Spatial Statistics
Point Process Data Unit 3

PM569 Spatial Statistics

Lecture 10: November 6, 2015

Review of point processes

- Testing for CSR:
  - Adjusting for edge effect
  - Testing for CSR with Ripley's K.
  - Testing for CSR based on inter-event distances, H(h).
  - Testing for CSR based on nearest-neighbour distances, G(h).
- Spatial processes, Poisson processes are the building block:
  - Homogeneous Poisson process (constant intensity).
  - Inhomogeneous Poisson process (intensity varies across domain).
  - Poisson Cluster process (intensity varies for parents and/or children forming clusters), Cox process when the cluster intensity is spatially varying.
  - Simple inhibition processes, Markovian processes (Strauss and pairwise interaction) for regular patterns.
- In this lecture we will focus on fitting point process models and on methods for detecting clusters.

Point Pattern Data

Fitting point process models

- Given our set of observed point events $\{x_1, ...x_n\}$ in region A we wish to fit a model (which is stationary and isotropic)
- Model fitting is approached by estimating the parameters of the particular process
  - Example: fitting the parameters ρ, μ and σ^2 of a clustered process
  - Example: fitting a parametric form of the intensity of an inhomogeneous poisson process
- We use familiar fitting methods: Least squares, Maximum Likelihood and non-parametric methods.

Fitting point process models: Least Squares

- We start with $K(h)$ and the estimator $\hat{K}(h)$ (or L, or G, or H) for parameter fitting
- This is useful when the mathematical form of $K(h)$ is known either explicitly or as an integral (which is true for some point processes)
- If $K(h)$ is not known we use the simulated realizations
- Example, to fit a homogeneous poisson cluster process we have parameters $\theta = (\rho, \sigma)$ and the Ripley's K function is:

$$K(h, \theta) = \pi h^2 + \frac{1}{\rho}(1 - \exp(-h^2/(4\sigma^2)))$$

- And we estimate $\hat{K}(h)$ from the data

Fitting point process models: Least Squares

- Given the theoretical K-function and the estimator $\hat{K}(h)$ we minimize the deviance:

$$D(\theta) = \int_0^{h_0} [(\hat{K}(h))^c - (K(h,\theta))^c]^2 dh$$

- Where h_0 is the maximum distance which is typically chosen as $1/3$ to $1/2$ of the width of a rectangular region, and c is the power transformation
- The power transformation controls the sampling fluctuations in $\hat{K}(h)$ which can increase with h and have influence on $\hat{\theta}$ (i.e. it is a variance stabilizer)
- Examples of c are c=0.5 for a pattern that is not too different from CSR, c=0.25 for cluster patters. However, choose a variety of c values in practice in order to see how sensitive the results are

Fitting point process models: Least Squares Estimation Steps

1. Compute the edge corrected $\hat{K}(h)$

$$\hat{K}(h) = \frac{|A|}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} I(h_{i,j} \leq h)$$

2. Choose a theoretical model for $K(h, \theta)$ where θ are the parameters of the model
3. Find $\hat{\theta}$ that minimizes the deviance for a given c

$$D(\theta) = \int_0^{h_0} [(\hat{K}(h))^c - (K(h, \theta))^c]^2 dh$$

Fitting point process models: Least Squares Estimation

- When $K(h, \theta)$ is unknown because there is no closed form, use the simulated method (for s simulations):

$$\bar{K}_s(h, \theta) = \frac{1}{n} \sum_{i=1}^{s} \hat{K}_i(h, \theta)$$

- Finding $\bar{K}_s(h, \theta)$ for each value of θ can be prohibitive computationally
  1. Start with a small number of simulations, s
  2. Find a first approximation of $\hat{\theta}$
  3. Repeat with a larger value for s

Fitting point process models: Least Squares Estimation Steps

- A weighted version of the deviance, shown to have asymptotic properties (consistency and asymptotic normality), is often used:

$$D(\theta) = \int_0^{h_0} w(h)[(\hat{K}(h))^c - (K(h,\theta))^c]^2 dh$$

- The weight $w(h)$ is a weight on the distance also controls the variance
- When $c = 0.5$ and $w(h) = 1$ we have the Poisson cluster process
- See Guam and Sherman, J R Stat Soc (2007) for asymptotic properties

Fitting point process models: Least Squares Estimation

- In R `spatstat`, cluster or Cox point process models are fit with least squares estimation through the `kppm` function with the `method="mincontrast"` option
- To fit a log-Gaussian Cox point process, use the function `lgcp.estK`
- To fit the Matern cluster process (type I or II), use the function `matclust.estK`

Fitting point process models: Maximum Likelihood

- To fit inhomogeneous Poisson, Strauss and pairwise interaction processes we need to rely on likelihood methods
- Recall for the inhomogeneous Poisson process:
  - $N(A)$ is Poisson with mean $\int_A \lambda(x)dx$
  - Conditional on $N(A) = n$, the n events in A form an independent random sample from A with a probability distribution function proportional to $\lambda(x)$
- We can define the process based on its conditional intensity
- Namely, the conditional probability of finding a point of the process inside an infinitesimal neighbourhood du of the location u given the complete point pattern $\mathbf{x}$ is $\lambda(u, \mathbf{x})du$

Fitting point process models: Conditional Intensity

- For example, CSR has conditional intensity $\lambda(u, \mathbf{x}) = \lambda$
- The IPP has conditional intensity $\lambda(u, \mathbf{x}) = \lambda(u)$
- Sometimes the IPP trend is denoted as $\beta(u)$ and indicates "spatial trend"
- The Strauss process has conditional intensity $\lambda(u, \mathbf{x}) = \beta^n \gamma^p$ where β is the intensity, γ is the interaction parameter, and p is the number of points of $\mathbf{x}$ that lie within a distance δ of u (i.e. pairs of neighbours)
- For example, the Strauss process with $\gamma < 1$ dependence between points is reflected in the fact that the conditional probability of finding a point of the process at the location u is reduced if other points of the process are present within a distance δ. And when $\gamma = 0$, the conditional probability of finding a point at u is zero if there are any other points of the process within a distance δ of this location.

Fitting point process models: Pseudolikelihood

- Because maximum likelihood is difficult for point process models, the log of the pseudolikelihood is maximized, using the conditional intensity
- For a point process governed by parameter θ the pseudolikelihood is:

$$PL(\theta; x) = \prod_{i=1}^{n} \lambda_\theta(x_i; x) \exp(\int_A \lambda_\theta(u; x) du)$$

- The maximum pseudolikelihood estimate of θ minimizes the above equation

Fitting point process models: Pseudolikelihood

- We need to take the log of the pseudolikelihood equation and approximate the integral using a "quadrature" scheme (see Berman and Turner, 1992)

$$\int_A \lambda_\theta(u; x) du \approx \sum_{j=1}^{m} \lambda_\theta(u_j, x) w_j$$

- Where u_j are "quadrature points" in A and $w_j \geq 0$ are the "quadrature weights"

Fitting point process models: Pseudolikelihood

- The quadrature points can be chosen as all data points, x_i and the addition of some dummy points u_j, i.e. $\{x_1, ..., x_n\} \subset \{u_1, ..., u_m\}$
- Then the log pseudolikelihood can be written:

$$\log PL(\theta; x) = \sum_{j=1}^{m} z_j \log \lambda_\theta(u_j; x) - w_j \lambda_\theta(u_j; x)$$

- Where $z_j = 1$ if u_j is a data point, and $z_j = 0$ if u_j is a dummy point

Fitting point process models in R

- In R `spatstat` the function `ppm` fits models by pseudolikelihood based on the conditional intensity $\lambda_\theta(u, x)$
- The model must be loglinear in the parameters θ
- For example, the Strauss process can be written:

$$\log \lambda(u, x) = \log \beta + \log \gamma p$$

- So $\theta = (\log \beta, \log \gamma)$ are the "regular parameters" and the parameter driving the interaction, p is the "irregular parameter"

Fitting point process models in R

- Thus in `spatstat` `ppm` the conditional intensity is split into first and higher order terms:

$$\log \lambda_\theta(u, x) = \eta S(u) + \phi V(u, x)$$

- The first order term $S(u)$ describes the spatial inhomogeneity of the intensity (including covariate effects) and the higher order term $V(u, x)$ describes the interactions between points

Fitting point process models in R

- The general form of `ppm` is `ppm(X, trend, interaction,...)`
- The trend argument specifies any spatial trend or covariate effects and is written as an R formula
- The default trend formula is ~ 1, which indicates $\lambda(u) = 1$, corresponding to a process without spatial trend or covariate effects. The formula $\sim x$ indicates the vector statistic $\lambda(x, y) = (1, x)$ corresponding to a spatial trend of the form $\exp(\alpha + \beta x)$, where α, β are coefficient parameters to be estimated, while $\sim x + y$ indicates $\lambda(x, y) = (1, x, y)$ corresponding to $\exp(\alpha + \beta x + \gamma y)$

Fitting point process models in R

- The general form of `ppm` is `ppm(X, trend, interaction,...)`
- The interaction term represents the interaction function $V(u, x)$
- For example, the Strauss function with interaction distance $\delta = 0.1$ is fit with `ppm(X, 1, Strauss(r=0.1))`
- Note that the `ppm` with a specified higher order term calls the first order term for the intensity β rather than λ
- `spatstat` automatically creates a quadrature scheme but it can be controlled by the user through the function `quadscheme`

Point Pattern Data: Detecting Clusters

- So far we have simulated point processes, found statistics that indicate a global measure of what pattern there may be, and fit models to specific types of point patterns.
- If we want to find where clusters of observations are located, we need a different set of tools called scan statistics.
- Goals of scan statistics:
  - To determine areas where the number of events is inconsistent with the number observed over the rest of the study area.
  - Compare local rates of events (or case/control ratios) to detect clusters.

Point Pattern Data: Detecting Clusters

- First attempts at cluster detection methods were developed in the 1980s: geographical analysis machine and the cluster evaluation permutation procedure.
- Early cluster detection methods:
  - Graphical in nature
  - Divides the region up into a fine grid
  - Uses a search window, which is a circle of predefined radius, larger than the grid spacing in order for circles to overlap
  - Centers the circle over each grid cell, then moves across the region
  - The number of cases occurring within the search window are counted
  - The circle is drawn on the map if the count observed within the circle exceeds some tolerance level
  - The tolerance level may be defined as the observed count exceeding all of the counts associated with that circle under random selection (N=499)

- The circle in these cases is considered a circular uniform kernel
- In kernel estimation, the kernel is centered on the data locations, in scan statistics, the circle is centered on the grid

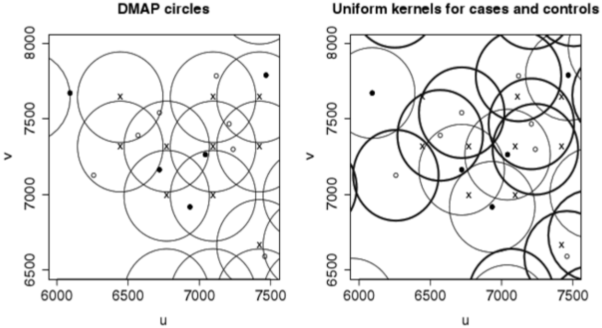

FIG. 6.8 Equivalence between the case/control ratio within circles surrounding grid points (left-hand plot) and the ratio of intensity (density) functions based on circular uniform kernels (right-hand plot). Case locations appear as filled small circles, control locations as open small circles, and grid points as "×" symbols. In the left-hand plot, large circles represent radii of 300 units around each grid point. In the right-hand plot, dark circles represent control kernel radii, and lighter circles represent control kernel radii (both 300 units).

Point Pattern Data: Spatial Scan Statistics

- A scan statistics involves the definition of a moving window and a statistical comparison measurement (count or rate) within the window to the same measurement outside the window.
- Kulldorff (1995, 1997) defines a spatial scan statistic that is similar to the geographical analysis machine, but with an inferential framework.
- Scan statistics tend to consider circular windows with variable radii ranging from the smallest distance between a pair of cases to the user-defined upper bound.
- The circles can be centered on either grid locations (like earlier methods)or the set of locations. Different results will be seen depending on what is chosen.

Point Pattern Data: Spatial Scan Statistics

Setup:

Let $N_{1,in}$ represent the number of case locations and $N_{in} = N_{0,in} + N_{1,in}$ be the total number of people at risk (or number of case and control locations) inside a particular window.

Let $N_{1,out}$ represent the number of case locations and $N_{out} = N_{0,out} + N_{1,out}$ be the total number of people at risk (or number of case and control locations) outside the window.

The test statistic is:

$$T_{scan} = \max\left(\frac{N_{1,in}}{N_{in}}\right)^{N_{1,in}}\left(\frac{N_{1,out}}{N_{out}}\right)^{N_{1,out}} I\left(\frac{N_{1,in}}{N_{in}} > \frac{N_{1,out}}{N_{out}}\right)$$

I(.) is the indicator function (i.e. we only maximize over windows where the observed rate inside the window exceeds that outside the window)

Point Pattern Data: Spatial Scan Statistics

- The maximum observed likelihood ration statistic provides a test of overall general clustering and an indication of the most likely clusters with significance determined by Monte Carlo testing of the constant risk hypothesis.
- SaTScan is a software package that enables this, and we can access it through R satscan (both must be downloaded).