-a_1 r)\, dx\, dy$.

$a_1 \exp(-a_1 r) / (2\pi r)$ is a spatial density, because
$\iint a_1 \exp(-a_1 r) / (2\pi r)\, dx\, dy = \int_0^{2\pi} \int_0^\infty a_1 \exp(-a_1 r) / (2\pi r)\, r\, dr\, d\phi$
$= \int_0^\infty a_1 \exp(-a_1 r)\, dr$
$= 1$.

So I should fit **lambda_p (z | z_1, ..., z_k) =**
mu + alpha x + beta y + gamma SUM_{i = 1 to k} a1/2π exp{-a1 D(z_i,z)}/D(z_i,z).
This is in the current version of day08.r.

2. Problems.

Suppose you observe a Poisson process with rate μ on the space-time window [0,1]x[0,1] x [0,10], and it happens to have 5 points.

S T.

What is the log-likelihood, $l(\mu)$?

a) $5\,\mu + 10\,\exp(\mu)$.
b) $5\,\log(\mu) - 10\,\mu$.
c) $5 + 10\,\log(\mu)$.
d) $5\,\exp(\mu) + 5\,\log(\mu)$.

Problems.

Suppose you observe a Poisson process with rate μ on the space-time window [0,1]x[0,1] x [0,10], and it happens to have 5 points.

S T.

What is the log-likelihood, $l(\mu)$?

a) $5\,\mu + 10\,\exp(\mu)$.

b) $5\,\log(\mu) - 10\,\mu$.

c) $5 + 10\,\log(\mu)$.

d) $5\,\exp(\mu) + 5\,\log(\mu)$.

Problems.

Suppose you observe a Poisson process with rate 3t on the space-time window [0,1]x[0,1] x [0,10].

S T.

How many points do you expect to observe?

a) 50.
b) 100.
c) 150.
d) 200.

Problems.

Suppose you observe a Poisson process with rate 3t on the space-time window [0,1]x[0,1] x [0,10].

S T.

How many points do you expect to observe?

a) 50.
b) 100.
c) 150.
d) 200.

$\int\int\int 3t\, dx\, dy\, dt = \int 3t\, dt = 3t^2/2\ ]_0^{10} = 300/2 - 0 = 150.$

Problems.

Suppose you observe a Hawkes process with conditional intensity
$\lambda(t,x,y) = 2 + 0.6 \int f(t-t') g(x-x',y-y') dN(t',x',y')$, on the space-time window
$[0,1]x[0,1] \times [0,10]$,
S T,
where f(t) is a density like $f(t) = 4\exp(-4t)$, and g(x,y) is a planar density like
$g(x,y) = 3 \exp(-3r) / (2\pi r)$, where $r = \sqrt{(x^2+y^2)}$.

How many points do you expect to observe?

a) 50.
b) 100.
c) 150.
d) 200.

Problems.

Suppose you observe a Hawkes process with conditional intensity
$\lambda(t,x,y) = 2 + 0.6 \int f(t-t')\, g(x-x',y-y')\, dN(t',x',y')$, on the space-time window
$[0,1]x[0,1] \times [0,10]$,
S T,
where f(t) is a density like $f(t) = 4\exp(-4t)$, and g(x,y) is a planar density like
$g(x,y) = 3 \exp(-3r) / (2\pi r)$, where $r = \sqrt{(x^2+y^2)}$.

How many points do you expect to observe?

a) 50.
b) 100.
c) 150.
d) 200.

$20 + 20 \times .6 + 20 \times .6^2 + 20 \times .6^3 + \ldots = 20/(1-.6) = 20/.4 = 50.$

Problems.

Suppose you observe a Hawkes process with conditional intensity
$\lambda(t,x,y) = 2 + 0.6 \int f(t-t') g(x-x',y-y') dN(t',x',y')$, on the space-time window
$[0,1]x[0,1] \times [0,10]$,
S T,
where $f(t) = 4\exp(-4t)$, and
$g(x,y) = 3 \exp(-3r) / (2\pi r)$, where $r = \sqrt{(x^2+y^2)}$.

You observe 2 points, at (t,x,y) = (1,.5,.5) and (3,.5,.6). What is the log-likelihood?

a) $\log(2) + \log(2 + 36 \exp(-8.3)/\pi) - 21.2$.
b) $\log(3.2) + \log(2 + 36 \exp(-8.3)/\pi) - 20$.
c) $\log(3.2) + \log(2 + 36 \exp(-8.3)/\pi) - 20$.
d) $2\log(2) - 20$.

Problems.

Suppose you observe a Hawkes process with conditional intensity
$\lambda(t,x,y) = 2 + 0.6 \int f(t-t')\, g(x-x',y-y')\, dN(t',x',y')$, on the space-time window
$[0,1]x[0,1] \times [0,10]$,
S T,
where $f(t) = 4\exp(-4t)$, and
$g(x,y) = 3 \exp(-3r) / (2\pi r)$, where $r = \sqrt{(x^2+y^2)}$.

You observe 2 points, at (t,x,y) = (1,.5,.5) and (3,.5,.6). What is the log-likelihood?

a) **$\log(2) + \log(2 + 36 \exp(-8.3)/\pi) - 21.2$.**
b) $\log(3.2) + \log(2 + 36 \exp(-8.3)/\pi) - 20$.
c) $\log(3.2) + \log(2 + 36 \exp(-8.3)/\pi) - 20$.
d) $2\log(2) - 20$.
$\sum\log(\lambda) - \int\lambda d\mu = \log(2) + \log\{2 + .6(4\exp(-8))(3\exp(-.3)/(.2\pi))\} - 20 - .6 - .6$
$= \log(2) + \log(2 + 36 \exp(-8.3)/\pi) - 21.2$.

Nonparametric estimation of Hawkes and ETAS processes.

Let **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).

David Marsan

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 i A happens to be large, must then into a more subtle "mainshock C1, which triggered C2, ..., wh B." This has paramount conse physical mechanism that cause gering (static or dynamic str fluid flow, afterslip, etc.) cann by looking at aftershocks th directly triggered by the mains over, if indirect triggering is imp overall aftershock budget ($I-3$), 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

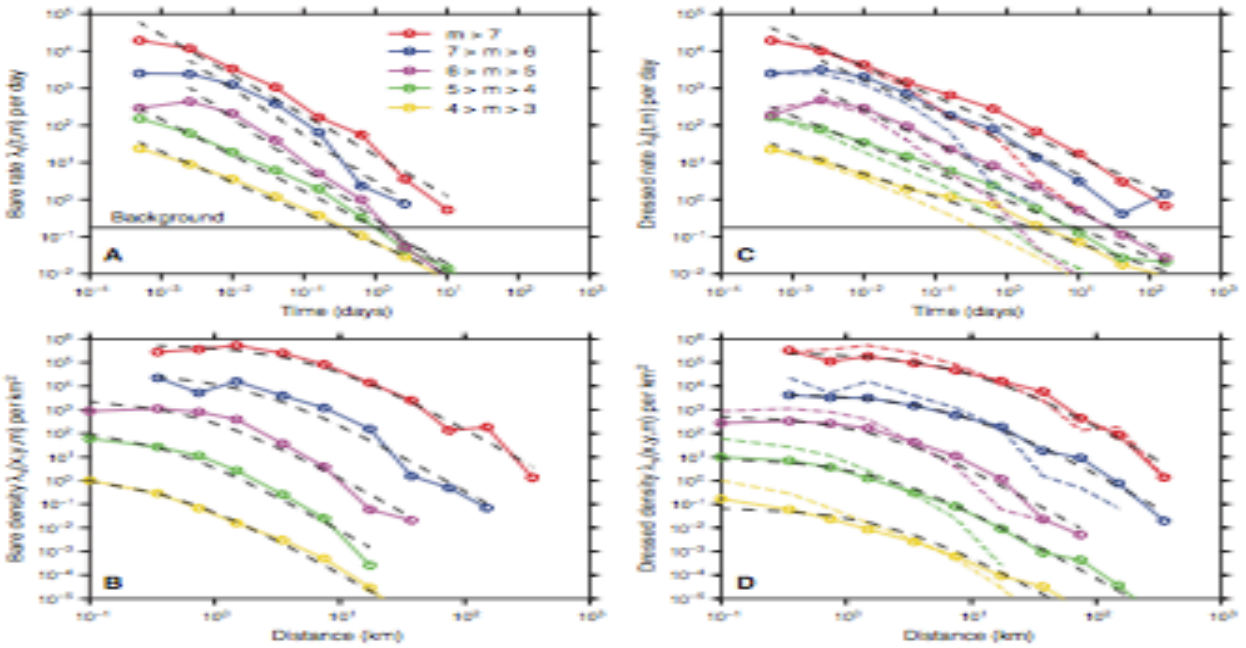


Fig. 1. Estimated rates and densities for California. (**A** and **B**) Bare kernels; (**C** and **D**) dressed kernels. The best power laws for the temporal rates $\lambda_t(t, m)$ and the best $[1 + (r/L)]^{-\delta}$ laws for the densities $\lambda_r(x, y, m)$ are shown as black dashed lines. The background 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.

Nonparametric estimation of Hawkes and ETAS processes.

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 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.

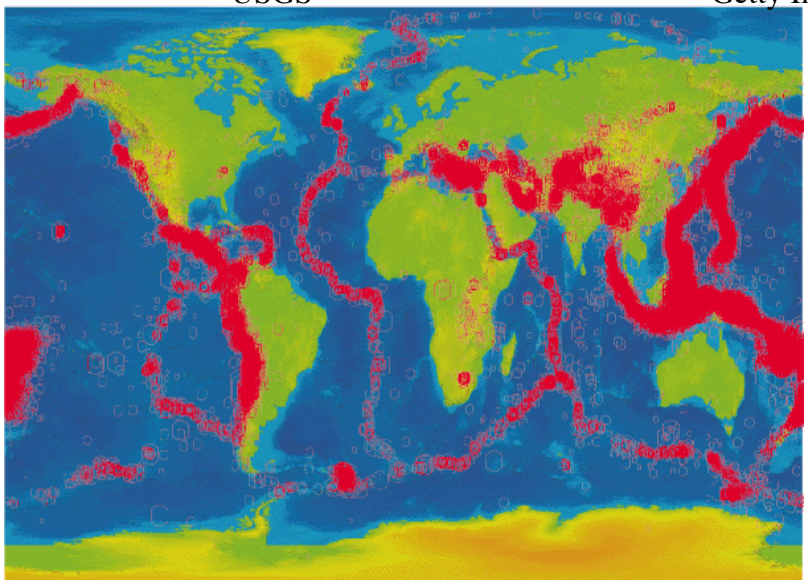
Adelfio and Chiodi (2015) use a semi-parametric method where the background rate λ is estimated nonparametrically and the triggering function g parametrically.

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.

Applications to earthquakes and US plague.

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 $M \geq 3$ within 15 months.

(Google images)

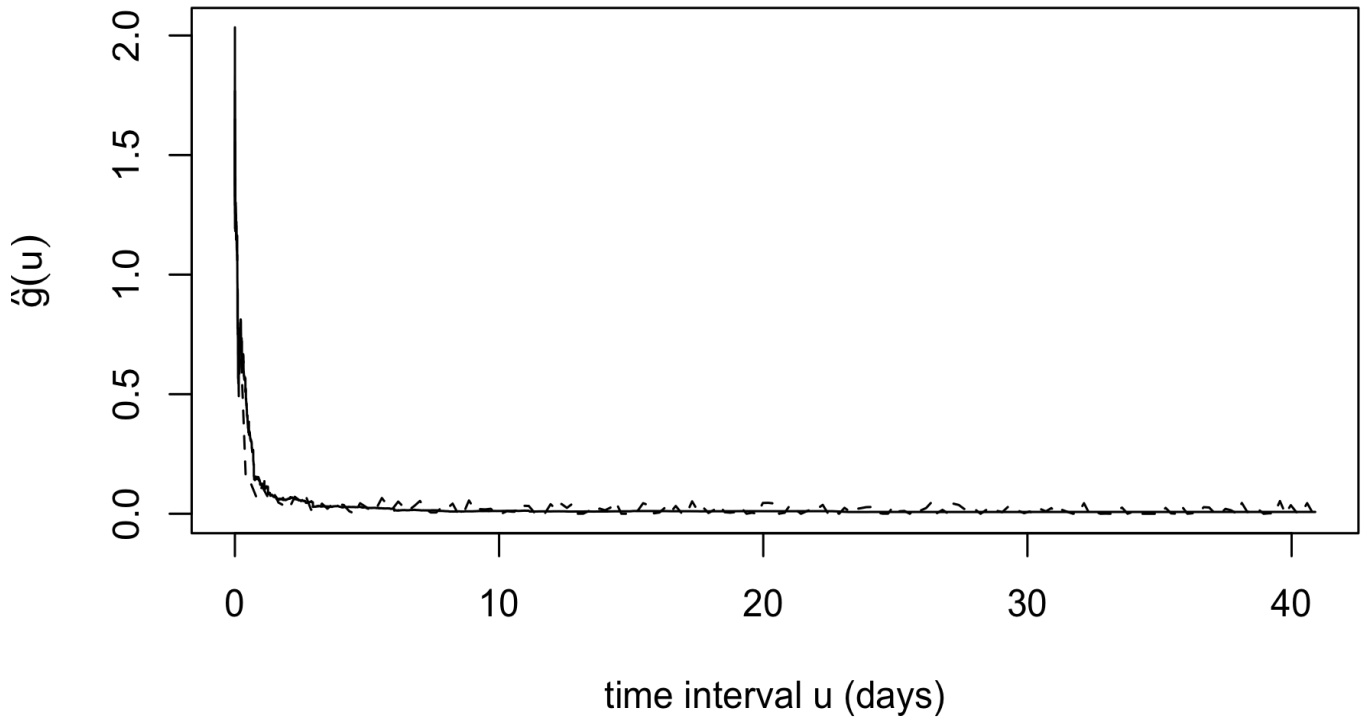
Application to Loma Prieta earthquake data. (SCEC.ORG)

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.



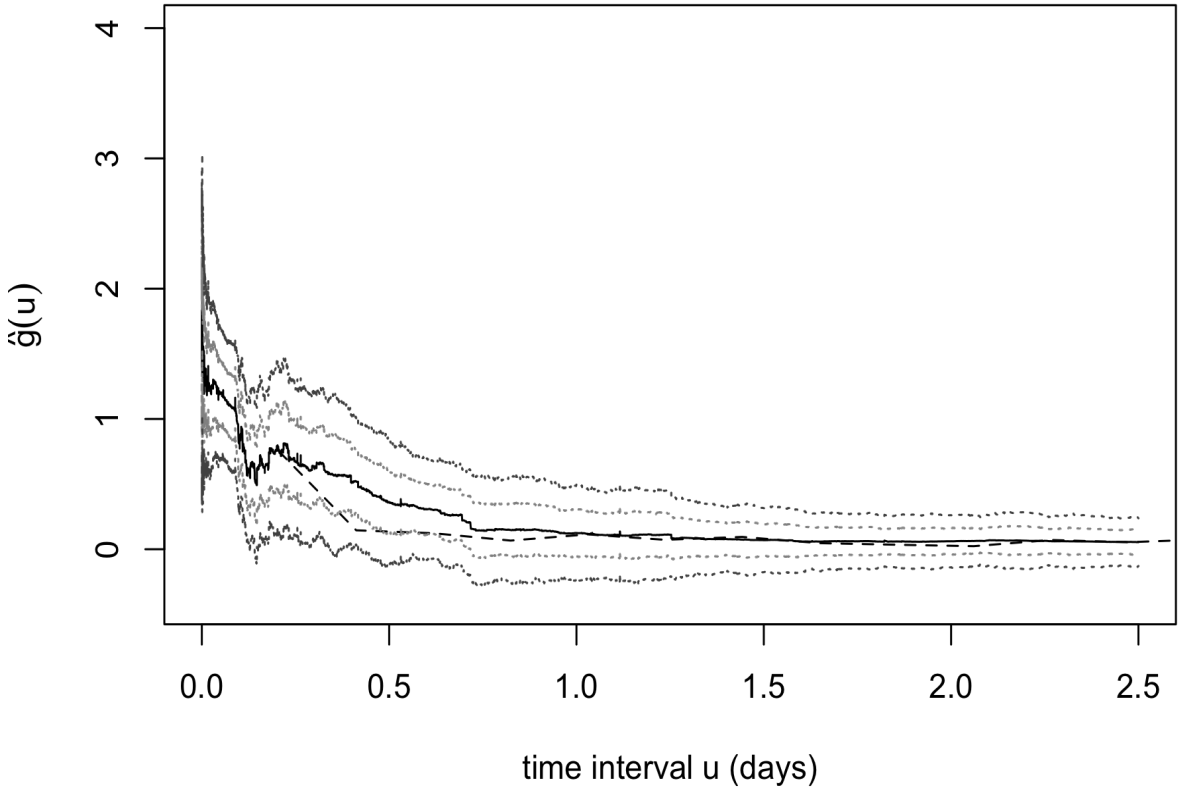
Application to Loma Prieta earthquake data.

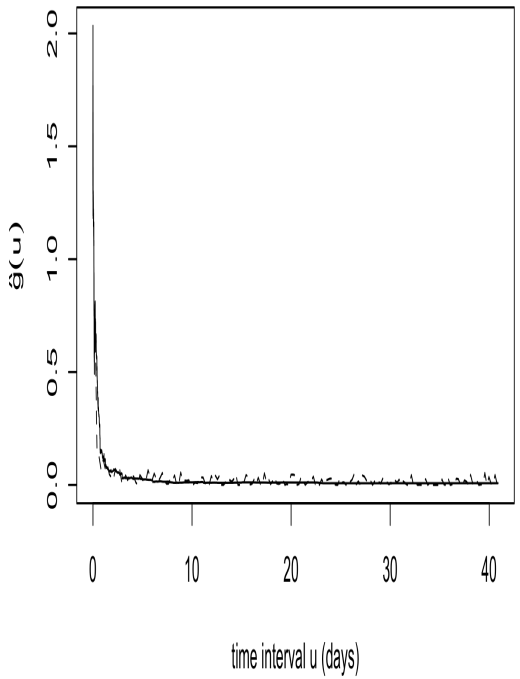
Estimated triggering function for Loma Prieta seismicity M ≥ 3, 10/16/1989 to 1/17/1990.

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 US plague data.

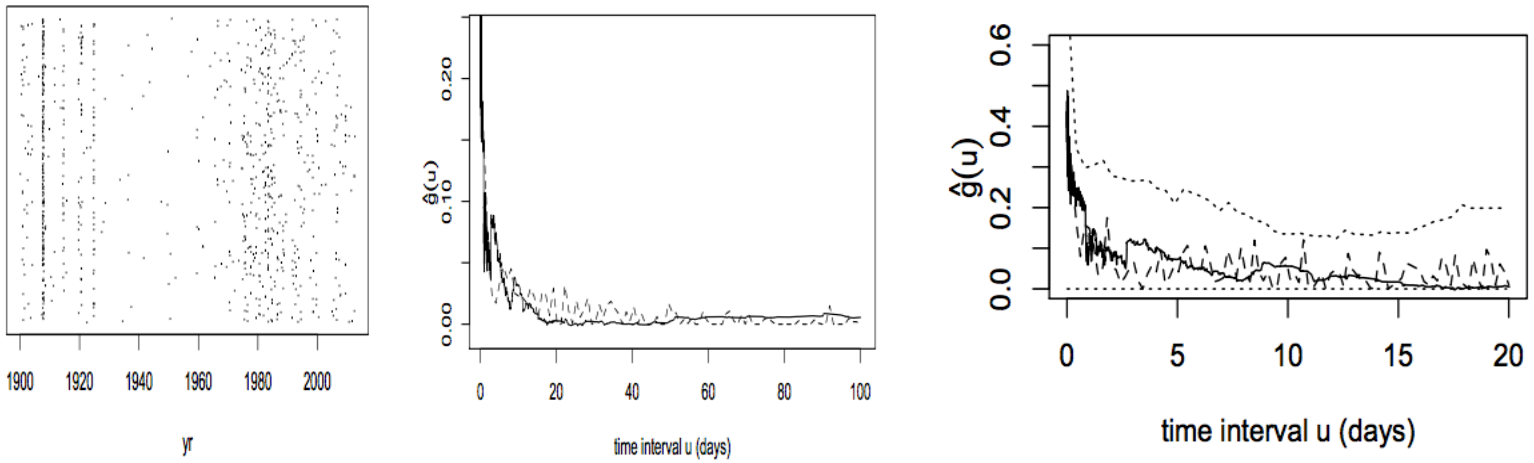


Figure 4: (a) Onset dates of reported and confirmed occurrences of plague in the United States from 1900-2012, according to data from the CDC. The y-coordinates are scattered uniformly at random on the y-axis for ease of visualization. (b) Estimated triggering function, $\hat{g}$, for the reported onset times of U.S. plague cases. (c) Estimated triggering function $\hat{g}$, for U.S. plague data, for intervals up to 20 days. In (b) and (c), the solid curves correspond to equation (9), the dashed curves result from the method of Marsan and Lengliné (2008), and the dotted curves are the middle 95% range for $\hat{g}$ from equation (9) resulting from simulating Hawkes models where the true triggering function is that estimated from the data using equation (9).