1. Examples of spatial-temporal point processes

1a. A point process is a random collection of points falling in some metric space. For a spatial-temporal point process, the metric space is a portion of space-time, $S = R^d \times R$.

1b. Examples include incidence of disease, sightings or births of a species, occurrences of fires, earthquakes, lightning strikes, tsunamis, or volcanic eruptions.

1c. For a marked point process, each point has some mark or random variable associated with it.

2. Characterizations of point processes.

2a. STOCHASTIC PROCESS.
Point processes on the line were originally characterized as examples of stochastic processes on the line, that are:
* Non-decreasing,
* Z+ valued. non-negative integer value
So define a point process as any non-decreasing, Z+ valued stochastic process.
N(x,y) = the number of points (x_1, y_1) with $x_1 \leq$ x and $y_1 \leq$ y.

2b. A LIST OF POINTS.
For any finite collection of points, simply as a finite list of points, N = { x_1, x_2, ... }.

2c. RANDOM MEASURE.
A Z+ valued random measure includes a wide range of processes on the line and extends readily to space-time.
The measure N(A) represents the number of points falling in the region A of space-time.

3. Integration.

N(B) = $\int_B$ dN is the number of points in B.
$\int t \int x \int y$ f(t,x,y) dN is simply $\sum_i f(t_i, x_i, y_i)$. $dN = \lambda(t,x,y)dt\,dx\,dy$

Exercise 1

1. Suppose the spatial-temporal point process *N* has points at
time 1.2, x=2, y=3.
time 2.4, x=3, y=0.5.
time 8.7, x=2, y=1.
Let *B* = [0, 10] (time) x **[1.5, 2.5]** x [0, 5].

What is $\int_B$ dN? =# of points in domain B, 2.

2. Suppose the spatial-temporal point process *N* has points at
time 1.2, x=2, y=3.
time 2.4, x=3, y=0.5.
time 8.7, x=2, y=1.
Let *B* = [0,10] (time) x [0,5] x [0,5].
What is $\int_B (t+x^2 y)$ dN? $\int t \int x \int y$ f(t,x,y) dN = $\sum_i f(t_i, x_i, y_i)$.

Sum up all points in domain B using the function $t+x^2y$: **13.2 + 6.9 + 12.7 = 32.8.**

4. Simple and orderly point processes.

4a. A point process is *stationary* (or *homogeneous)* if, for any k and any collection of measurable sets, $B_1, B_2, ..., B_k$) (*intervals*) , the joint distribution of the collection $\{N(B_1 + \Delta), N(B_2 + \Delta), ..., N(B_k + \Delta)\}$ doesn't depend on Δ.

Stationary = no region of data has more features than other regions

4b. A point process is called *simple* if all the points are distinct,
i.e. P(there are indices i and j where $\tau_i = \tau_j$) = 0.

Simple = all points happen in different time and location, time cannot overlap.

4c. A point process is *orderly* if for any time t and any spatial interval B,
$\lim_{\Delta t \to 0}$ P(N([t, t+ Δt) × B)>1)/(Δt |B|)=0.

Orderly = no cluster of points, no a bunch of points in a time, no same prob. mass of two points.

4d. **If *N* is simple and stationary, then it is orderly.**

All points distinct and λ same everywhere (homogenous), no cluster of points.

4e. The times $\{t_1, t_2, ..., t_n\}$ are sometimes said to form the *ground* process.
N has a *simple ground process* if all the *times* are distinct, with prob. 1.

5. Conditional intensity, λ.

5a. Fix any space-time location (t, x, y). λ(t,x,y) is the limiting expected rate of accumulation of points around (t,x,y), given all points prior to t.

$\lambda(t,x,y) = \lim_{\Delta t, \delta \to 0} E(N([t, t+\Delta t) \times B(x, y, \delta)) \mid H_t) / (\Delta t \, \pi \delta^2)$, where B(x,y, δ) is a circle of radius δ around (x,y), and H_t is the history of the process up to but not including time t.

➔ If N is orderly, then $\lambda(t,x,y) = \lim_{\Delta t, \delta \to 0} P(N([t, t+\Delta t) \times B(x, y, \delta)) > 0 \mid H_t) / (\Delta t \, \pi \delta^2)$.

Note that λ is random, depending on what points have occurred previously, and might be different with every realization.

5b. Fix some spatial interval, B. The integral of λ, $A(t) = \int t \iint_B \lambda(t,x,y) \, dx \, dy \, dt$, is called the *compensator* of *N*. For any B, N(t, B) – A(t) is a *martingale*.

➔ define as a derivative of A, if it exists. $A(t) = \int t \iint_B \lambda(t,x,y) \, dx \, dy \, dt \Rightarrow \lambda = \frac{\partial A(t)}{\partial t \, \partial x \, \partial y}$

5c. λ is *predictable*. Predictable = not know # points happened, but know # points expected to happen, rate predictable. Predictable: slight generalization of left-continuous.

5d. If for any (t,x,y), the limit doesn't exist or is ∞, then λ doesn't exist.
λ uniquely determines the distributions of any simple point process.

6. Poisson processes. poisson random variable: rare prob. among large population

6a. If *N* is a simple point process with conditional intensity λ, where λ does not depend on what points have occurred previously, then *N* is a *Poisson process* (most basic models for point

processes).
For such a process, for any set B, N(B) has a Poisson distribution. (assume that B is measurable for N(B)), $P(N(B) = k) = e^{-A}A^k/k!$, for k = 0, 1, 2, ...,
where $A = \int_B \lambda(t,x,y)\, dt\, dx\, dy$ A is the total rate (expected points) of set B

6b. The mean of N(B) is A and the variance is also A. E(N(B)) = A , var(N(B)) = A
Note that a Poisson process does not have to be stationary.

6c. If λ is constant for all t,x,y, and *N* is simple, then *N* is a stationary Poisson process, and is sometimes called *completely random.*

6d. A stationary poisson process with $\lambda(t,x) = 2.5$ on [0,1] x [0,10]. completely random
2.5 points per subspace, λ same everywhere.
A Poisson process with $\lambda(t,x) = 1.5 + 10t + 2x$.
As t and x go up, rate higher, not depend on what happened before.

6e. The key thing about poisson processes is their complete independence.
For a Poisson process *N*, $N(B_1)$ and $N(B_2)$ are independent for any disjoint sets B_1 and B_2.
complete independence: any subsets of points are indep of others, completely indep in any dimensions.

Exercise 2
1. Suppose *N* is generated as follows. For each integer i = 1,2,..., *N* has a point at (i, i, i) with probability 1/i, independently of the other points, and *N* has no other points.
What is $\lambda(2,2,2)$?
a) 2. b) 1. c) 1⁄2. **d) does not exist.**

$$\lambda(t,x,y) = \lim_{\Delta t,\, \delta \to 0} E\,(N([t, t+\Delta t) \times B(x, y, \delta)) \mid H_t) / (\Delta t\, \pi \delta^2)$$

λ = expected # points per unit space; a point there or no point, with prob. of ½, so E = ½
Size of the region ($\Delta t\, \pi \delta^2$) goes to 0 for the specific point (2, 2, 2). ½ / 0 = infinity.
If for any (t,x,y), the limit doesn't exist or is ∞, then λ doesn't exist.
Usually, λ on some specific point, then λ doesn't exist.

2. Suppose *N* is generated as follows. For each integer i = 1,2,..., *N* has a point at (i, i, i) with probability 1/i, independently of the other points, and *N* has no other points.
N is
a) simple but not orderly. b) orderly but not simple.
c) simple and orderly. d) neither simple nor orderly. Simple = distinct, orderly = no cluster

All points are well spread out, no two points overlapped.

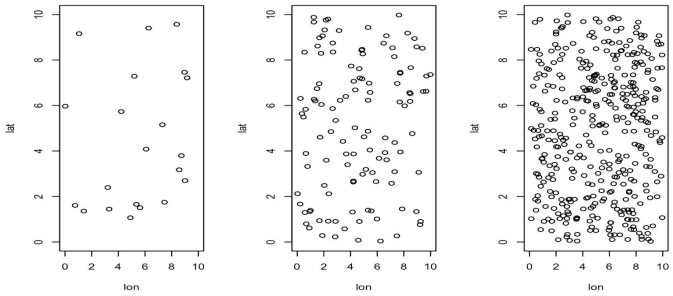
3. Suppose *N* is a Poisson process with $\lambda(t,x,y) = 1.5 + 10t + 2x$ on B = [0,1] x [0,10] x [0,1].
What is EN(B)? expected points in domain B

$\int_B \lambda(t,x,y)\, dt\, dx\, dy = 1.5(10) + 10(10)(1^2)/2 + 2(1)(10^2)/2 = 15+50+100=165.$

EN(B) = A = $\int_B \lambda(t,x,y)\, dt\, dx\, dy$, take integral of λ over time and space.

7. Mixed Poisson processes.

7a. Suppose $\lambda(t,x,y) = c$, where c is a random variable. For example, c might be Poisson or exponential, or half normal, or something constrained to be positive. Then conditional on c, N(B) is Poisson distributed. Then N is a *mixed Poisson process.*



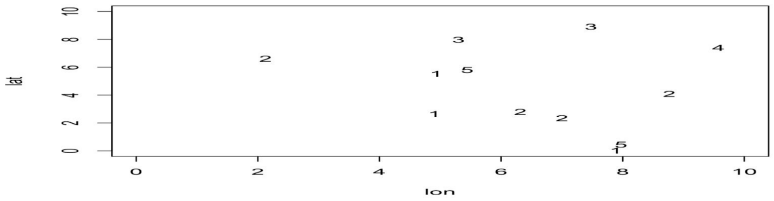
c fixed, rate keeps the same every subspace.

E(N(B)|c) = V(N(B)|c) = c|B|, *rate* × *|size of region B|*
but unconditionally, N(B) is not Poisson distributed.

7b. E(N(B)) = $\int$ E(N(B)|c) f(c) dc = $\int$ c|B| f(c) dc = |B|E(c). E(N(B)) = A
E($N(B)^2$) = $\int$ E $(N(B)|c)^2$ f(c) dc = $\int$ [c^2 $|B|^2$ + c|B|] f(c) dc = $|B|^2$ E(c^2) + |B|E(c),
E($N(B)^2$) = $A^2 + A = |B|^2$ E(c^2) + |B|E(c)
so V(N(B)) = $|B|^2$ E(c^2) + |B|E(c) - $|B|^2$ $[E(c)]^2$ = E(N(B)) + $|B|^2$ V(c).
So, V(N(B)) ≥ E(N(B)). V ≥ E for mixed poisson process, V = E for poisson process

8. Compound Poisson process. e.g. bread available in supermarket

8a. Suppose N is not simple (points not distinct, overlap). First generate a stationary Poisson process M with intensity c, and then for each point τ_i of M, N will have some non-negative number Z_i of points right at τ_i, where Z_i are all iid and independent of M. Then N is a compound Poisson process.
Start with M (location of markets), each point m_i generated, there are Z points come along.
For a compound Poisson process, again the variance ≥ the mean.



8b. Variance of the compound Poisson process.
Fix B. Let M denote M(B). For a compound Poisson process,
EN(B) = Σ $E(N(B)|m)\, f(m)$, where the sum is from m = 0, 1, 2, …
= Σ $E(Z_1 + Z_2 + \ldots + Z_m)\, f(m)$ $N(B) = (Z_1 + Z_2 + \ldots + Z_m) = \Sigma_i Z_i$
= Σ $(m\, E(Z))\, f(m)$ start with stationary poisson process M with rate c, for

$= E(Z) \Sigma\, m f(m)$ each point m, put random points Z_i, Z_i's have the same dist.
$= \mathrm{E(Z)\ E(M) = c|B|\ E(Z).}$
$\mathrm{E}(N(B)^2) = \Sigma\, E(N(B)^2|m) f(m)$
$= \Sigma\, E(Z_1 + Z_2 + \ldots + Z_m)^2 f(m)$
$= \Sigma\, [\, mE(Z^2) + (m^2 - m)\, E(Z)^2]\, f(m)$
$= E(Z^2) \Sigma\, m f(m) - E(Z)^2\, \Sigma\, m f(m) + E(Z)^2 \Sigma\, m\ f(m) + E(Z)^2\, \Sigma\, m^2 f(m)$
$= E(Z^2)\, E(M) - E(Z)^2 E(M) + E(Z)^2 E(M^2)$
$= V(Z)\, E(M) + E(Z)^2 E(M^2)$

So $\mathrm{V(N(B)) = E}(N(B)^2) - [EN(B)]^2$
$= V(Z)\, E(M) + E(Z)^2 E(M^2) - E(Z)^2 E(M)^2$
$= V(Z)\, E(M) + E(Z)^2 [E(M^2) - E(M)^2]$
$= V(Z)\, E(M)\ +\ E(Z)^2\ V(M)$

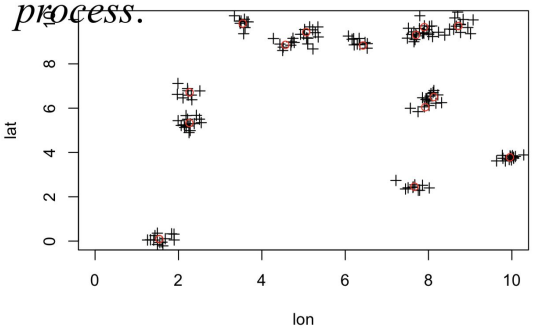
M is Poisson, so $\mathrm{E(M) = V(M) = c|B|}$, so
$V(N(B)) = c|B| \cdot [V(Z) + E(Z)^2] = c|B|\, E(Z^2) \geq EN(B)$, since $E(Z^2) \geq E(Z)$.

9. Poisson cluster processes. e.g. earthquakes

Another extension of the Poisson process is the Poisson cluster process.
First generating *parent* points *M* according to a Poisson process. Then for each parent point τ_i, generate some random number Z_i of offspring points, and these offspring points are scattered spatially and temporally, independently of each other, with some distribution centered at τ_i.
Let *N* be the collection of just the offspring, not the parents. *N* is called *Poisson cluster process*.
Usually *M* is assumed *stationary* Poisson.
Parent points assumed stationary poisson; offspring points are in poisson cluster process

➔ In the particular case where the Z_i are iid Poisson random variables independent of *M*, the process is called a *Neyman-Scott cluster process*.

process.

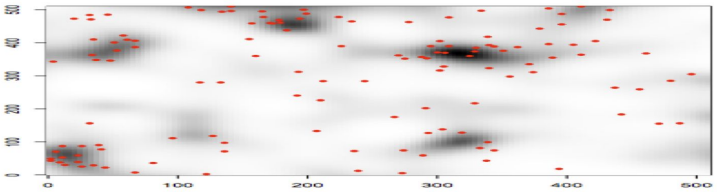


Parent points uniformly scattered, as red dots, like main shocks.
Child points cluster around parents, aftershocks generated randomly.

10. Cox process.

Suppose generate a stochastic process $\lambda(t,x,y)$ such that $\lambda(t,x,y) \geq 0$ for all t, x, and y. Then let *N* be a Poisson process with intensity $\lambda(t,x,y)$.
So $\lambda(t,x,y)$ can be random, but conditional on λ, *N* is a Poisson process.
So *N* is a *Cox* process or equivalently a *doubly stochastic Poisson process.*

Cox processes arise in practice when modeling events depending on some other random phenomenon. For instance, the points of N might be the times and locations of flu epidemics, which might depend on the temperature and might in turn be modeled as evolving stochastically.



λ like temperature (grey dots) is random.

11. Gibbs process. e.g. look for some specific species in forest

For any finite collection $(\tau_1, \tau_2, ..., \tau_n)$ of points in space-time, if the joint density is $C(\theta)\, exp[-\theta \{\sum_i \Psi_1(\tau_i) + \sum_{i,j} \Psi_2(\tau_i, \tau_j)\}]$, then N is a Gibbs process.

Often $\Psi_2(\tau_i, \tau_j)$ can be written $\Psi(r)$, where $r = |x_i - x_j|$. interaction term

$\Psi_2(\tau_i, \tau_j)$ high when points are not close, density small, making clusters of points less likely.

$\Psi_2(\tau_i, \tau_j)$ low, density high.

11a. When $\Psi(r) = 0$, there are no interactions, and the process is an inhomogeneous Poisson process with intensity $\Psi_1(x)$. mean term

11b. $\Psi(r) = -log[1 - e^{-(r/\sigma)^2}]$ defines a *soft-core* model. Weak repulsion.

$\Psi_2(r)$ is called the *interaction potential*. Joint density small, less likely

11c. $\Psi(r) = \infty$ for $r \leq \sigma$ σ some #

$= 0$ for $r > \sigma$ can't have 2 points super close together

defines a *hard-core* process.

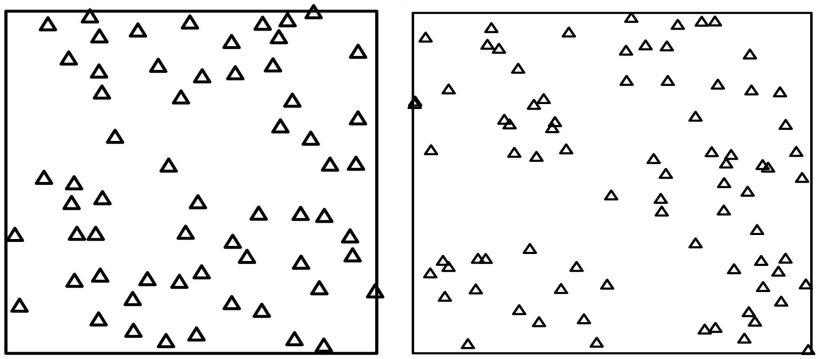
11d. $\Psi(r) = (\sigma/r)^n$ is an intermediate choice between the soft-core and hard-core models.

Unlikely to get many points close together, but a little possible.

11e. Strauss process. $\Psi_1(r) = \alpha$, and $\Psi_2(r) = \beta$, for $r \leq R$, β some positive #

$\Psi_2(r) = 0$, for $r > R$.

less likely to have paired points, rare compared with poisson process; unlikely to have 2 points together, but possible



Intermediate choice strauss process

12. Matern process.

12a. The Matern(I) process is generated as e.g. large trees can't really close to others

a) Generate M according to a stationary Poisson process. must be spread out, well-spaced

b) Let N be all points of M that are not within some fixed distance r of any other point of M.

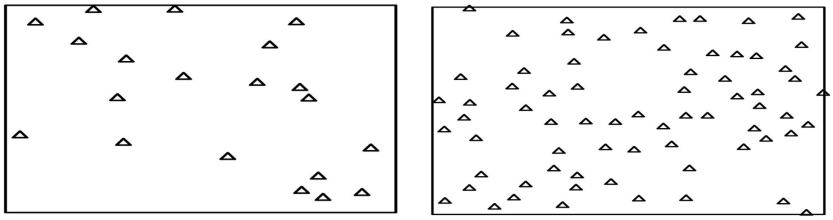
Two points close, get rid of both

12b. The Matern(II) process is generated a bit differently.

a) Generate points $\tau_1, \tau_2, \ldots$ according to a stationary Poisson process.

b) For i = 1, 2, ..., keep point i if there is no *previous* kept point τ_j with $|\tau_i - \tau_j| \le r$.

Two points close, get rid of one of them



Matern I Matern II

Exercise 3

1. A mixed Poisson process is a Cox process where

a. $\lambda = E(\lambda)$ in every realization.

b. $\lambda(t,x,y) = \lambda(t',x',y')$, for any locations (t,x,y) and (t',x',y').

c. The cluster sizes are Poisson distributed with mean λ.

d. $\lambda = 1$.

a. means λ is a constant, so *N* is a stationary Poisson process.

d. Also defines a stationary Poisson process, with rate 1.

Mixed process, conditional on c, N(B) poisson dist.; roll a dice to generate λ, r.v $\lambda(t, x, y) = c$; $\lambda = c$, same everywhere, same in every subspace.

Cox process, conditional on λ(random), N poisson process; λ more varying in space, spatially variate, λ not the same for every subspace.

The difference of these two processes is λ.

13. Hawkes process.

A Hawkes process or *self-exciting* process branching process with branching behavior

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

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

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

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

percent/portion of background points out of total points is 1- k; percent/portion of triggered points out of total points is k.
background rate μ(x,y) only depends on location, like mainshocks, mainshocks trigger aftershocks, aftershocks trigger aftershocks.

14. ETAS process.

An *ETAS* process is a marked version of the Hawkes process, where points have different productivities depending on their magnitudes.

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

where μ(x,y) is estimated by smoothing observed large earthquakes, $h(m) = k\, e^{\alpha(m-m_0)}$

where m_0 is the catalog cutoff magnitude, and $g(t,x,y) = g_1(t)g_2(r^2)$, where $r^2 = \|(x,y)\|^2$

15. Likelihood.

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

16. Maximum likelihood estimation.

16a. Find $(\widehat{\theta})$ maximizing $l(\theta) = \sum \log(\lambda(\tau_i)) - \int \lambda(t,x,y)\, dt\, dx\, dy$.

$\widehat{\theta}$, under standard conditions, is asymptotically unbiased, $E(\widehat{\theta}) \to \theta$,

consistent, $P(|\widehat{\theta} - \theta| > \varepsilon) \to 0$ as $T \to \infty$, for any $\varepsilon > 0$,

asymptotically normal, $\widehat{\theta} \to_D$ Normal as $T \to \infty$,

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

16b. Even if the process is not Poisson, under some circumstances 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.

16c. λ is completely separable if $\lambda(t,x,y;\ \theta) = \theta_3\lambda_0(t;\ \theta_0)\cdot\lambda_1(t,\ x;\ \theta_1)\cdot\lambda_2(t,\ y;\ \theta_2)$.
Suppose N has marks too. λ is separable in mark (or coordinate) *i* if
$\lambda(t, x, y,\ m_1, ..., m_k;\ \theta) = \theta_2\ \lambda_i(t,\ m_i;\ \theta_i)\ \lambda_{-i}(t,\ x,\ y,\ m_{-i}\ ;\ \theta_{-i})$

16d. In maximizing $l(\theta) = \sum log(\lambda(\tau_i)) - \int \lambda(t,x,y)\, dt\, dx\, dy$,
Typically straightforward to compute the sum, but the integral can be tricky esp. when the conditional intensity is very volatile.
For a Hawkes process where $\lambda(t,x,y) = \mu(x,y) + k\,\Sigma_{\{t',x',y':\, t'<t\}}\, g(t-t',x-x',y-y')$, where g is a density, and ∫μ(x,y)dxdy = μ, how far away from space and time

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

Proof: assume the space-time region B = [0,T] x S.

$\lambda(t,x,y) = \mu\,\rho(x,y) + K\sum_{i:t_i<t} g(t-t_i,x-x_i,y-y_i)$, where ρ and g are densities.

$$\int_0^T \iint \lambda(t,x,y)dx\, dy\, dt = \int_0^T \iint \mu\,\rho(x,y)dx\, dy\, dt + \int_0^T \iint K \sum_{i:t_i<t} g(t-t_i,x-x_i,y-y_i)\, dxdydt$$

$= \mu\, T + \int_0^T \iint K \int_B 1_{\{t'<t\}}\, g(t-t',\ x-x',\ y-y')\, dN(t',x',y')\, dxdydt$ interchanging the integrals

$= \mu T + K \int_B \int_0^T \iint 1_{\{t'<t\}}\, g(t-t',\ x-x',\ y-y')\, dxdydt\, dN(t',x',y')$ changing coordinates, letting u = t-t', v = x-x', w = y-y',

$= \mu T + K \int_B \int_0^T \iint_{S-(x',y')} g(u,v,w)\, dudvdw\, dN(t',x',y') \sim \mu T + K \int_B (1)\, dN(t',x',y')$

$= \mu T + K N(B)$.

approximate because typically $\int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty g(u,v,w)\, du\, dv\, dw = 1$, but instead, we have

$\int_0^{T-t'} \iint_{S-(x',y')} g(u,v,w)\, dudvdw$, which is often close to 1.

Exercise 4

1. 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.

ETAS: marked version of Hawkes; different productivitives depend on magnitude

2. 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.

a. not possible, $\Psi_2(r) = \infty$ for $r \le R = .01$

b. make the two points close less likely, but still possible, $\Psi_2(r) = \beta$ for $r \le R = .01$

c,d. two points too close, get rid of one or both of them

17. Purely spatial processes, Papangelou intensity and Georgii-Zessin Nguyen formula.

17a. the Papangelou intensity, $\lambda(x,y)$, is the conditional rate of points around location (x,y), given information on everywhere else. Letting $l(\theta) = \sum \log(\lambda(\tau_i)) - \int \lambda(t, x, y)\, dt\, dx\, dy$, where $\lambda(x, y)$ is the Papangelou intensity, $l(\theta)$ is called the *pseudo-loglikelihood*.

17b. A key formula for space-time point processes is called the *martingale formula*: for any predictable function f(t,x,y), $\int t \int x \int y\, f(t,x,y)\, dN = \sum_i f(t_i, x_i, y_i)$

$E \int f(t,x,y)\, dN = E \int f(t,x,y)\, \lambda(t,x,y)\, d\mu = E \sum_i f(t_i, x_i, y_i) = E \int f(t,x,y)\, \lambda(t,x,y)\, dt\, dx\, dy$

17c. For spatial point processes the corresponding formula,

$E \int f(x,y)\, dN = E \int f(x,y)\, \lambda(x,y)\, dx\, dy$ is called the Georgii-Zessin-Nguyen formula.

When f = 1, means $EN(B) = E \int 1\, d\mu$.

Exercise 5

1. Suppose N is a Poisson process with intensity $\lambda(t,x,y) = \exp(-3t)$ over t in [0,10], x in [0,1], y in [0,1].N happens to have points at (1.5, .4, .2), (2, .52, .31), (4, .1, .33), (5, .71, .29). What is the log-likelihood of this realization?

-4.5-6-12-15 - $\iiint$ exp(-3t) dt dx dy $l(\theta) = \sum \log(\lambda(\tau_i)) - \int \lambda(t,x,y)\, dt\, dx\, dy$

= -37.5 - $\int_0^{10}$ exp(-3t) dt, because x and y go from 0 to 1, no relation to each specific point

= -37.5 - exp(-3t) / (-3) $|_0^{10}$ - ⅓ exp(-3t) $|_0^{10}$ because of integrals
= -37.5 + exp(-30)/3 - exp(0)/3
= -37.5 + exp(-30)/3 – 1/3
~ -37.83.

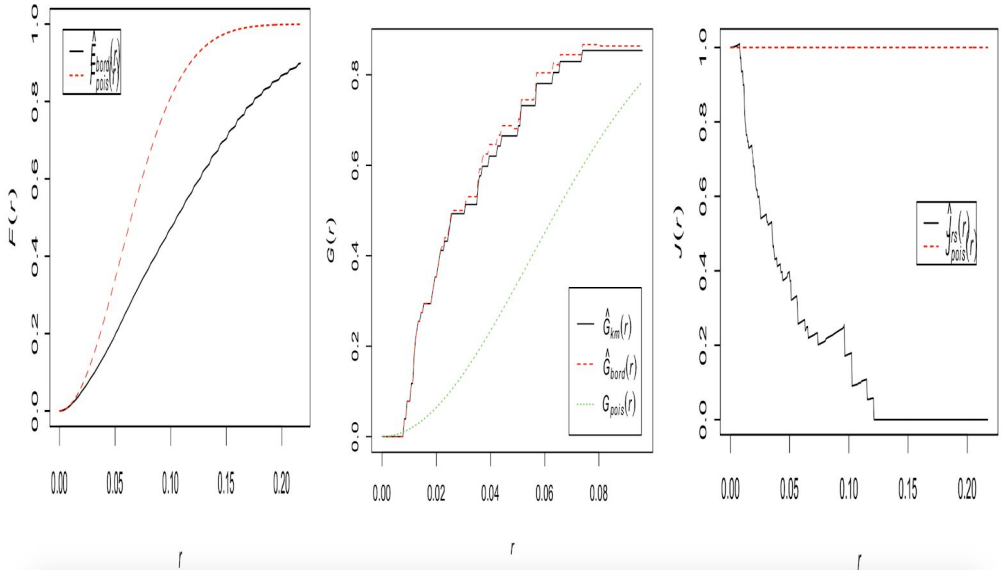
2. Which of the following is not typically true of the MLE of a spatial-temporal point process?
a. It is unbiased. asymptotically unbiased
b. It is consistent. consistent includes asymptotically
c. It is asymptotically normal. MLE = $\widehat{\theta}$
d. It is asymptotically efficient.

18. Kernel smoothing. summarize the main feature of data, density

A simple way to start summarizing a spatial point process is by kernel smoothing.
Suppose observation region is B. Let k(x,y) be a spatial density function, called a kernel, and construct, for each location (x,y), $\widehat{\lambda}(x,y) = \int_B k((x',y') - (x,y))\, dN(x',y') / \rho(x,y)$,
where $\rho(x,y) = \int_B k((x',y') - (x,y))\, dx'\, dy'$ is an edge correction term.

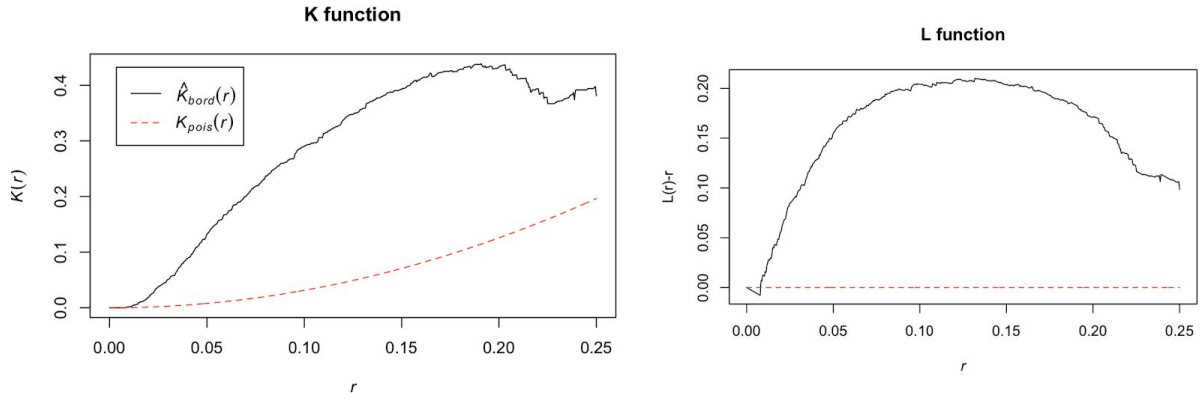
19. F, G, J, K, and L functions.

Let F(r) be the probability that the distance from a randomly chosen location to its nearest *point* of the process is ≤ r.
Let G(r) be the probability that the distance from a randomly chosen *point* to its nearest neighbor is ≤ r.
F is the empty space function and G is the nearest neighbor distribution function.
For a homogeneous Poisson process, $F(r) = G(r) = 1 - \exp(-\lambda \cdot \pi\, r^2)$.
Let J(r) = (1-G(r)) / (1-F(r)), for any r such that F(r) <1.
J > 1 indicates inhibition, and J < 1 indicates clustering.



Lower F, Higher G or Lower J indicate clustering, compared with stationary poisson process
red line is stationary poisson, 1 for J-function

For a stationary Poisson process with rate μ, let K(r) = 1/μ · E(# of other points within distance *r* of a randomly chosen point).
For a stationary Poisson process in R^2, K(r) = πr^2, so one may consider L(r) = $\sqrt{(K(r)/\pi}$ = r
For a stationary Poisson process in R^2, L(r) – r = 0 and $\widehat{L}(r)$ -r should be approx. 0.



Higher K or Higher L indicate clustering, compared with stationary poisson process
red line is stationary poisson, πr^2 for K-function, 0 for L-function
choose L over K, L is more powerful than K, making difference between two lines more obvious

Exercise 6
1. Suppose N is a spatial point process with clustering for distances ≤ d.
Let F(r) be the empty space function and let G(r) be the nearest neighbor distribution function.
Which of the following is true? mean G larger, F smaller compared to stationary poisson
a. F(d) = G(d).
b. F(d) < G(d).
c. F(d) > G(d)

F is the empty space function and G is the nearest neighbor distribution function.
F and G are prob. F(r) is abouts some random space, G(r) is about space that origin at some point

20. Marked G and J functions. e.g. bears in forest, mark be bear species (about association, magnitude), different kinds of bears affect each other (cluster or inhibit)
G(r) = P_0 (point within r), where P_0 means given a point 0, estimated with
$\widehat{G}(r) = \frac{1}{n}\Sigma_i 1(there\ is\ j : |\tau_i - \tau_j| \le r) = \frac{1}{n}\Sigma_i 1(min_{i \neq j} |\tau_i - \tau_j| \le r)$
Could alternatively compute a *marked* G-function $\frac{1}{n_1}\Sigma_i 1(min_j |\tau_i - \tau_j| \le r)$
where the sum is over the n_1 points τ_i with mark in some range M_1 (bear 1 area), and the minimum is over the points τ_j with mark in some range M_2 (bear 2 are). This is the *marked* or *cross* G-function.
Similarly define a marked or cross J-function as J(r) = (1-G(r)) / (1-F(r)) accordingly, plugging in the corresponding G function. ↓ ↑

21. Weighted K function.
For a stationary Poisson process with rate μ, K(r) = 1/μ E(# of other points within distance *r* of a

randomly chosen point).
Estimated via $K_4(r) = 1/(\widehat{\lambda}\, n) \sum_{i \neq j} (|\tau_i - \tau_j| \le r)\, w(\tau_i, \tau_j)$, where $\widehat{\lambda}$ = n/|S|, and $w(\tau_i, \tau_j)$ = 1/proportion of circle centered at i going through j that is in S = border correction term.
If N is inhomogeneous, can instead weight each point by 1/λ, obtaining
$K_w(r) = 1/n \sum_{i \neq j} (|\tau_i - \tau_j| \le r)\, w(\tau_i, \tau_j) / \lambda(\tau_i) / \lambda(\tau_j)$.
$K_w(r) \sim N(\pi r^2, 2\pi r^2 |S| / E(n)^2)$, if inf $\lambda = 1$.
orderly K-function: H_0 : *data points uniformly scattered out*
vs. H_0 : *how model fitted the cluster/inhibition* (*test cluster/inhibition out by the model*)
weighted K-function should plot as curve πr^2 if model correct

22. $\gamma(t) = \rho(0) - \rho(t)$**, for 2nd order stationary processes.** Why?
2nd order stationary = weakly stationary and means
$E(X_t^2) < \infty$, $E(X_0) = E(X_1) = \ldots = E(X_t)$ *for all t*,
and $Cov(X_0, X_t) = Cov(X_1, X_{t+1}) = Cov(X_2, X_{t+2})$, etc., for any t.
Cov for two points at every t distance is the same
If 2nd order stat., then letting t = 0, $Var(X_0) = Var(X_1) = \ldots = Var(X_t)$ for all t.
The semivariogram $\gamma(t) = Var(X_t - X_0)/2$.
The covariogram $\rho(t) = Cov(X_0, X_t)$.
So $\rho(0) - \rho(t) = Cov(X_0, X_0) - Cov(X_0, X_t) = Var(X_0) - Cov(X_0, X_t)$.
$\gamma(t) = Var(X_t - X_0)/2 = Cov(X_t - X_0, X_t - X_0)/2$
$= \{Cov(X_t, X_t) + Cov(X_0, X_0) - 2Cov(X_0, X_t)\}/2$
$= Var(X_t)/2 + Var(X_0)/2 - Cov(X_0, X_t) = Var(X_0) - Cov(X_0, X_t)$
$= \rho(0) - \rho(t)$

23. Nonparametric estimation of Hawkes and ETAS processes.
Let **x** mean spatial coordinates = (x,y).
Hawkes processes have $\lambda(t,x) = \mu(x) + K \sum_i g(t - t_i, x - x_i)$.

- 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

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

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

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.

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.

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

Exercise 7

1. 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. $S \times T$

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

a) 5 μ + 10 exp(μ).

b) 5 log(μ) - 10 μ.

c) 5 + 10 log(μ).

d) 5 exp(μ) + 5 log(μ).

$l(\theta) = \sum \log(\lambda(\tau_i)) - \int \lambda(t,x,y)\, dt\, dx\, dy = 5\, log(\mu) - \mu\, |_0^1\, |_0^1\, |_0^{10} = 5\, log(\mu) - 10\, \mu$

2. Suppose observe a Poisson process with rate 3t on space-time window [0,1]x[0,1] x [0,10]. How many points do you expect to observe? $S \times T$

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.**

E(N(B)) = A, Var(N(B)) = A for poisson process

so $A = \int_B \lambda(t,x,y)\, dt\, dx\, dy = \int_0^1 \int_0^1 \int_0^{10} 3t\, dt\, dx\, dy = 3t^2/2\, |_0^1\, |_0^1\, |_0^{10} = 3(10^2)/2 - 0 = 150$

3. 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], where f(t) is a density like f(t) = 4exp(-4t), and g(x,y) is a planar density like g(x,y) = 3 exp(-3r) / (2π r), where $r = \sqrt{x^2 + y^2}$. μ=2, k=0.6

How many points do you expect to observe?

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

$20 + 20 \times .6 + 20 \times .6^2 + 20 \times .6^3 + \ldots = 20/(1 - .6) = 20/.4 = 50$ $\mu\, T = 2 \times 10 = 20$

Alternatively, <u>60% of the pts are expected to be triggered (k=.6 interpretation)</u>, so 40% are background, and we expect 20 background points, so 20 = 40% of x, so x = 20/.4 = 50.

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

$\int \lambda(t,x,y)\, dt\, dx\, dy \;=\; \mu T \;+\; k \int \sum g(t-t', x-x', y-y')\, dt\, dx\, dy$

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

4. 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], 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) $2\log(2) - 20$.

$\sum \log(\lambda) - \int \lambda\, d\mu = \log(2) + \log\{2 + .6(4\exp(-8))(3\exp(-.3)/(.2\pi))\} - 20 - .6 - .6$ k=0.6, N=2
$= \log(2) + \log(2 + 36\exp(-8.3)/\pi) - 21.2$. *log(1st point) + log(2nd point) − (μT + k N)*

1st point with $t' = 2$, 2nd point with $t = 3$. $r = \sqrt{x^2 + y^2} = \sqrt{(0.5-0.5)^2 + (0.6-0.5)^2} = 0.1$

24. Deviance residuals.

Given two competing models, can consider the difference between residuals, number of observed points – number expected, over each pixel. Divide by the estimated SE to get *Pearson residuals.* Problem: hard to interpret.

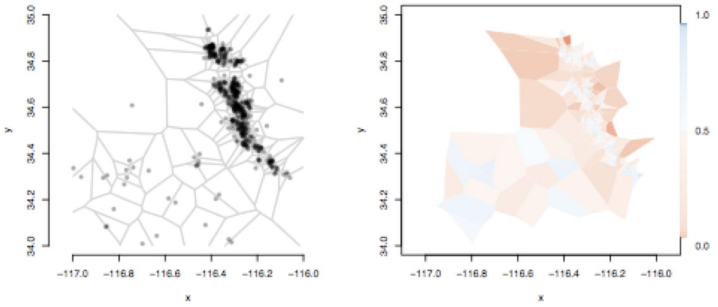
With two competing models, it is better to consider the difference between *log- likelihoods*, in each pixel. The result may be called *deviance residuals*, ~ resids from gen. linear models.

$$R_D(B_i) = \sum_{i:(t_i,x_i,y_i)\in B_i} \log(\hat{\lambda}_1(t_i,x_i,y_i)) - \int_{B_i} \hat{\lambda}_1(t,x,y)\, dt dx dy$$
$$- \left(\sum_{i:(t_i,x_i,y_i)\in B_i} \log(\hat{\lambda}_2(t_i,x_i,y_i)) - \int_{B_i} \hat{\lambda}_2(t,x,y)\, dt dx dy \right).$$

25. Voronoi residuals. measure how model for λ fits

A Voronoi tessellation divides a space into cells Ci, where Ci contains all locations closer to event i than any other observed event. Within each cell, calculate residuals.

$$\hat{r}_i := 1 - \int_{C_i} \hat{\lambda}\, d\mu$$
$$= 1 - |C_i|\bar{\lambda},$$



spatially adaptive and nonparametric.

cells of different sizes; more zero (red) residuals, more fitted the model for λ is

With 2 models, can compare loglikelihoods across pixels or Voronoi cells.

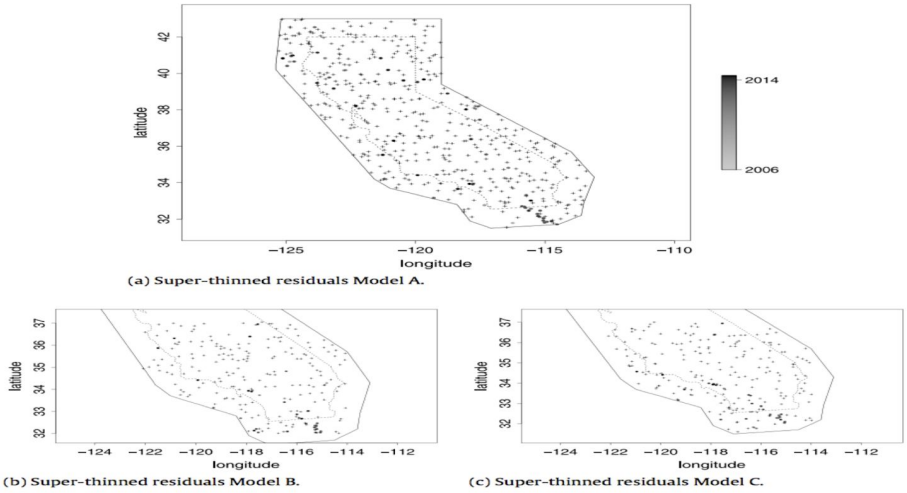
26. Superthinning. test how well model for λ fits

Choose some number c ~ mean($\widehat{\lambda}$). want it to be stationary poisson after superthning

Superpose: where $\widehat{\lambda}(t, x, y) < c$, add in points of a simulated Poisson process of rate c - $\widehat{\lambda}(t, x, y)$

Thin: where $\widehat{\lambda}(t_i, x_i, y_i) > c$, keep each point $\widehat{\lambda}$(ti,xi,yi) with prob. c / $\widehat{\lambda}(t_i, x_i, y_i)$.

Result is Poisson with rate c, if the model for λ is correct.



gap: model for λ is too high;

if λ too small, end up with clustering

Exercise 8

1. Suppose N_1 is a Poisson process with rate 3,
and N_2 is a Poisson process with rate 2 + x + 4t, independent of N_1, and both are on [0,10] x [0,1] x [0,1]. T S

Let M = $N_1 + N_2$. Is M a Poisson process? What is its intensity?

For any disjoint measurable sets $B_1, B_2, \ldots$, M(B_i) = $N_1(B_i) + N_2(B_i)$ is independent of $\{N_1(B_j)$, j≠i} and $\{N_2(B_j)$, j≠i}, and thus is independent of $\{N_1(B_j) + N_2(B_j)$, j≠i}.

So yes, M is a Poisson process and since EM(B) = $EN_1(B) + EN_2(B)$, M has rate 5 + x + 4t.

2. Suppose N is homogeneous Poisson process with rate 1, and M is a clustered Hawkes process. Both M and N have 40 points on B = [0,10] x [0,1] x [0,1]

Let v1 = the average size of a Voronoi cell in a Voronoi tessellation of N, and v2 = the average size of a Voronoi cell in a Voronoi tessellation of M. partition of B into cells

Which is bigger, v1 or v2, or will they be the same?

The same, since v1 = v2 = 1/4. Each cell has one point, and the 40 cells occupy an area of size 10. *area B* = 10 × 1 × 1 = 10, *and* 40 *points means* 40 *cells in area B*; 10/40 = 1/4

bigger or smaller size, cells all over the space B

Review list.
1. PP as a random measure.
2. Integration, $\int f(t,x,y)\, dN$.
3. Simple and orderly.
4. Cond. intensity and Papangelou intensity.
5. Poisson processes.
6. Mixed Poisson processes.
7. Compound Poisson processes.
8. Poisson cluster processes.
9. Cox processes.
10. Gibbs and Strauss processes.
11. Matern processes.
12. Hawkes and ETAS processes.
13. Likelihood and MLE.
14. Covariance and variogram.
15. Kriging.
16. CAR, SAR models.
17. Simulation by thinning.
18. Martingale formula.
19. Kernel smoothing.
20. F,G,J,K, and L functions.
21. Marked G and J functions.
22. Weighted K function.
23. Nonparametric triggering function est.
24. Deviance, Voronoi, andsuperthinned residuals.

Exercise

1. For an Ornstein-Uhlenbeck process, a 2nd-order stationary process with covariance function $\rho(h) = \exp(-\beta |h|)/(2\beta)$, for all h, where $\beta > 0$.
What is the corresponding semivariogram? What are the nugget, sill, and partial sill?

Use the fact that $\gamma(h) = \rho(0) - \rho(h)$. See p15 of van Lieshout.
$\gamma(h) = 1/(2\beta) - \exp(-\beta |h|)/(2\beta)$.
The nugget effect is $\lim_{h \to 0} \gamma(h) - \gamma(0) = 0 - 0 = 0$.
The sill is $\lim_{h \to \infty} \gamma(h) = 1/(2\beta)$.
The partial sill is $\lim_{h \to \infty} \gamma(h) - \lim_{h \to 0} \gamma(h) = 1/(2\beta)$.

semi-variogram $\gamma = \frac{1}{2} var(X_t - X_0)$
Let $X = (Xt)_{t \in Rd}$ *be intrinsically stationary.*
Then the semi-variogram $\gamma : R^d \to R$ *is defined by* $\gamma = \frac{1}{2} var(X_t - X_0)$.
Note that $\gamma(t) = \rho(0) - \rho(t)$ for weakly stationary random fields.
In particular, $\gamma(0) = \rho(0) - \rho(0) = 0$.
The definition of a semi-variogram, however, requires only the weaker assumption of intrinsic stationarity.

In practice, there is often additional measurement error. To be specific, suppose that the observations are realisations from the linear model $Y_i = X_{t_i} + E_i,\ i = 1, ..., n,$
for independent, identically distributed zero mean error terms E_i that are independent of the intrinsically stationary random field X and have variance σ^2_E.
Then $\frac{1}{2}Var(Y_j - Y_i) = \gamma_X(t_j - t_i) + \frac{1}{2}Var(Ej - Ei) = \gamma_X(t_j - t_i) + \sigma^2_E 1\{i \neq j\}$
so that $\gamma_Y(t) = \gamma_X(t) + \sigma^2_E \ for\ t \neq 0;\ \gamma_Y(t) = \gamma_X(t)\ for\ t = 0$
is discontinuous in t = 0. This phenomenon is known as the *nugget effect*.
It is natural to assume that the dependence between sampled values fades out as the distance between them increases, that is, $lim_{||t||\to\infty}\ \rho(t) = 0$, provided it exists. In this case, the limit $lim_{||t||\to\infty}\ \gamma(t)$ is called the *sill*. Taking into account the nugget effect, the *partial sill* is defined as $lim_{||t||\to\infty}\ \gamma(t) - lim_{||t||\to 0}\ \gamma(t)$.

2. Which of the following is not true regarding the differences between a CAR model, a SAR model, and kriging?
a. For a CAR model, the errors are correlated with each other, whereas with SAR the errors are uncorrelated.
b. For a CAR model, the errors at one location are uncorrelated with the values of the random field at other locations, whereas with SAR the errors and the random field are correlated with each other.
c. With CAR and SAR, only neighboring values are used to predict a certain value of the random field, whereas kriging uses all the values and as a result under general conditions is optimal for prediction.
d. With CAR and SAR, typically the covariance function is zero unless two values are neighbors, whereas this is not typically assumed in kriging.
e. None of the above.

CAR: value on location, some covariance of near locations matters
CAR and SAR are different at the error term; both based on variogram, use variance, how one value depends on neighbors
Kriging uses all values, but CAR and SAR only use neighbor values
X = BX + E is an autoregression formula. Note, though, that the 'noise' field E may be spatially correlated. E is correlated in CAR models
Uncorrelated noise in SAR models

3. Suppose a point process is generated on B = [0,10] days x [0,1] x [0,1]. First one generates parent points according to a stationary Poisson process with rate 0.3. Then each parent point gives birth to exactly one child point placed uniformly within 1 day after the parent and anywhere in the unit square. The resulting process consists of both the parents and children points.
One realization of this process results in four points, at (3, 0.4, 0.5), (3.4, 0.7, 0.8), (7.2, 0.4, 0.5), and (7.5, 0.9,0.1). What is the loglikelihood, L?

L = $\sum$ log(λ) - $\int \lambda$(t,x,y) dt dx dy
= log(.3) + log(1.3) + log(.3) + log(1.3) – 0.3 x 10 – 2

~ -6.883.

For a Hawkes process where $\lambda(t,x,y) = \mu(x,y) + k \Sigma_{\{t',x',y':\, t' < t\}}\, g(t-t', x-x', y-y')$, where g is a density, and $\int \mu(x,y)\, dxdy = \mu$, how far away from space and time

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

This point process is like Hawkes process

(3, 0.4, 0.5), (3.4, 0.7, 0.8), (7.2, 0.4, 0.5), and (7.5, 0.9,0.1) are one parent point followed with one offspring point, and one parent point followed with one offspring point

1st point cannot be an offspring point, so log(0.3)

2nd point be log(0.3+1): space-time: unit square and 1 day

so $\sum \log(\lambda)$ be log(.3) + log(1.3) + log(.3) + log(1.3)

$\int \lambda(t,x,y)\, dt\, dx\, dy = 0.3 \times 10 - 1 - 1$

$= \mu T$ - 2 offspring points triggered by 2 parent points

Simulation by Thinning

One can simulate spatial-temporal point processes by *thinning*.

Simulation of nonhomogeneous poisson processes by thinning.

Suppose λ has some upper bound, B. $\lambda(t,x,y) \leq B$ everywhere.

First, simulate a stationary Poisson process *N* with intensity B.

For i = 1,2, ... keep point τ_i with probability $\lambda(\tau_i)/B$.

Boundary issues can be important in simulation. For Gibbs processes, for instance, the simulation can be biased because of missing points outside the observation region. For Hawkes processes, the simulation will tend to be biased by having too few points at the beginning of the simulation. One can have burn-in, by simulating points outside the observation region or before time 0, or in some cases some fancy weighting schemes can be done to achieve *perfect* simulation without burn-in.