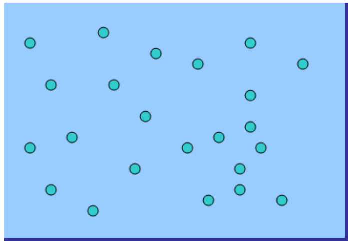
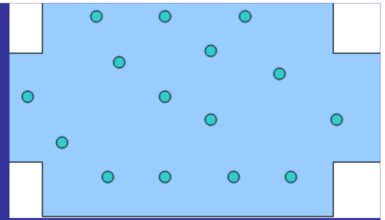
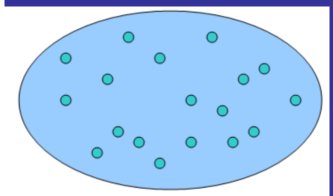
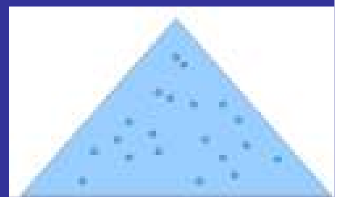




STATISTICS AND ITS APPLICATION IN POINT PATTERN ANALYSIS

Rizwanur Rahman
First year PhD Mathematics Student
University of Bristol



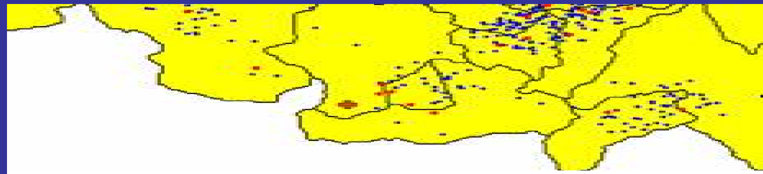


Presentation Outline

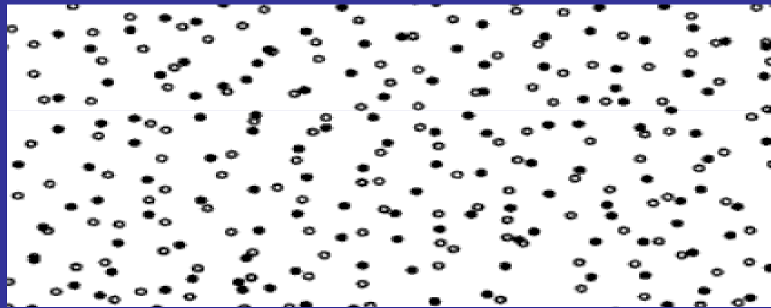
- Statistics and Point Patterns
  - Point Patterns
  - Important Applications of Point Pattern Analyses
  - Some Statistical Methods in Point Pattern Analyses
    - Exploratory
    - Testing for randomness

Statistics and Point Patterns

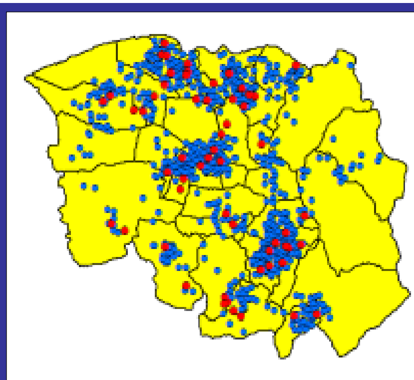
- Point pattern:
  - Event collection over region arising from generating process
  - Interest in whether events clustered or not (many others)
  - Types:
    - Spatial – at point locations, e.g. Larynx cancer cases (red) over district with incinerator
    - Bivariate – two types, e.g. Amacrine cells in rabbit's retina
    - Spatio-temporal – depends on time and space, e.g. Lung (blue) and Larynx (red) cancer cases



Larynx Cancer Cases in Lancashire

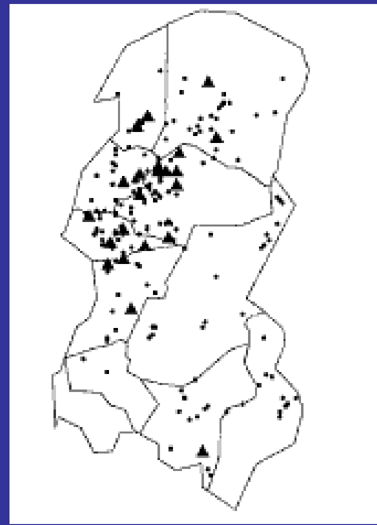


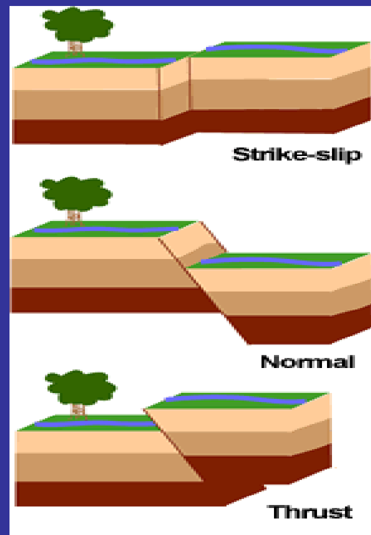
Distribution of Amacrine cells in the retina of a rabbit



Statistics and Point Patterns

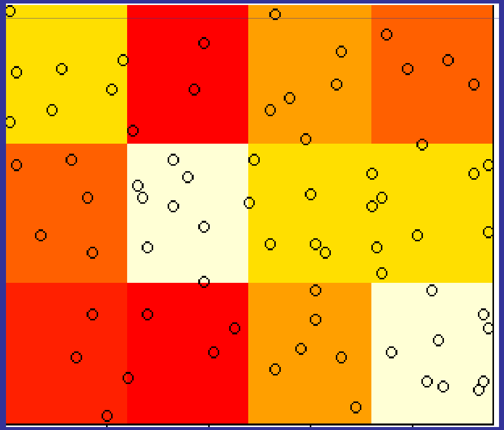
- Applications include:
  - Environmental epidemiology, e.g. are diseases such as Burkitt's Lymphoma infectious – spread with time?
  - Seismology e.g. predict spatio-temporal aftershock occurrence following earthquake based on results obtained for earthquake occurred





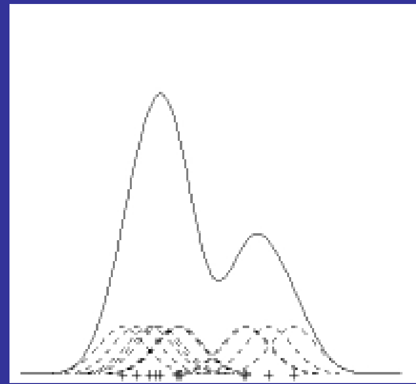
Statistics and Point Patterns

- Quadrat counts:
    - Summarize distribution of point pattern event locations
    - Fine grid of equally spaced squares placed over point pattern
    - # events in each grid square recorded
    - Density plot representing each grid square's event density obtained

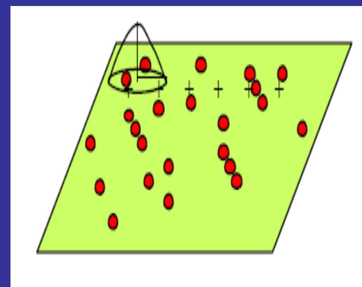


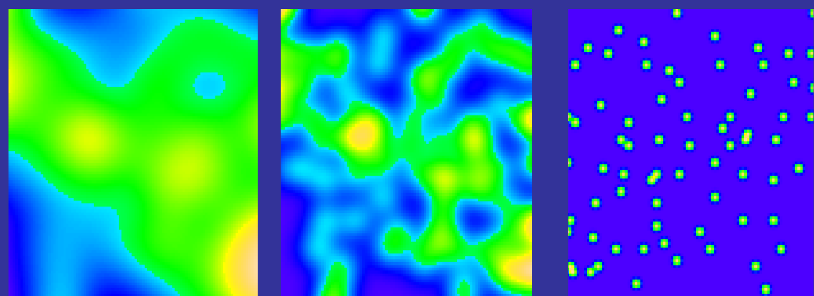
Statistics and Point Patterns

- Kernel estimation:
  - Obtain smooth estimate of uni- or multi-variate probability density from sample (smooth histogram in univariate case)



  - Typical expression for intensity estimate of events in point pattern is: $\hat{\lambda}_\tau(s) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{(s - s_i)}{\tau}\right)$, kernel *k*() is suitably chosen bivariate (2-d) probability density function, bandwidth τ represents smoothing degree, ***s*** is general point location and $\boldsymbol{s}_i$ is *i*th event's point location

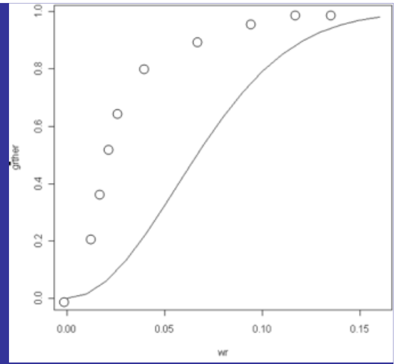


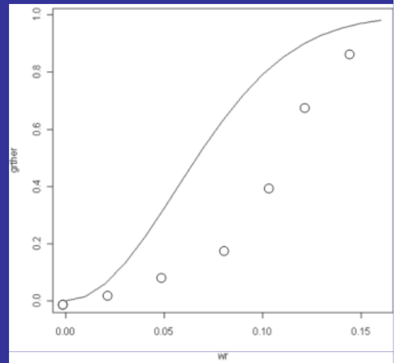


Statistics and Point Patterns

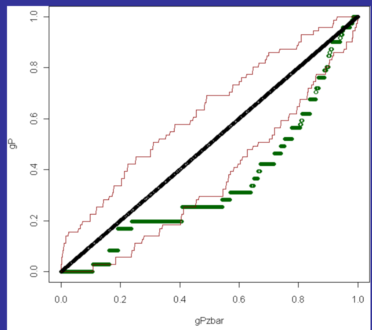
- Index of Dispersion:
  - Test statistic $s^2/\bar{x}$, $\bar{x}$ is mean number of events in each square of grid placed over point pattern and s^2 is its observed variance
  - For randomly distributed events, $s^2/\bar{x}$ should be close to 1
  - For clustered patterns, $s^2/\bar{x}$ should be less than 1
  - For a regular pattern, $s^2/\bar{x}$ should be greater than 1
  - Expect consistent results for point patterns comprising sufficiently many events

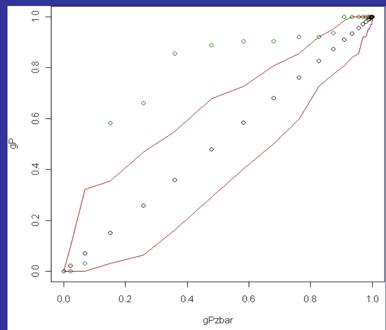
Statistics and Point Patterns





- Nearest Neighbour distances:
  - Randomly distributed events have theoretical cumulative distribution function for their nearest neighbour event distances given by $G(w) = 1 - e^{-\lambda \pi w^2}$, λ is intensity of events occurring in point pattern
  - Clustered patterns have more small than large nearest neighbour distances
  - Regular patterns have more large than small nearest neighbour distances
  - Assess significance of deviation from randomness by:
    - Upper and lower simulation envelopes: $\Upsilon(w) = \max_{i=1,\ldots,m} \{\hat{G}_i(w)\}$ $\Lambda(w) = \min_{i=1,\ldots,m} \{\hat{G}_i(w)\}$, $\hat{G}_i(w)$, for i=1…, m are empirical distribution functions from m independent simulations of n events in random spatial point pattern domain
    - Plot estimate of empirical cumulative distribution for $G(w)$, $\hat{G}(w) = \frac{\#(w_i \leq w)}{n}$, vs $\bar{G}(w) = \sum \hat{G}_i(w)/m$ (green) and superimpose simulation envelopes (red)





Statistics and Point Patterns

- The Clark-Evans test:
  - •The mean nearest neighbour distance for random point pattern containing *m* events arises from

$$N\left(\frac{1}{2\sqrt{\lambda}}, \frac{(4-\pi)}{4\lambda\pi m}\right)$$

  - •Expect clustered patterns to have observed mean nearest neighbour distance values significantly smaller than theoretical mean nearest neighbour event distances
  - •Expect regular patterns to have observed mean nearest neighbour distance values significantly larger than theoretical mean nearest neighbour event distances

Statistics and Point Patterns

- *K* function:
  - $\lambda K(d)$ = E(#(events within distance *d* of arbitrary event)),
  λ = intensity or mean number of events per unit area
  - Suitable *K* function estimate:

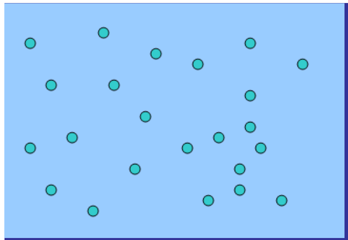
$$\frac{1}{\lambda^2 R}\sum_{i\neq j}\sum I_d(d_{ij})$$

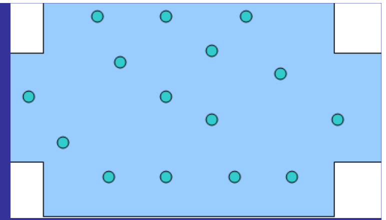
  R is point pattern's area and $I_d(d_{ij})$ is indicator function for *i*-th event within distance *d* of *j*-th event
  - Expected number of events occurring within arbitrary distance *d* of given event from random point pattern is $\lambda\pi d^2$
  - For clustered pattern, $K(d)$ greater than πd^2
  - For regular pattern, $K(d)$ less than πd^2

Summary and Conclusions

- Looked at:
  - Point Patterns:
    - Definition
    - Types
    - Applications
    - Statistical methods
- Further potential work:
  - Modelling point patterns
  - Edge effects and correcting for them
- References:

- ‘*Drivers of bacterial colonization patterns in stream biofilms*’, 2010, Auspurger C.
- ‘*Assessing spatiotemporal predator-prey patterns in heterogeneous habitats*’, 2010, Birkhofer, K.





THE END