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 was going to have 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 day07.r.

3. 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) $5 \exp(\mu) + 5 \log(\mu)$.

3. Problems.

Suppose you observe a Poisson process with rate μ on the space-time window $[0,1]\text{x}[0,1]$ 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
$\underbrace{[0,1]\times[0,1]}_{S} \times \underbrace{[0,10]}_{T}$,

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
$\underbrace{[0,1]\times[0,1]}_{S} \times \underbrace{[0,10]}_{T}$,

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.

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

Alternatively, 60% of the pts are expected to be triggered, so 40% are background, and we expect 20 background points, so 20 = 40% of x, so x = 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
$\underbrace{[0,1]\text{x}[0,1]}_{S}$ x $\underbrace{[0,10]}_{T}$,

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.

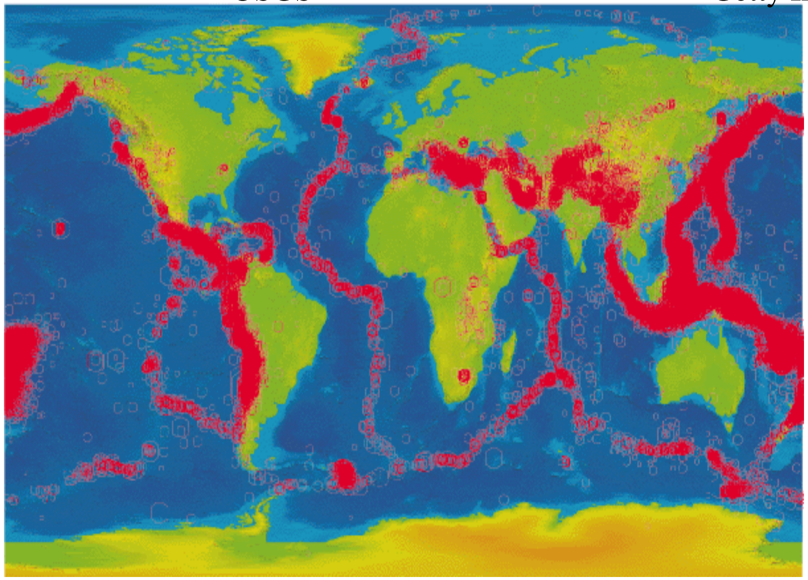
Suppose you observe a Hawkes process with conditional intensity
$\lambda(t,x,y) = 2 + 0.6 \int f(t-t')\, g(x-x',y-y')\, dN(t',x',y')$, on the space-time window
[0,1]x[0,1] x [0,10],
S T,
where $f(t) = 4\exp(-4t)$, and
$g(x,y) = 3 \exp(-3r) / (2\pi r)$, where $r = \sqrt{(x^2+y^2)}$.

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

Applications to earthquakes and US plague.

USGS

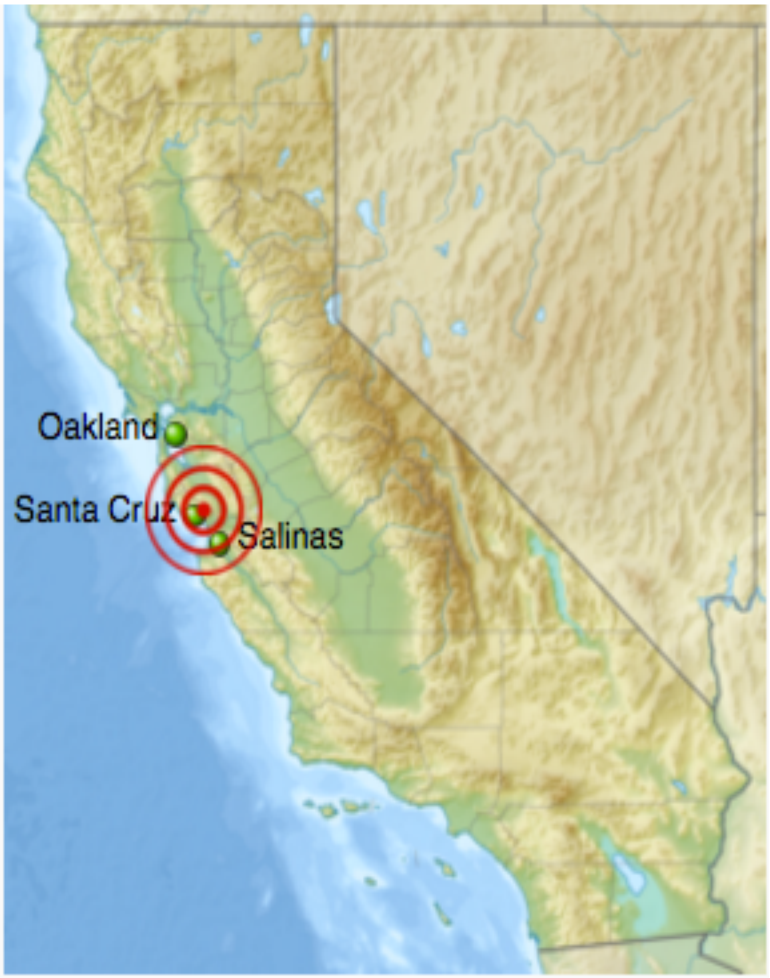


Getty Images

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)

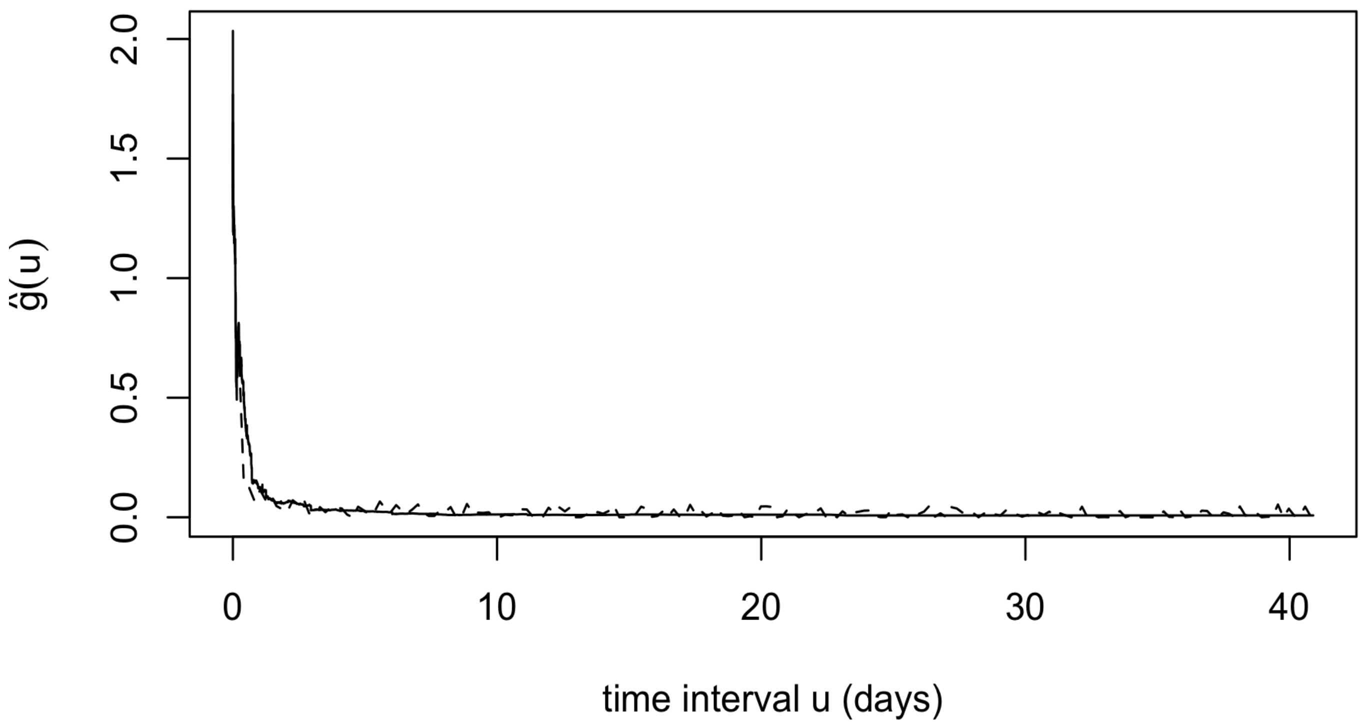
Application to Loma Prieta earthquake data. (SCEC.ORG)

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.



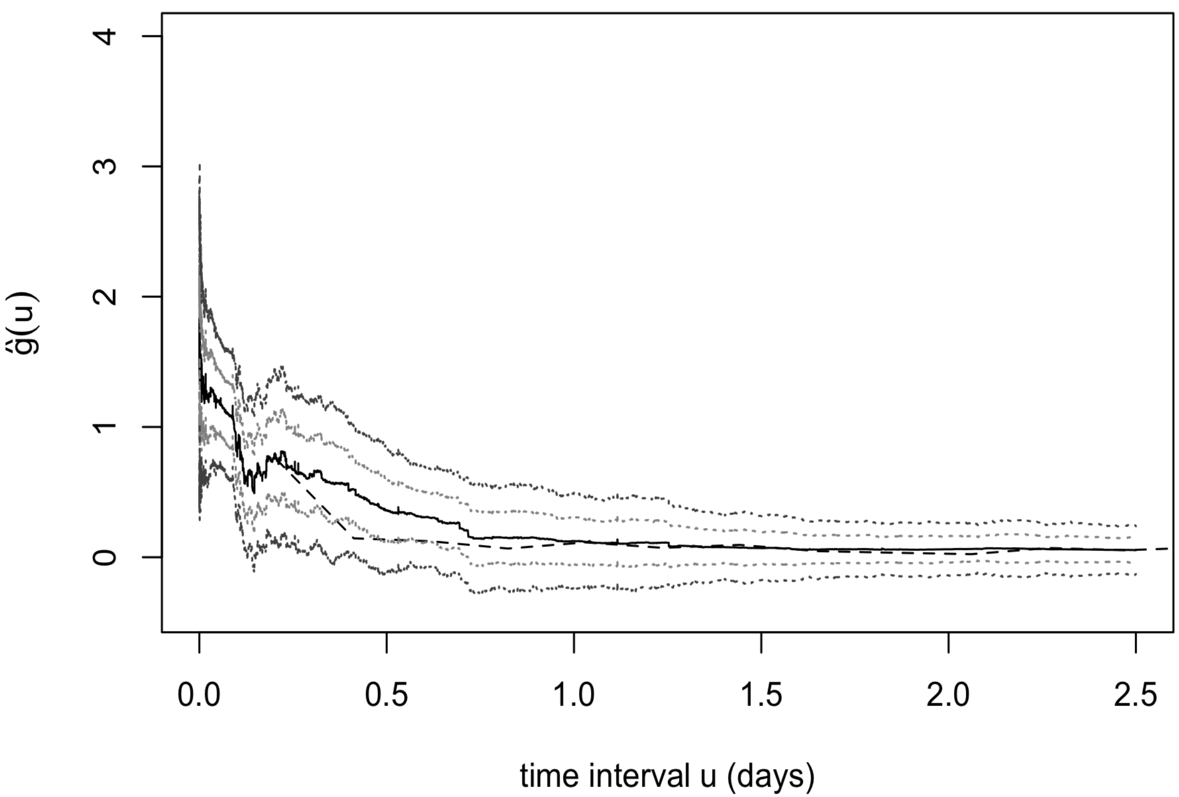
Application to Loma Prieta earthquake data.

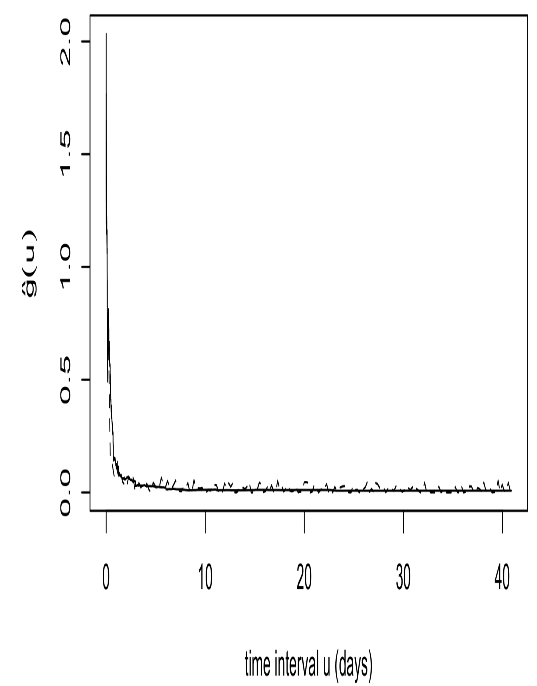
Estimated triggering function for Loma Prieta seismicity M ≥ 3, 10/16/1989 to 1/17/1990.

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.

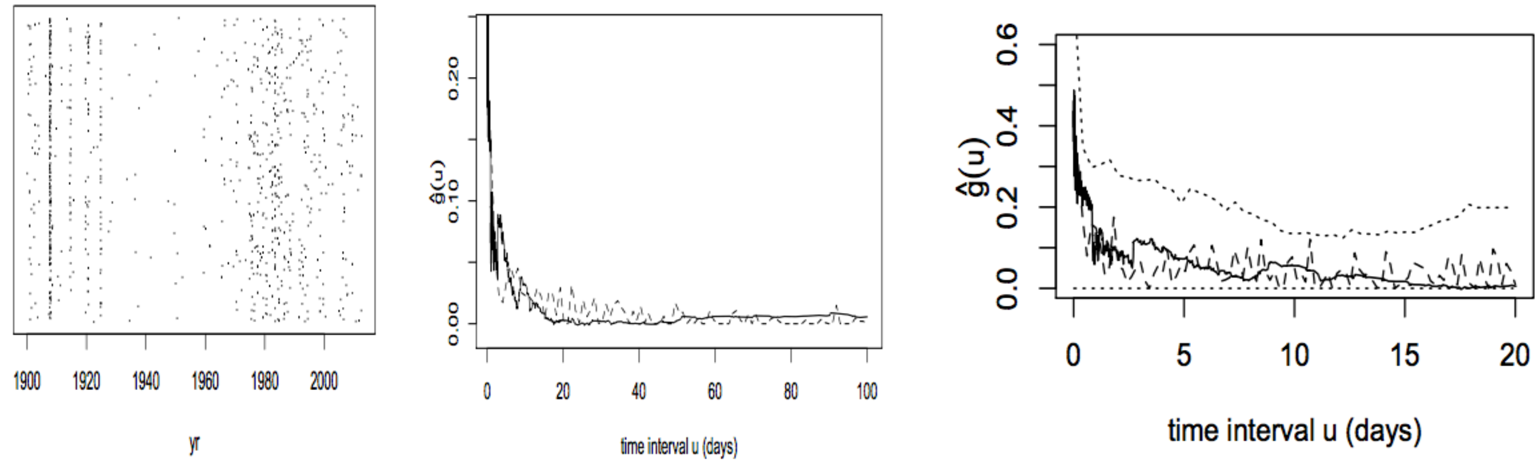


Figure 4: (a) Onset dates of reported and confirmed occurrences of plague in the United States from 1900-2012, according to data from the CDC. The y-coordinates are scattered uniformly at random on the y-axis for ease of visualization. (b) Estimated triggering function, $\hat{g}$, for the reported onset 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).