Statistics 222, Spatial Statistics.

Outline for the day:

1. Variance of compound Poisson.
2. Hawkes process.
3. ETAS.
4. Likelihood.
5. Maximum likelihood estimation.

1. Variance of the compound Poisson processes, from last time.

Fix B. Let M denote M(B). For a compound Poisson process,

$EN(B) = \sum E(N(B)|m) f(m)$, where the sum is from m = 0, 1, 2, ...,

$= \sum E(Z1 + Z2 + ... + Zm) f(m)$

$= \sum (m E(Z)) f(m)$

$= E(Z) \sum m f(m)$

$= E(Z) E(M) = c|B| E(Z)$.

$E(N(B)^2) = \sum E(N(B)^2|m) f(m)$

$= \sum E(Z1 + Z2 + ... + Zm)^2 f(m)$

$= \sum (mE(Z^2) + (m^2-m) E(Z)^2)f(m)$

$= E(Z^2)\sum mf(m) - E(Z)^2 \sum m f(m) + E(Z)^2 \sum m^2 f(m)$

$= E(Z^2) E(M) – E(Z)^2 E(M) + E(Z)^2 E(M^2)$

$= V(Z)E(M) + E(Z)^2 E(M^2)$.

So $V(N(B)) = E(N(B)^2) – (E(N(B)))^2$

$= V(Z)E(M) + E(Z)^2 E(M^2) - E(M)^2 E(Z)^2$

$= V(Z) E(M) + E(Z)^2 (E(M^2) – E(M)^2)$

$= V(Z) E(M) + E(Z)^2 V(M)$.

M is Poisson, so $E(M) = V(M) = c|B|$, so

$V(N(B)) = c|B| (V(Z) + E(Z)^2) = c|B| E(Z^2) \geq EN(B)$, since $E(Z^2) \geq E(Z)$.

2. Hawkes process.

A Hawkes process or *self-exciting* process has conditional intensity

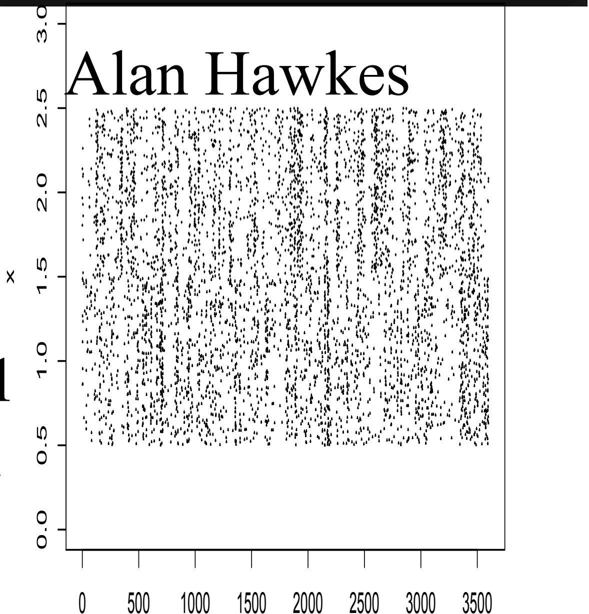
$$\lambda(t,x,y) = \mu(x,y) + \kappa\int_{t'<t} g(t-t',x-x',y-y')\, dN(t',x',y')$$

$$= \mu(x,y) + \kappa\sum_{\{t',x',y':\ t' < t\}} g(t-t',x-x',y-y').$$

g is called the *triggering function* or *triggering density* and κ is the *productivity*.

If g is a density function, then κ is the expected number of points triggered directly by each point. Each background point, associated with $\mu(x,y)$, is expected to generate $\kappa + \kappa^2 + \kappa^3 + \ldots = 1/(1-\kappa) - 1$ triggered points, so the exp. fraction of background pts is $1-\kappa$.

Alan Hawkes



3. ETAS process.

An *Epidemic-Type Aftershock Sequence (ETAS)* process is a marked version of the Hawkes process, where points have different productivities depending on their magnitudes. Ogata (1988, 1998) introduced

$$\lambda(t,x,y)=\mu(x,y) + \sum_{\{t',x',y':\ t' < t\}} g(t-t',x-x',y-y')h(m'),$$

where $\mu(x,y)$ is estimated by smoothing observed large earthquakes,
$h(m) = \kappa\, e^{\alpha(m-m0)}$,
where m0 is the catalog cutoff magnitude,
and $g(t,x,y) = g_1(t)\, g_2(r^2)$, where $r^2 = ||(x,y)||^2$,
and g_1 and g_2 are power-law or *Pareto* densities,
$g_1(t) = (p-1)\, c^{p-1}\, (t+c)^{-p}$. $g_2(r^2) = (q-1)\, d^{q-1}\, (r^2+d)^{-q}$.
An alternative is where g2 is exponential or sum of exponentials.

Yosihiko Ogata

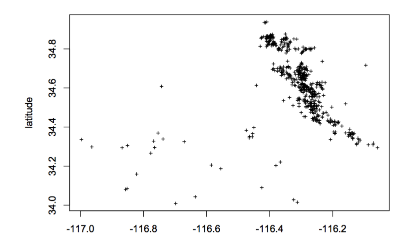


Figure 3: Recorded epicenters of Hector Mine $M \geq 3.0$ earthquakes from 10/16/1999 to 12/23/2000, from SCSN.

4. Likelihood.

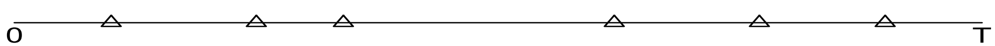
For iid draws $t_1, t_2, \ldots, t_n$, from some density $f(\theta)$, the likelihood is simply
$L(\theta) = f(t_1 ; \theta) \times f(t_2 ; \theta) \times \ldots \times f(t_n ; \theta)$
$= \prod f(t_i ; \theta)$.

This is the probability density of observing $\{t_1, t_2, \ldots, t_n\}$, as a function of the parameter θ.

For a stationary Poisson process with intensity $\lambda(\theta)$,
on $[0,T]$, the likelihood of observing the points $\{\tau_1, \tau_2, \ldots, \tau_n\}$ is simply
$\lambda(\tau_1) \times \lambda(\tau_2) \times \ldots \times \lambda(\tau_n) \times$
$\exp\{-A(\tau_1)\} \times \exp\{-(A(\tau_2)-A(\tau_1))\} \times \ldots \times \exp\{-(A(T)-A(\tau_n))\}$,
$= \prod \lambda(\tau_i) \exp\{-A(T)\}$,
where $A(t) = \int_0^t \lambda(t)dt$.
$P\{k \text{ points in } (\tau_1, \tau_2)\}$ is $\exp(-B)\, B^k/k! = \exp(-B)$ for $k = 0$,
where $B = \int_{\tau 1}^{\tau 2} \lambda(t)dt$.

Likelihood, continued.

For a stationary Poisson process with intensity $\lambda(\theta)$,
on $[0,T]$, the likelihood of observing the points $\{\tau_1, \tau_2, \ldots, \tau_n\}$ is simply
$\lambda(\tau_1) \times \lambda(\tau_2) \times \ldots \times \lambda(\tau_n) \times$
$\exp\{-A(\tau_1)\} \times \exp\{-(A(\tau_2)-A(\tau_1))\} \times \ldots \times \exp\{-(A(T)-A(\tau_n))\}$,
$= \prod \lambda(\tau_i) \exp\{-A(T)\}$,
where $A(t) = \int_0^t \lambda(t)dt$.
$P\{k \text{ points in } (\tau_1, \tau_2)\}$ is $\exp(-B)\, B^k/k! = \exp(-B)$ for $k = 0$,
where $B = \int_{\tau 1}^{\tau 2} \lambda(t)dt$.



So the log likelihood is $\sum \log(\lambda(\tau_i)) - A(T)$.
In the spatial-temporal case, the log likelihood is simply
$\sum \log(\lambda(\tau_i)) - \int \lambda(t,x,y)\, dt\, dx\, dy$.

5. Maximum likelihood estimation.

Find $\hat{\theta}$ (= θˆ) maximizing $l(\theta) = \sum \log(\lambda(\tau_i)) - \int \lambda(t,x,y)\, dt\, dx\, dy$.

Ogata (1978) showed that the resulting estimate, θˆ, is, under standard conditions,

asymptotically unbiased, E(θˆ) → θ,

consistent, P(|θˆ - θ| > ε) → 0 as T → ∞, for any ε > 0,

asymptotically normal, θˆ $\to_D$ Normal as T → ∞,

and asymptotically efficient, min. variance anong asymptotically unbiased estimators.

Further, he showed standard errors for θˆ can be constructed using the diagonal elements of the inverse of the Hessian of L evaluated at θˆ .

```
sqrt(diag(solve(loglikelihood$hess)))
```

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. Ann. Inst. Statist. Math. 30, 243-261.

Maximum likelihood estimation continued.

The conditions of Ogata (1978) can be relaxed a bit for Poisson processes [1], and for certain spatial-temporal process in general [2].

Even if the process is not Poisson, under some circumstances [3] the parameters governing the unconditional intensity, $E\lambda$, can be consistently estimated by maximizing $L_P(\theta) = \sum \log(E\lambda(\tau_i)) - \int E\lambda(t,x,y)\, dt\, dx\, dy$. Basically pretend the process is Poisson.

Suppose you are missing some covariate that might affect λ. Under general conditions, the MLE will nevertheless be consistent, provided the effect of the missing covariate is small [4].

[1] Rathbun, S.L., and Cressie, N. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. Adv. Appl. Probab. 26, 122–154.

[2] Rathbun, S.L., (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *JSPI* 51, 55–74.

[3] Schoenberg, F.P. (2004). Consistent parametric estimation of the intensity of a spatial-temporal point process. *JSPI* 128(1), 79--93.

[4] Schoenberg, F.P. (2016). A note on the consistent estimation of spatial-temporal point process parameters. *Statistica Sinica,* 26, 861-879.

Maximum likelihood estimation continued.

λ is completely separable if $\lambda(t,x,y;\theta) = \theta_3\, \lambda_0(t;\theta_0)\, \lambda_1(t,x;\theta_1)\, \lambda_2(t,y;\theta_2)$.
Suppose N has marks too. λ is separable in mark (or coordinate) *i* if
$\lambda(t, x, y, m_1, m_2, ..., m_k ; \theta) = \theta_2\, \lambda_i(t, m_i; \theta_i)\, \lambda_{-i}(t, x, y, m_{-i} ; \theta_{-i})$.

Suppose you are neglecting some *mark* or coordinate of the process. Under some conditions, the MLE of the other parameters will nevertheless be consistent [1].

In maximizing $L(\theta) = \sum \log(\lambda(\tau_i)) - \int \lambda(t,x,y)\, dt\, dx\, dy$,
it is typically straightforward to compute the sum, but the integral can be tricky esp. when the conditional intensity is very volatile. One trick noted in [2] is that, for a Hawkes process where $\lambda(t,x,y) = \mu(x,y) + \kappa \sum_{\{t',x',y':\ t' < t\}} g(t-t',x-x',y-y')$, where g is a density, and $\int \mu(x,y)dxdy = \mu$,

$$\int \lambda(t,x,y)\, dt\, dx\, dy = \mu T + \kappa \int \sum g(t-t',x-x',y-y')\, dt\, dx\, dy$$
$$= \mu T + \kappa \sum \int g(t-t',x-x',y-y')\, dt\, dx\, dy$$
$$\sim \mu T + \kappa N.$$

[1] Schoenberg, F.P. (2016). A note on the consistent estimation of spatial-temporal point process parameters. *Statistica Sinica,* 26, 861-879.
[2] Schoenberg, F.P. (2013). Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America,* 103(1), 601-605.

6. Questions.

The difference between ETAS and a Hawkes process is ...

a) an ETAS process is more strongly clustered.
b) the points of an ETAS process occur at different locations.
c) the points of an ETAS process have different productivities.
d) the points of an ETAS process have different triggering functions.

Questions.

The difference between ETAS and a Hawkes process is ...

a) an ETAS process is more strongly clustered.
b) the points of an ETAS process occur at different locations.
c) the points of an ETAS process have different productivities.
d) the points of an ETAS process have different triggering functions.

Which of the following can possibly have two points within distance .01 of each other?
a) a hardcore process with $\sigma = .01$.
b) a Strauss process with $R = .01$.
c) a Matern I process with $r = .01$.
d) a Matern II process with $r = .01$.

Questions.

The difference between ETAS and a Hawkes process is ...

a) an ETAS process is more strongly clustered.
b) the points of an ETAS process all occur at different locations.
c) the points of an ETAS process all have different productivity.
d) the points of an ETAS process all have different triggering functions.

Which of the following can possibly have two points within distance .01 of each other?

a) a hardcore process with $\sigma = .01$.
b) a Strauss process with R = .01.
c) a Matern I process with r = .01.
d) a Matern II process with r = .01.

Code.

```
install.packages("spatstat")
library(spatstat)

## STRAUSS process
z = rStrauss(100,0.7,0.05)
plot(z, pch=2,cex=.5)

## HARDCORE process
z = rHardcore(100,0.05)
plot(z, pch=2,cex=.5)

## MATERN(I).
z = rMaternI(20,.05)
plot(z, pch=2,cex=.5)
```

Code.

```
## MATERN(II)
z = rMaternII(100,.05)
plot(z,pch=2,cex=.5)

## HAWKES.
install.packages("hawkes")
library(hawkes)
lambda0 = c(0.2,0.2)
alpha = matrix(c(0.5,0,0,0.5),byrow=TRUE,nrow=2)
beta = c(0.7,0.7)
horizon = 3600
h = simulateHawkes(lambda0,alpha,beta,horizon)
plot(c(0,3600),c(0,3),type="n",xlab="t",ylab="x")
points(h[[1]],.5+runif(length(h[[1]])),pch=2,cex=.1)
points(h[[2]],1.5+runif(length(h[[2]])),pch=3,cex=.1)
```