Surveillance of disease Maps

BMTRY 763

Background

- Surveillance is about detecting **change**
- This is not what most disease mapping is about
- Most disease mapping studies are *retrospective*
  - *The data have been collected and hence the task is to model or investigate the process history*
  - Space-time disease mapping examples and associated models you have seen are all retrospective : Ohio respiratory cancer, Knorr-Held model; SC Flu models; FMD in Cumbria
- Surveillance should be *prospective*

Reference

- Lawson, A. B. and Kleinman. K. (2005) (eds) *Spatial and Syndromic Surveillance for public Health*. Wiley
- Brookmeyer, R. and Stroup, D. (2004) (eds) *Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance*, Oxford University Press, London

- recent reviews:

Robertson, C., Nelson, T.A., MacNab, Y.C., Lawson, A.B. (2010) Review of methods for space-time disease surveillance. *Spatial and Spatio-temporal Epidemiology,* 1: 105-116.

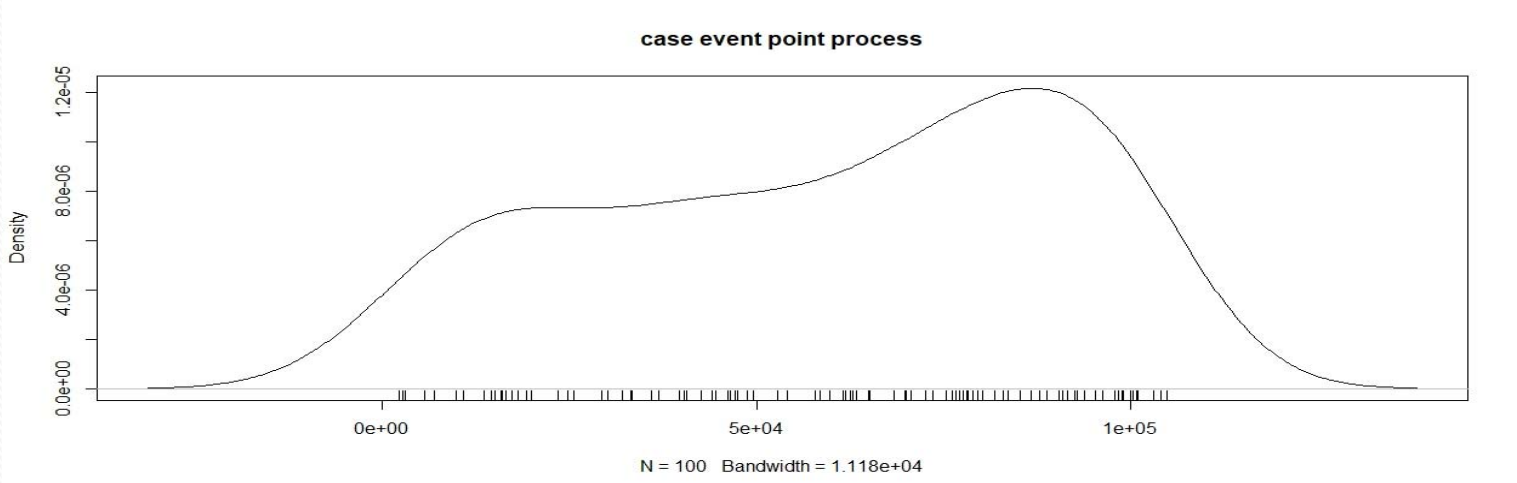
S. Unkel and C. P. Farrington and P. Garthwaite and C. Robertson and N. Andrews (2012) Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society,* 175, 49-82

Some Basic surveillance issues

- Industrial process control was the basis for early work.
- This usually sets up tests for changes in process
  - Examples would be thickness of cloth produced y a machine
    - Simple assumption of Gaussian behavior and testing for unusual Z scores (outliers)
    - Many related measures have been developed
  - Not appropriate for health surveillance for 2 reasons:
    - 1) disease occurs within populations and so there is no constant level to be maintained (and no intervention directly)
    - 2) disease events are discrete (usually counts or case events)

Example

- Case events: disease events occur as a point process.
- In time they form a sequence (Poisson process with exponential gaps shown)



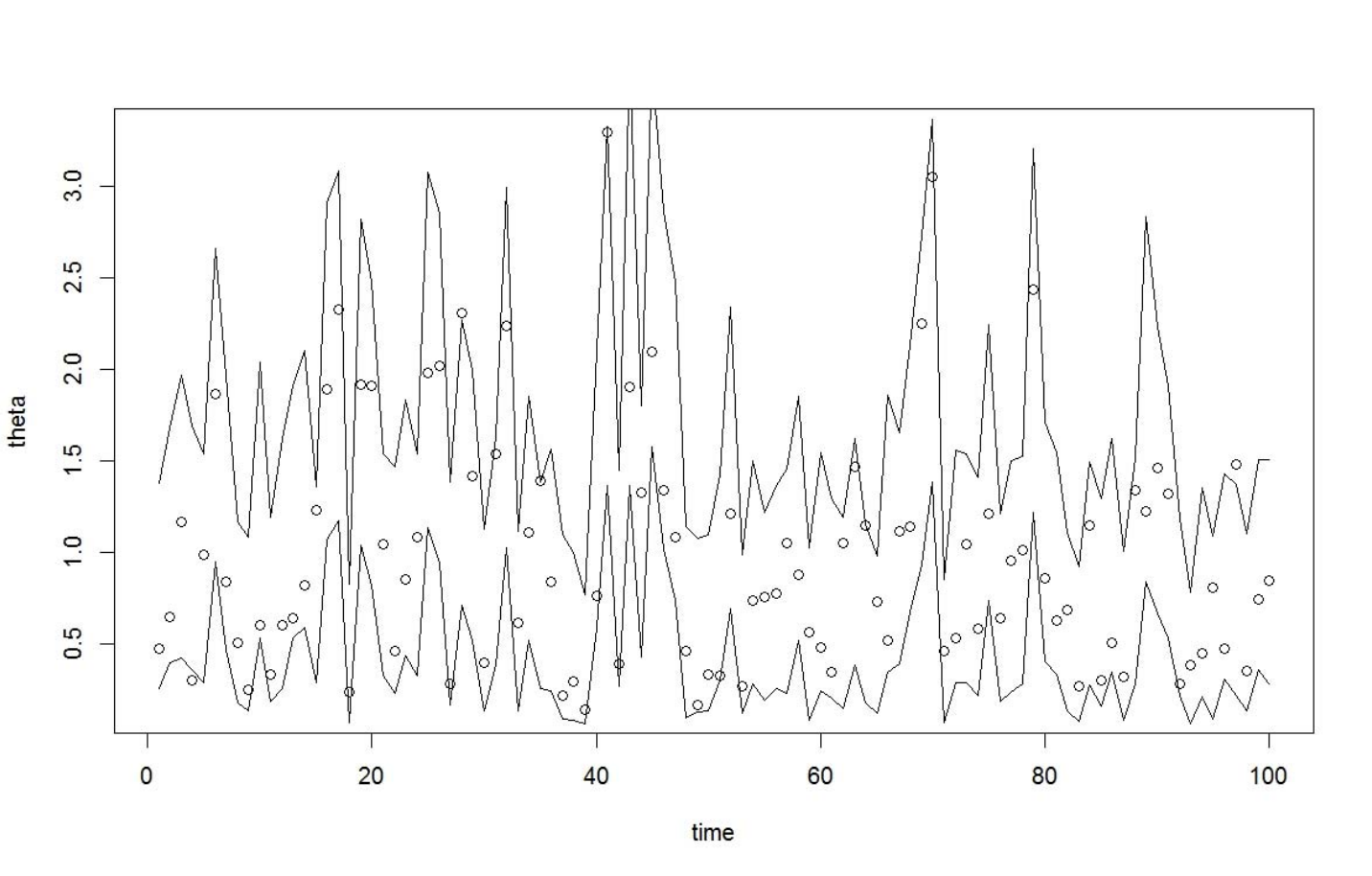
Counts of events

- Simpler to deal with counts within J defined time units
- Define the following model for disease counts:

$$y_j \sim Pois(e_j \lambda_j); \;\; j = 1,\ldots,J$$

- Figure below represents a simulated sequence of counts where the expected rates are generated from N(5,1) and the relative risk is thought to follow an Gamma prior distribution with Ga(3,3). The 95% quantiles are from GA(y+3,e+3) as solid lines

Time series of simulated relative risks



Temporal modelling and measures

- Various measures have been used in surveillance of disease in time:
  - Gap tests for case events
  - Scan tests
  - Cusum measures
- For modeling, Bayesian methods (amongst others) have been used as they allow easy temporal updating (recursive Bayesian learning)
  - Posterior functionals can be used:
    - Surveillance CPO (SCPO)
    - Surveillance residuals

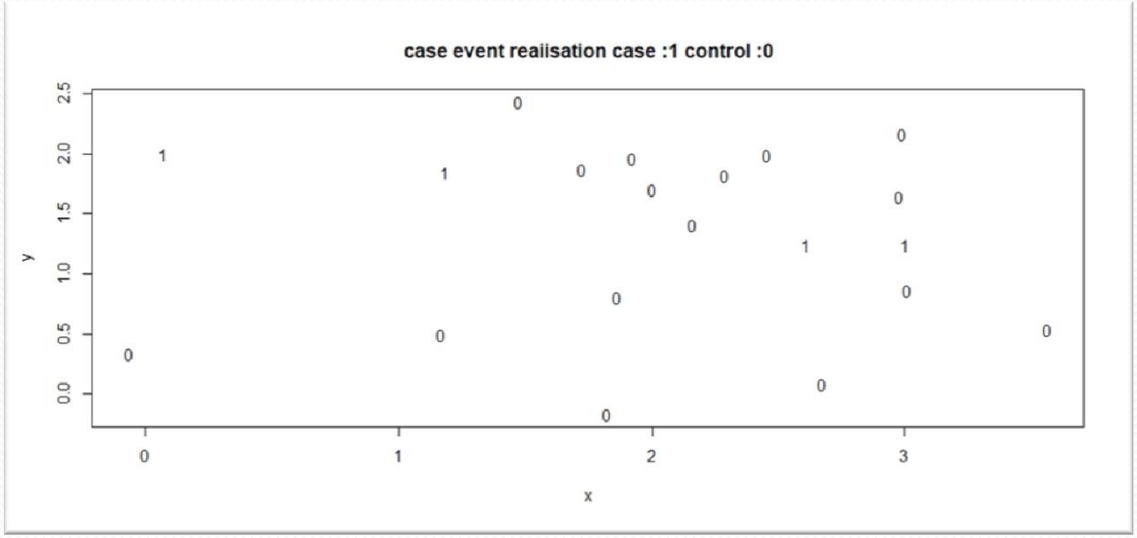
Extension to spatial domain

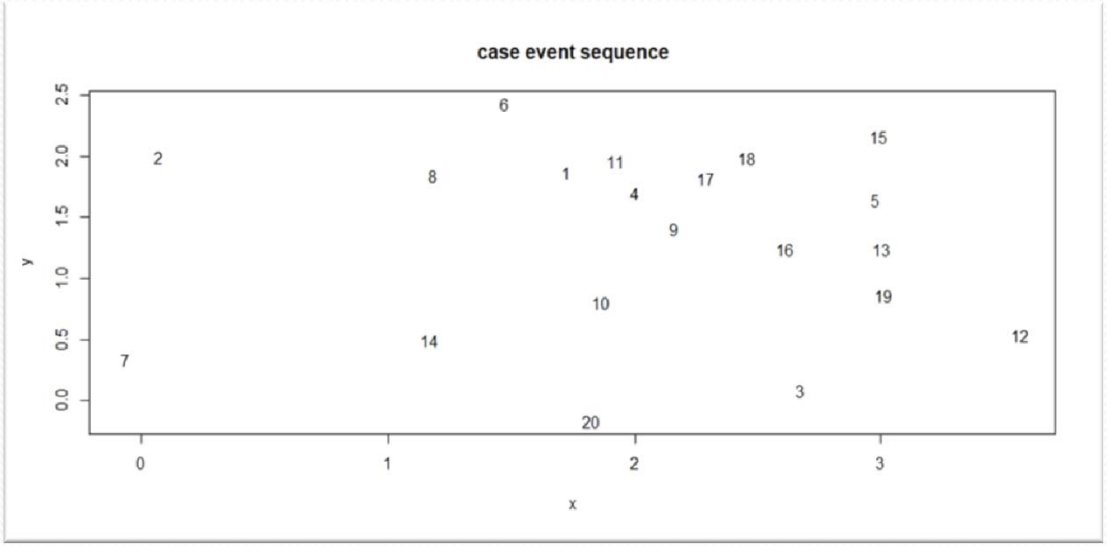
- Assume that we have a study window and observe it over a fixed period
- Events can occur in sequence within and without the window
  - This is a problem as this means that we only observe a partial sequence of events ie some are censored as they didn't happen within the window: disease ignores boundaries
- Assume that we only want to model events within the window and only record events wihtin.

Sequence is important

- As the disease progresses new cases could arise anywhere eg a cluster could develop and new cases arise close together, or they arise 'randomly' around the study area
- Need to have a control disease as well as case disease
  - e. g. birth example: normal birth(control), abnormality birth (case)

Case event sequence





Case event sequence

- Cases are 2, 8, 13, 16
- The ordering is important:
  - Cases can arise within existing incidence or can push the boundary of the convex hull
  - case 2 pushed the hull
  - Case 8 didn't
  - Case 13 almost did but 16 didn't
  - Hence case 2 spread the incidence but 8,13,16 increased the local density of cases

Surveillance of case events

- Complex unless you convert the problem:
  - Assume the cases and controls are treated as labels and we model the label sequence (conditional on the whole realisation having been observed).
  - Define

$$y_j = \begin{cases} 1 & \text{if case} \\ 0 & \text{if control} \end{cases}$$

and

$y_j \sim Bern(p(s_j))$ where s_j is the location of the j th event

  - Can model $\text{logit}\, p(s_j)$

Count surveillance

- Often counts within space-time units are only available and so the exact sequence of case incidence is lost due to aggregation.
- Define

$$y_{ij}, e_{ij}, \lambda_{ij}, \; i = 1,\ldots,m; \; j = 1,\ldots,J$$

and

$$y_{ij} \sim Pois(e_{ij}\lambda_{ij})$$

- Often a Poisson data model is assumed
- If a finite population is assumed then a binomial may suffice

Issues

- What are the prospective issues we must address:
  - What is best model?
  - How will fitting be done?
  - How to monitor change?
  - How to assess performance?

Best model

- There are state of the art models that fit ST data
  - Knorr-Held model is very good for retrospective data

$$\log \lambda_{ij} = \alpha + v_i + u_i + \gamma_j + \psi_{ij}$$

  - This might fit TOO well….as we need to detect changes and do not want to model them out
  - Might want to drop the temporal random effect and leave the interaction term

$$\log \lambda_{ij} = \alpha + v_i + u_i + \psi_{ij}$$

Best model

- For infectious disease there is also the possibility of considering a two part model:
  - With Endemic and epidemic phases
- The endemic phase is simply the normal background
- The epidemic phase is an outbreak
- Usually we wan to predict the start of the epidemic phase
- Example

$$\log \lambda_{ij} = \alpha + EN_{ij} + \rho EP_{ij}$$

$$EN_{ij} = v_i + u_i + \psi_{ij}$$

$$EP_{ij} = \log(y_{i,j-1}).....$$

How to Fit?

- As time goes on then data size increases
- parameter size could increase also

- Options:
  - Refit model to new enlarged data every time a new time point is reached
    - Could be inefficient and computationally burdensome
  - Use a sliding lag window which fixes the sample size and effective parameter set
  - Use dynamic updating such as Sequential MC

How to monitor change

- Could look at ratios of parameter estimates after refitting
- Could design in specific parameters which detect change
- Could use posterior functionals designed to detect change

Eg τ_j / τ_{j-1}; $\lambda_{ij} / \lambda_{i,j-1}$; ρ_j / ρ_{j-1}

- SCPO, MSCPO, EWMA, SKL, MSKL

How to assess performance

- How well do any metrics perform in detecting change?
  - ARL_0 is average run length under normal circumstances and should be as long as possible
  - Pr(false alarm) = $1/ARL_0$
- These are used to assess how well metrics perform
- Other metrics available