

# Spatial Statistics

## Point Process Data Unit 2

PM569 Spatial Statistics

Lecture 9: November 11, 2016

## Review of point processes

- ▶ Simple stochastic models for point patterns do not have tractable distributions.
- ▶ To test models against data we use Monte Carlo tests (simulation-based).
- ▶ Monte Carlo steps:
  - ▶ Let  $u_1$  be the observed value of a statistic  $U$
  - ▶ Let  $u_i$  be the values of the statistic  $U$  generated by independent random sampling from the distribution of  $U$  under a simple hypothesis  $H_0$  (the null hypothesis)
  - ▶ Let  $u_{(j)}$  denote the  $j$ th largest among the  $u_i$ ,  $i = 1, \dots, s$
  - ▶ Then, under  $H_0$ ,  $P\{u_1 = u_{(j)}\} = s^{-1}$ ,  $j = 1, \dots, s$  and rejection of  $H_0$  on the basis that  $u_1$  ranks  $k$ th largest or higher gives an exact one sided test of size  $k/s$

## Review of point processes

- ▶ Monte Carlo methods are not precisely replicable since they rely on simulated data.
- ▶ An independent set of simulated realizations will result in a different estimated p-value than the first set of realizations.
- ▶ The larger number of simulations the more stable the resulting estimates.
- ▶ We use Monte Carlo methods to test whether our observations are a CSR with homogenous or inhomogeneous Poisson process, cluster process, regular process.

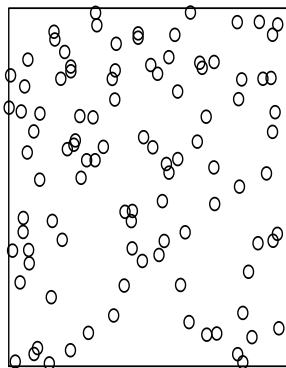
## Review of point processes

- ▶ Testing for CSR:
  - ▶ Often want to adjust for edge effects.
  - ▶ We test for CSR with Ripley's  $K$ , which involves a search window with bandwidth  $h$ .
  - ▶ We test for CSR based on inter-event distances being less than a threshold  $\delta$  with  $H(h)$ .
  - ▶ We test for CSR based on nearest-neighbour distances with  $G(h)$ .
- ▶ The types of spatial processes where the Poisson processes is the building block are:
  - ▶ Homogeneous Poisson process (constant intensity), used for testing CSR.
  - ▶ Inhomogeneous Poisson process (intensity varies across domain), used for testing CSR.
  - ▶ Poisson Cluster process (intensity varies for parents and/or children forming clusters), used for testing clustered patterns.
  - ▶ Simple inhibition processes, Markovian processes (Strauss and pairwise interaction), used for testing regular patterns.

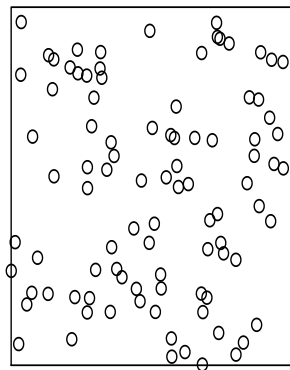
# Point Pattern Data

## Homogeneous Poisson Process (CSR)

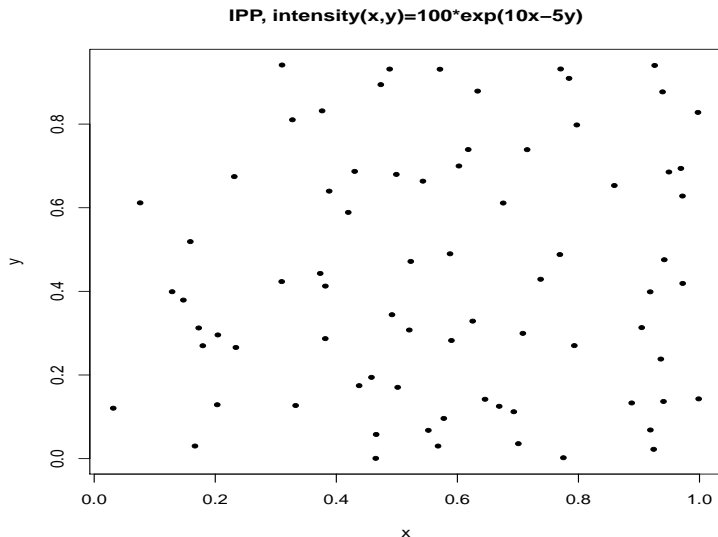
intensity = 100, unit square



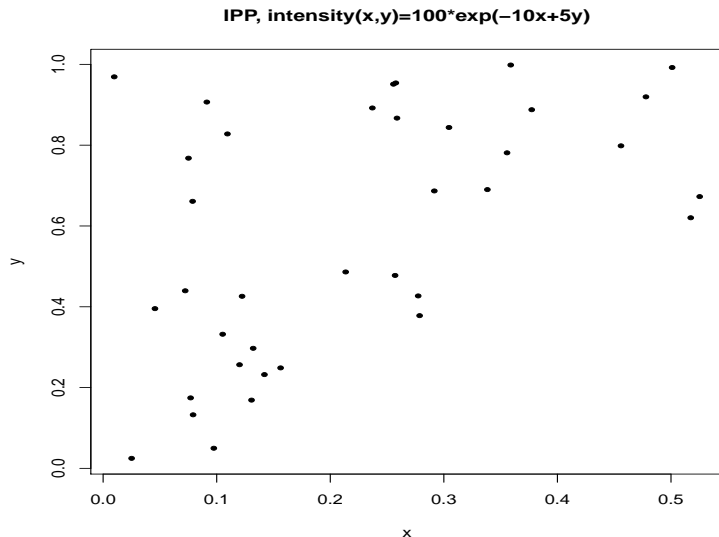
intensity = 1, 10 x 10 square



## Inhomogeneous Poisson Process



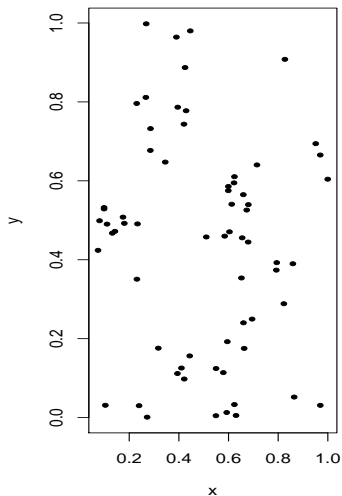
## Inhomogeneous Poisson Process



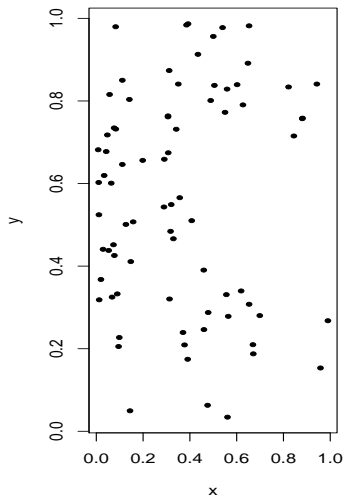
# Point Pattern Data

## Poisson Clustered Process

PCP, (P,O,Spread)=(25,4,0.0025)



PCP, (P,O,Spread)=(25,4,0.005)

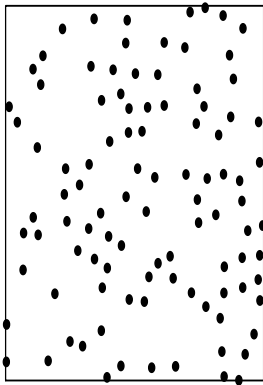




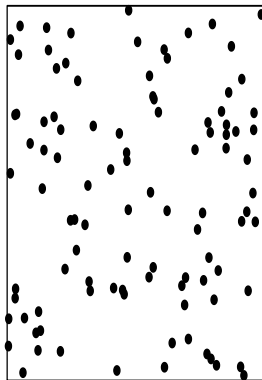
# Point Pattern Data

## Simple Inhibition Process

SIP, distance 0.05



SIP, distance 0.005



# Point Pattern Data

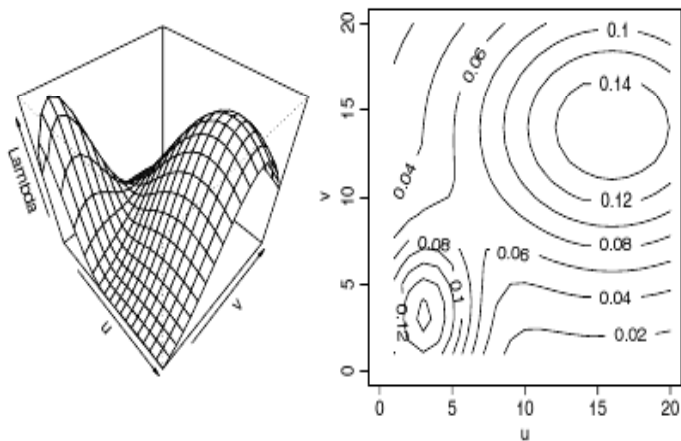
- ▶ Inhomogeneous Point Processes, where the intensity,  $\lambda$ , is not constant.
- ▶ Properties of a spatial point process in terms of the intensity function.
  - ▶ First order properties are described by the intensity function.

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \frac{E[N(dx)]}{|dx|}$$

- ▶ The first order properties are the mean properties of the random process, describing the expected density of events in any location of the region
- ▶ Clusters appear in areas of high intensity
- ▶ Under IPP and CPC, clusters occur due to heterogeneities in the intensity function and individual event locations remain independent of one another

- ▶ First order properties are described by the intensity function
- ▶ Example: consider the constant risk hypothesis
  - ▶ Each person has the same risk of disease, but we expect more cases in areas with more people at risk
  - ▶ Clusters of cases in high population areas will violate CSR but not the constant risk hypothesis
  - ▶ We are interested in clustering of disease events after accounting for known variations in population density
  - ▶ This requires a generalization of the intensity where we define it as a spatially varying function over the study area
  - ▶ As population size increases, so should the expected number of cases

## Inhomogeneous Poisson Process intensity function



**FIG. 5.5** Example intensity function,  $\lambda(s)$ , for a heterogeneous Poisson point process defined for  $s = (u, v)$  and  $u, v \in (0, 20)$ .

- ▶ The inhomogeneous Poisson process shows lack of events between the modes
- ▶ More events around the mode  $(16,14)$  and a narrower peaked area around  $(3,3)$
- ▶ Collections of events suggest areas of higher intensity
- ▶ Single realizations make it hard to identify the specific areas of these modes
- ▶ Useful to simulate multiple realizations of the process

- ▶ Second order properties are described by the inter-relationships between events

$$\lambda(x, y) = \lim_{|dx|, |dy| \rightarrow 0} \frac{E[N(dx)N(dy)]}{|dx||dy|}$$

- ▶ This allows us to describe how often events occur within a given distance of other events
- ▶ The second order properties are similar to variance/covariance of the process
- ▶ Allows us to summarize the spatial dependence between events over a wide range of possible spatial scales
- ▶ The Ripley's K function is a second-order statistic

- ▶ Recall the K function for distance h:

$$K(h) = \frac{E[\# \text{ events within } h \text{ of randomly chosen event}]}{\lambda}$$

- ▶ The second order properties gives us insight into the global aspects of the point pattern
- ▶ Are there general patterns of clustering or regularity with respect to CSR or another pattern?

## Cox processes

- ▶ Spatial clustering with a spatially varying intensity function of the inhomogeneous Poisson process
- ▶ Varying  $\lambda(x)$  and  $\lambda(x)$  is a realization of a stochastic process
- ▶ Property 1) it is a non-negative valued stochastic process

$$\{\Lambda(x); x \in \mathbb{R}^2\}$$

- ▶ Property 2) the events for an inhomogeneous poisson process with intensity function  $\lambda(x)$

$$\{\Lambda(x) = \lambda(x); x \in \mathbb{R}^2\}$$



## Cox processes

- ▶ The Cox process is homogeneous iff  $\Lambda(x)$  is homogeneous:

$$E[\Lambda(x)] = \lambda \forall x$$

$$E[\Lambda(x)\Lambda(x+h)] \text{ depends only on } ||h||$$

## Cox processes

- ▶ The Cox process is linked to the clustered Poisson process
- ▶ Aggregation into clusters may be a result of environmental heterogeneity
- ▶ Clusters of events in regions of high intensity
- ▶ Cox processes are considered doubly stochastic, intensity is heterogeneous but also may be a random quantity
- ▶  $\lambda(x)$  can be drawn from some probability distribution of possible intensity functions over the study area

## Cox processes

$$\Lambda(x) = \mu \sum_{i=1}^{\infty} h(x - X_i)$$

- ▶  $\mu > 0$ ,  $h(\cdot)$  is a bivariate pdf, and  $X_i$  are points from a Poisson process
- ▶ The Cox process can also be thought of as a specific case of a Poisson cluster process with number of offspring having intensity  $\mu$  and dispersion around parents with pdf  $h(\cdot)$

## Cox processes

- ▶ The log-Gaussian Cox process is another form of the Cox process

$$\Lambda(x) = \exp(Z(x))$$

- ▶  $Z(x)$  is a Gaussian process.
- ▶ If  $Z(x)$  is stationary with mean  $\mu$ , variance  $\sigma^2$  and correlation  $\rho(h)$ :
  - ▶  $\lambda = \exp(\mu + 0.5\sigma^2)$
  - ▶  $\gamma(h) = \exp(\sigma\rho(h))$
- ▶ The log-Gaussian Cox process can be fit in R spatstat with the `rLGCP()` function

## Simple Inhibition Process

- ▶ This process is used to describe regular patterns
- ▶ Often related to interactions or contagions where the occurrence of an event raises or lowers the probability of subsequent events nearby
- ▶ Useful for modeling the spread of infectious disease (contagion) or an application where an event precludes the occurrence of other events in a nearby area such as animal territories (inhibition)
- ▶ **Contagion** typically refers to the increased likelihood of events occurring near other events
- ▶ **Inhibition** may be absolute, where there is a specified distance around which *no* other events may occur, or it may be probabilistic where there is small but positive probability of an event occurring near other events

## Simple Inhibition Process

- ▶ Models for inhibition or contagion processes are Markov point processes or Gibbs processes
- ▶ The general idea is to take a CSR and "delete" points within a distance less than a threshold  $\delta$
- ▶ Under a Markov process, the existence of an event in a region depends on the locations of events in a neighbourhood (where neighbourhoods are within regions)
- ▶ There are two ways to do this: 1) to simulate CSR then delete all within a distance  $\delta$ , and 2) to simulate CSR, record when event was simulated, then delete an event if it is within distance  $\delta$  of an older event

## Simple Inhibition Process

- ▶ We use the packing intensity to describe simple inhibition processes:

$$\tau = \lambda\pi(\delta/2)^2$$

Where  $\lambda$  is the intensity, giving  $\tau$  to be the proportion of the region  $A$  covered by non-overlapping discs of diameter  $\delta$

## Simple Inhibition Process

- ▶ For simple inhibition process 1) we take a a Poisson process with intensity  $\rho$  and thin it by the deletion of pairs of events that are less than  $\delta$  apart
- ▶ In this case, the probability that an event "survives" is  $\exp(-\pi\rho\delta^2)$  giving the intensity of a simple inhibition process as:

$$\lambda = \rho \exp(-\pi\rho\delta^2)$$

- ▶ The second order properties can be expressed as:

$$\lambda(h) = \rho^2 \exp(-\rho U_\delta(h)) \quad h \geq \delta$$

- ▶  $\lambda(h) = 0$  when  $0 < h < \delta$ , and  $U_\delta(h)$  is the area of the union of two discs with equal radius  $\delta$  and centers distance  $h$  apart



## Simple Inhibition Process

- ▶ For simple inhibition process 2) we take a a Poisson process with intensity  $\rho$  and thin it by the deletion of pairs of "older" events that are less than  $\delta$  apart
- ▶ The expressions are the same as for process 1) but with the addition of the sequential piece (this process is referred to as the simple sequential inhibition process)
- ▶ Let  $X_i$  be a sequence of  $n$  events in  $A$ , and  $d(x, y)$  be the distance between two points  $x$  and  $y$ . Then:
  - ▶  $X_1$  is simulated from a uniform distribution in  $A$
  - ▶ Given (past)  $\{X_j = x_j, j = 1, \dots, (i - 1)\}$ , then  $X_i$  (present) is uniformly distributed on the intersection of  $A$  with  $\{y : d(y, x_j) \geq \delta, j = 1, \dots, (i - 1)\}$
- ▶ So the simple sequential inhibition process has packing intensity:

$$\tau = \frac{n\pi(\delta/2^2)}{|A|}$$

- ▶ In R (spatstat), the functions for thinning processes 1) and 2) described above are called `rMaternI` and `rMaternII`
- ▶ The simple sequential inhibition process, called `rSSI` is similar but slightly different:
  - ▶ Each new point is generated uniformly in the window and independently of preceding points
  - ▶ If a new point lies within distance  $\delta$  from an existing point then it is rejected and another random point is generated
  - ▶ The SSI process ends when no points can be added

## Markov point processes

- ▶ The general idea of a Markov point process lies in conditioning, whereby the existence of an event in a finite region  $A$  depends on the locations of events in a neighbourhood
- ▶ Inhibition processes are a special form of Markov process: the conditional intensity of an event at a point  $x$  given the realization of the process in the remainder of the region  $A$  depends on the existence (or otherwise) of an event within distance  $\delta$  of  $x$
- ▶ General Markov processes were introduced by Ripley and Kelly (1977)
- ▶ Markov point processes are characterized by the likelihood ratio with respect to a Poisson process of unit intensity

## Markov point processes

- ▶ Let's call the likelihood ratio  $f(\cdot)$
- ▶ If  $\mathbf{X} = \{x_1, \dots, x_n\}$  denotes a finite set of points in  $A$  then  $f(\mathbf{X})$  indicates how much more likely is the configuration of events  $\mathbf{X}$  than a homogeneous point process (with unit intensity)
- ▶ We can factorize the likelihood ratio to:

$$f(\mathbf{X}) = \alpha \prod_{i=1}^n g_i(x_i) \prod_{j>i} g_{ij}(x_i, x_j) \dots g_{12\dots n}(x_1, x_2, \dots, x_n)$$

- ▶ Where  $\alpha$  is a normalizing constant
- ▶ We also define two points  $x$  and  $y$  in  $A$  to be neighbours if  $d(x, y) < \delta$  for some  $\delta > 0$  where  $d(x, y)$  is the distance between  $x$  and  $y$
- ▶ We also define a clique (recall areal data) as a set of mutual neighbours, and the neighbourhood of  $x$  to be the set of points  $\{y \in A : 0 < d(x, y), \delta\}$

## Markov point processes

- ▶ The point process with these definitions is Markov with range  $\delta$  if the conditional intensity at the point  $x$  given the configuration of the other events in  $A$  depends only on the configuration in the neighbourhood of  $x$
- ▶ The  $g$ -functions from the above equation are unity *unless* the  $x$  form a clique

## Examples of Markov point processes: the Strauss process

$$f(\mathbf{X}) = \alpha \beta^n \gamma^p$$

- ▶ Where  $\alpha$  is the normalizing constant,  $\beta$  is the intensity of the process,  $\gamma$  is the interaction between neighbours, and  $p$  is the number of distinct pairs of neighbours in  $\mathbf{X}$
- ▶ If  $\gamma = 1$  then the Strauss process gives a Poisson process with intensity  $\beta$
- ▶ if  $\gamma = 0$  then the Strauss process gives a simple inhibition process because no two events may be neighbours
- ▶ In R spatstat, the Strauss process is simulated with `rStrauss`

## Examples of Markov point processes: the pairwise interaction process

$$f(\mathbf{X}) = \alpha \beta^n \prod_{i \neq j} h\{d(x_i, x_j)\}$$

- ▶ Where  $\alpha$  is the normalizing constant,  $\beta$  is the intensity of the process,  $h(d)$  is non-negative for all distances and the product is over all pairs of distinct points in  $\mathbf{X}$
- ▶ The additional restriction is that  $h(d)$  is bounded and that  $h(d) = 0$  for all distances less than some  $\delta > 0$
- ▶ This restriction limits the number of events in  $A$  by imposing a minimum allowable distance  $\delta$  between any two events
- ▶ The pairwise interaction process may be fit in R `spatstat` using the `rmh` function

## Examples of Markov point processes: the pairwise interaction process

- ▶ The pairwise interaction process may be simulated using the following steps (MCMC):
  1. For the initial realization, consider  $n$  points  $\{x_1, \dots, x_n\}$
  2. Delete one of the points in  $\{x_1, \dots, x_n\}$
  3. Generate a point  $y$  from a uniform distribution in  $A$ , and accept  $y$  with probability  $p(y)$
  4. Repeat 2-3 until the MCMC converges