Statistical Models for Case events

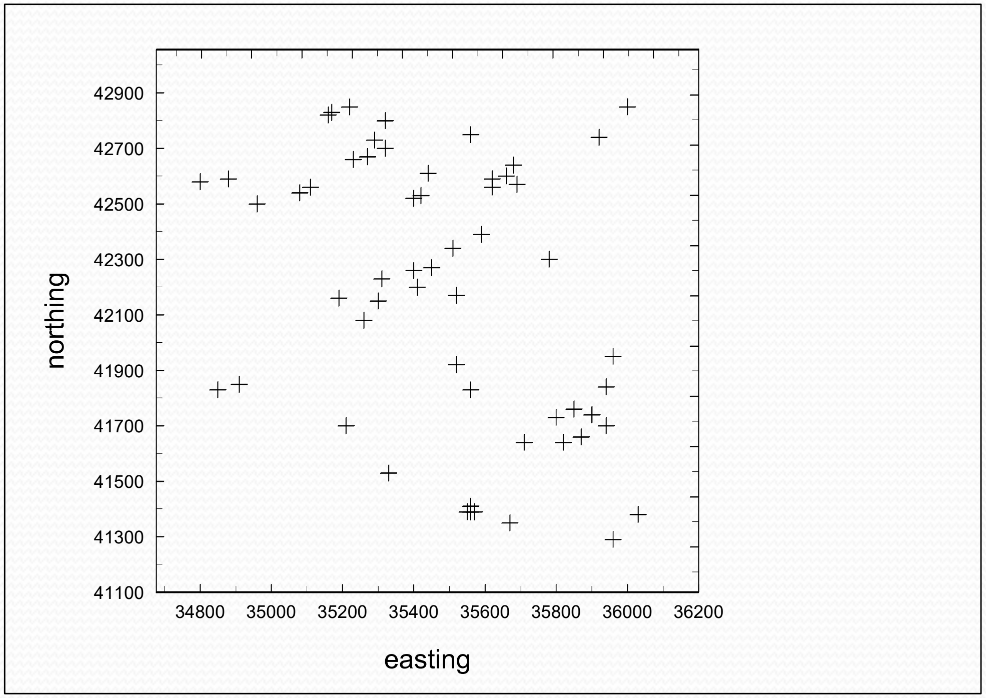
- We assume that within an area *A* we have observed a set of *m* case events at locations

$$\{s_i\}_{i=1,\ldots,m}$$

- We also assume that within the area A we observe a set of *n* controls

$$\{s_j\}_{j=m+1,\ldots,m+n}$$

Larynx cancer example



Poisson process models

- The first order intensity is defined as

$$\lambda(s) = \lambda_0(s).\lambda_1(s, \mathbf{x})$$

- This is the basic measure governing heterogeneous Poisson Process (HPP)
- Case events and controls can be thought to follow a HPP
- The conditional likelihood for a set of m case events is

$$L = \prod_{i=1}^{m} \left\{ \frac{\lambda(s_i)}{\int_A \lambda(\mathrm{u})\mathrm{du}} \right\}$$

Alternative conditioning

- The case events form a set $\{s_i\}_{i=1,\ldots,m}$
- Also the *n* controls form a set $\{s_j\}_{j=m+1,\ldots,m+n}$

- The case events have intensity $\lambda(s) = \lambda_0(s).\lambda_1(s,\mathbf{x})$
- The controls have intensity $\lambda_0(s)$
- If we merge these sets i.e. superimpose them, then the joint set has intensity $\lambda_0(s)[1+\lambda_1(s,\mathbf{x})]$
- We now ask the question: what is the probability of a case at a given location in the joint set ?

Case probability

- what is the probability of a case at a given location in the joint set ?
- This is just:

$$\lambda_0(s).\lambda_1(s,\mathbf{x}) / \lambda_0(s)[1+\lambda_1(s,\mathbf{x})] = \lambda_1(s,\mathbf{x}) / [1+\lambda_1(s,\mathbf{x})]$$

- What is the probability of a control at a given location?
- This is just: $\lambda_0(s) / \lambda_0(s)[1+\lambda_1(s,\mathbf{x})] = 1 / [1+\lambda_1(s,\mathbf{x})]$
- Hence, for short, $\text{Pr(case at s}_i\text{)} = p_i = \lambda_{1i} / [1+\lambda_{1i}]$

  and $\text{logit}(p_i) = \log(\lambda_{1i})$

Conditional Logistic Regression

- This is just a logistic regression where

$$y_i = \begin{cases} 1 \text{ if } s_i \text{ is a case} \\ 0 \text{ if } s_i \text{ is a control} \end{cases}$$

- and if the modelled intensity is

$$\lambda_1(s_i) = \exp(\eta_i)$$

$$\eta_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}$$

Logistic spatial model

- Hence we have a Bernoulli model for the outcome data and with a logistic link to a linear predictor we have a logistic spatial regression model:

$$y_i \sim Bern(p_i)$$

$$\text{logit}(p_i) = (\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + R_i)$$

$$R_i = \text{ random effects}$$

Space-Time (ST) extension

- Assume that case events arise at locations (s) and associated with a date of diagnosis (t) :
- Within *A* we observe *m* events: $\{s_i, t_i\}$
- Probability of case at s and t :

$$\Pr(\text{case at } s,t) = p(s,t) = \lambda_1(s,t) / [1 + \lambda_1(\text{s},\text{t})]$$

- and conditioning on the joint set of cases and controls then

$$y_i \sim Bern(p(s_i, t_i))$$
$$\text{logit}(p(s_i, t_i)) = (\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + R_i)$$
$$R_i = \text{ ST random effects}$$