Spatial Statistics

PM569 Lecture 11: Point Pattern 2

Meredith Franklin

Division of Biostatistics, University of Southern California

November 15, 2019

Point Pattern Data

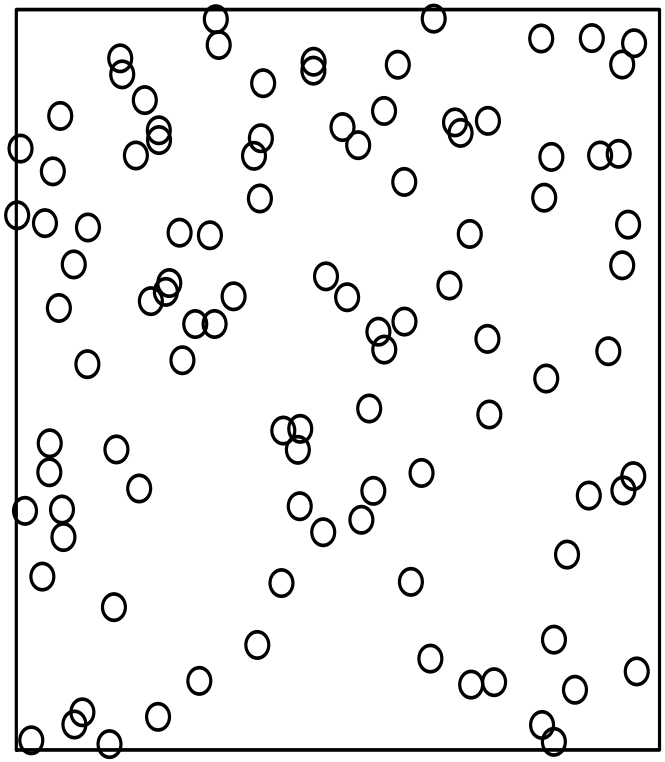
Review of point processes

- 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 (or radius r).
  - We test for for CSR based on nearest-neighbour distances with G(h).
  - We test for CSR based on inter-event distances being less than a threshold δ with H(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.

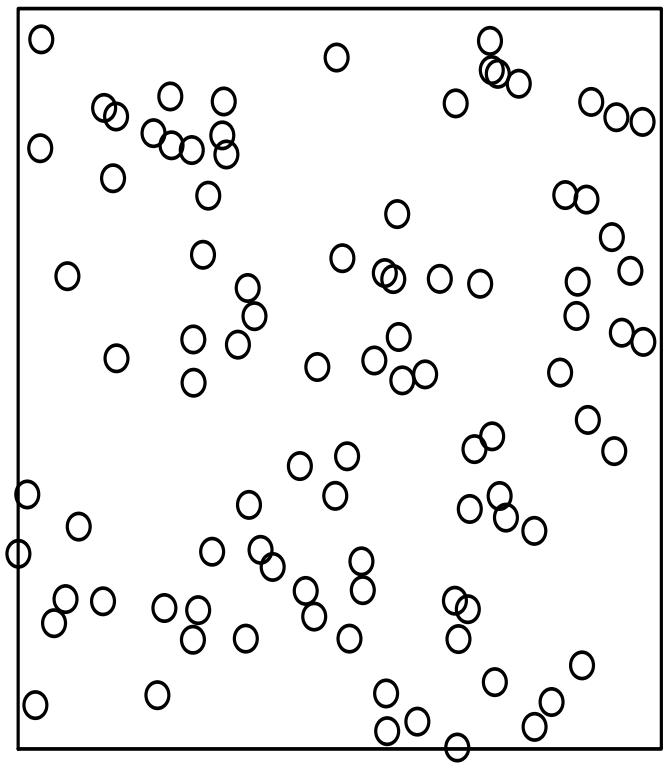
Point Pattern Data

Homogeneous Poisson Process (CSR)

ıtensity = 100, unit squar

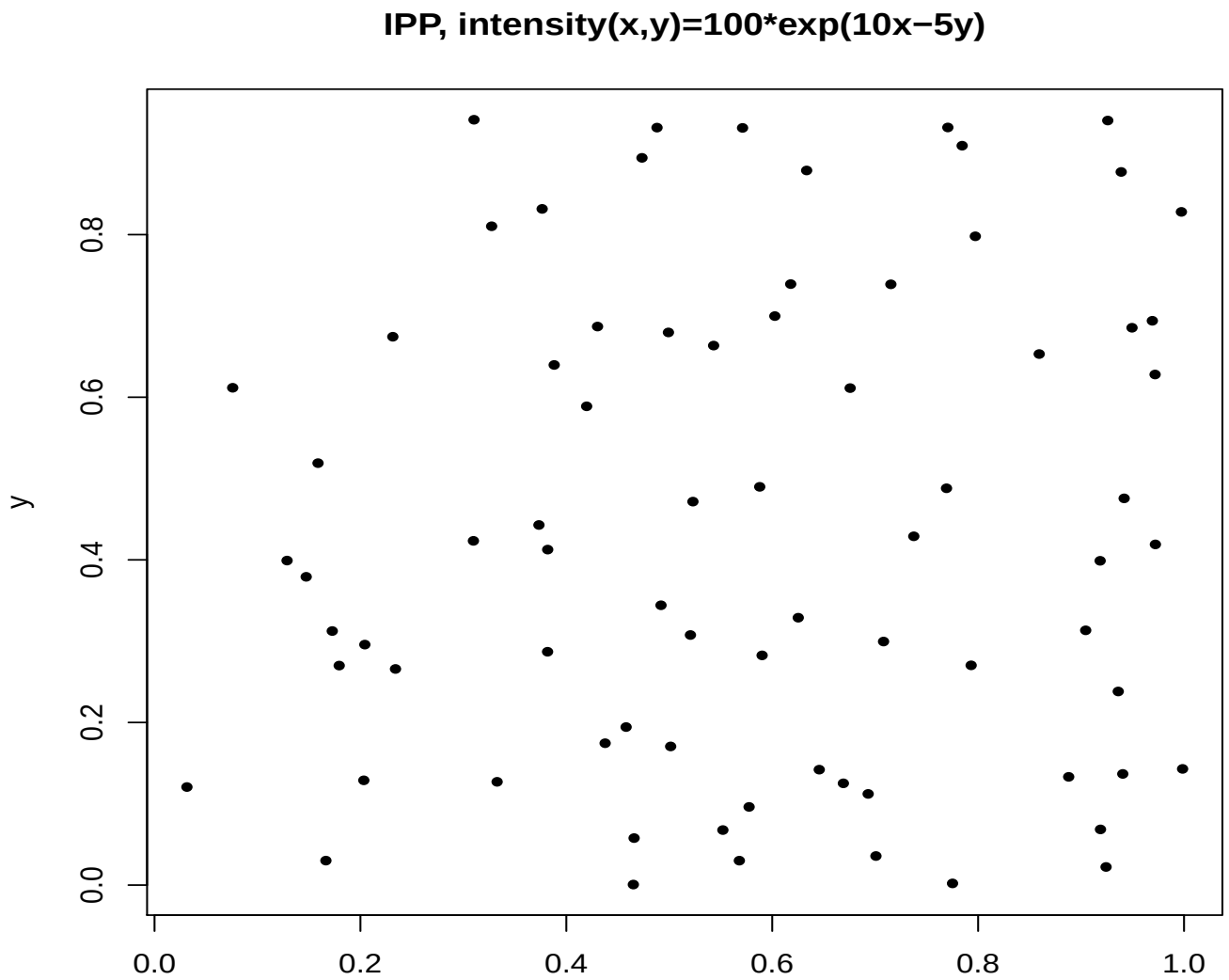


tensity = 1, 10 x 10 squa



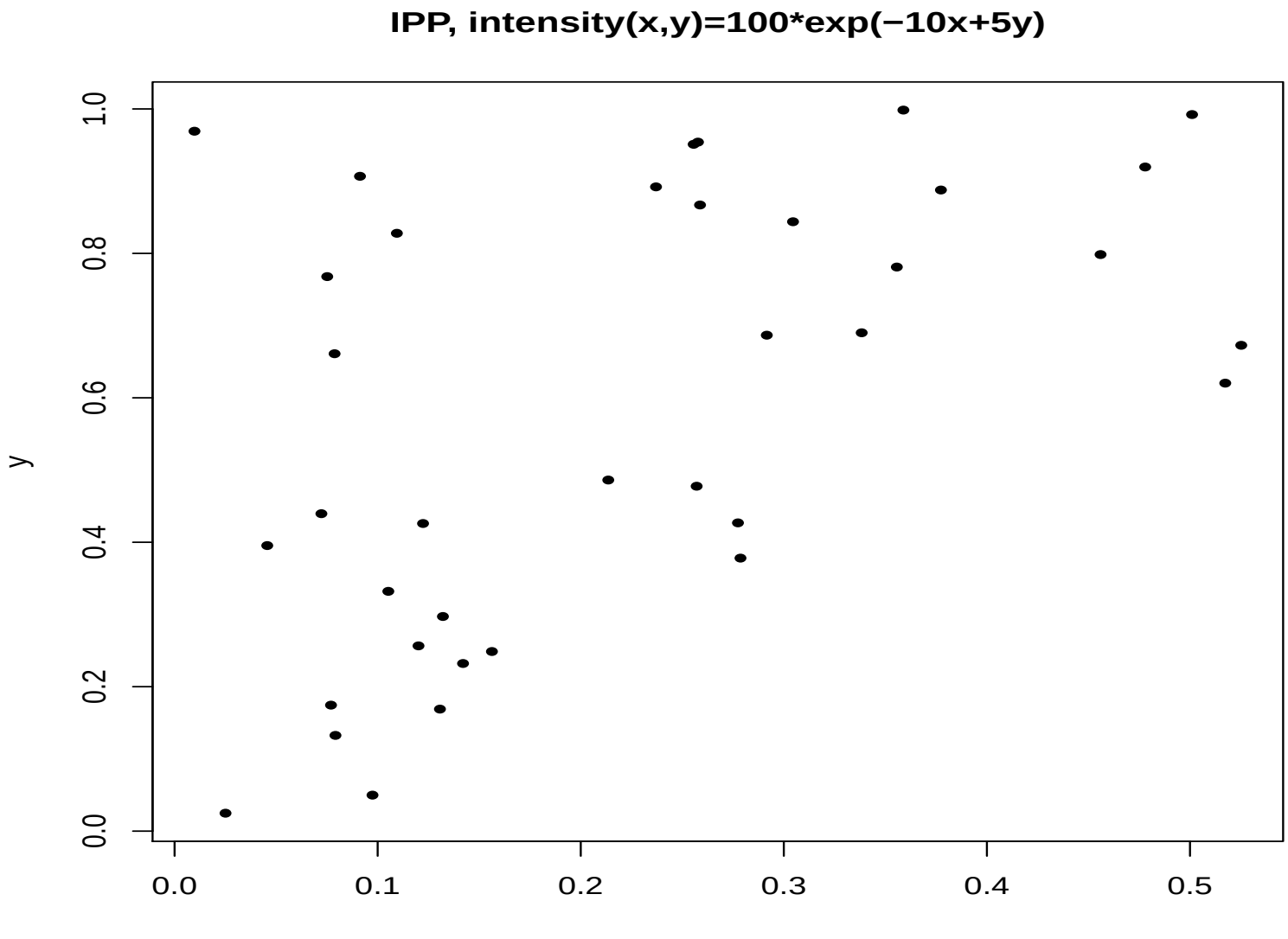
Point Pattern Data

Inhomogeneous Poisson Process



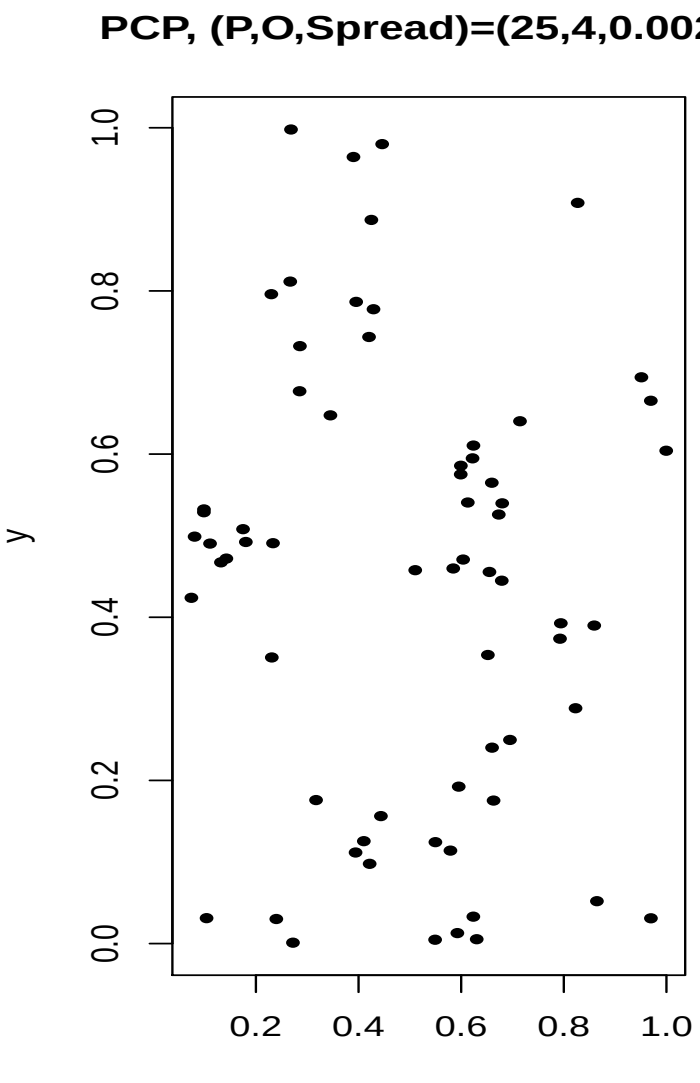
Point Pattern Data

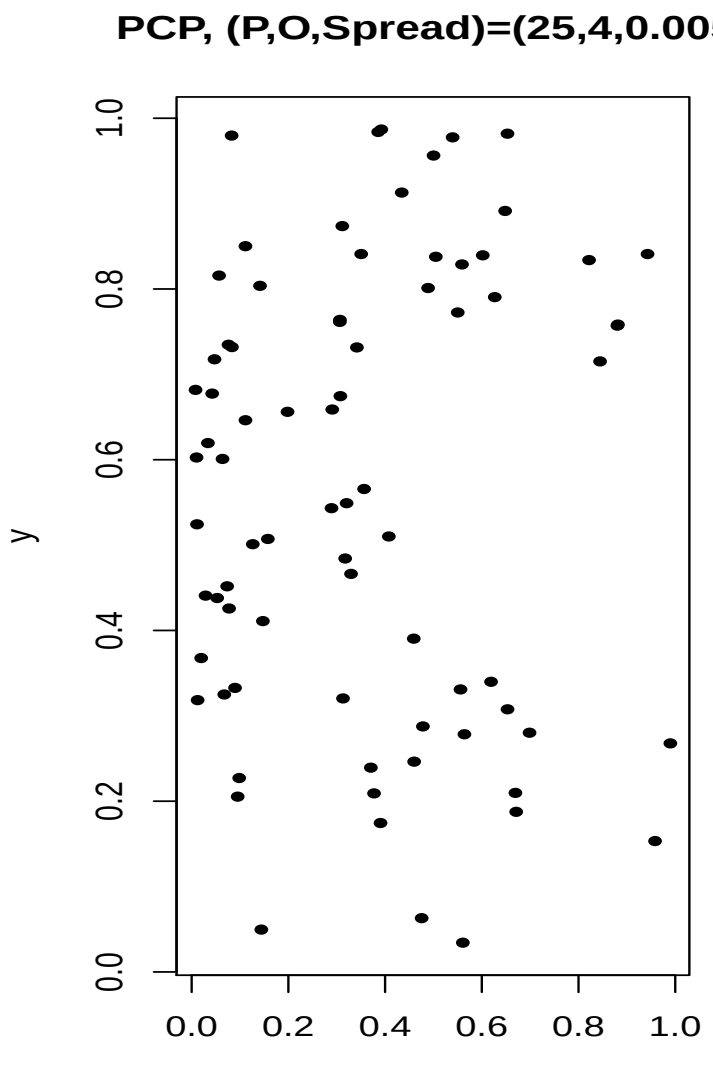
Inhomogeneous Poisson Process



Point Pattern Data

Poisson Clustered Process

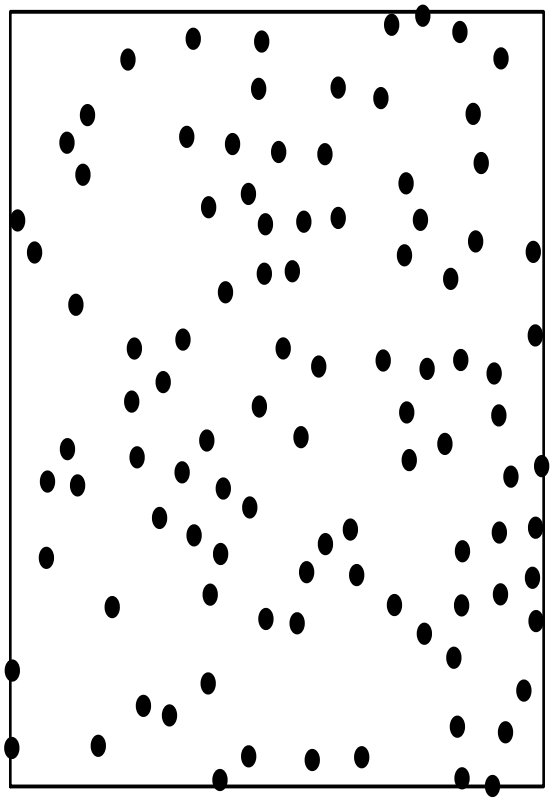




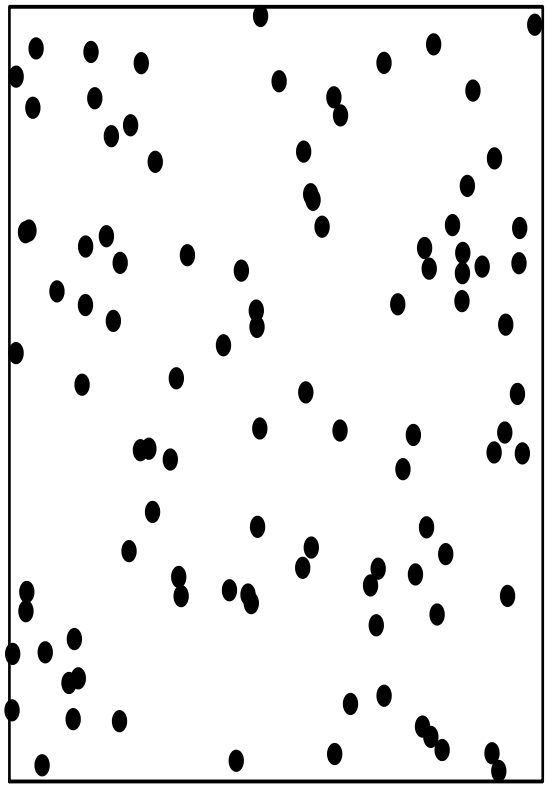
Point Pattern Data

Simple Inhibition Process

SIP, distance 0.05



SIP, distance 0.005



Point Pattern Data: Inhomogeneous Poisson Process

- A generalization of the HPP is the Inhomogeneous (heterogeneous) Poisson Process (IPP). The IPP occurs when the intensity λ is not constant over the region.
- Many cases homogeneity in intensity is not realistic, for example the locations of trees in a forest may be irregular due to geographic features such as soil, rock, slope or other terrain irregularities.
- In the case of IPP, the intensity is a function that varies spatially, $\lambda(s)$.
- The IPP does not define cluster process, but rather a

Point Pattern Data: Inhomogeneous Poisson Process

- Inhomogeneous Point Processes: intensity, λ, is not constant. It is a function $\lambda(s)$ of the locations $S \in D$.
- Properties of a spatial point process in terms of the intensity function.
  - First order properties are described by the intensity function.

$$\lambda(s) = \lim_{|ds| \to 0} \frac{E[N(ds)]}{|ds|}$$

- The first order properties are the mean properties of the random process that describe the expcted density of events in any location of the region.
- The number of events occurring within a finite region D is a random variable following a Poisson distribution with mean $\int_D \lambda(s)ds$
- given the total number of events N occuring in D, the N events represent an independent random sample of N locations with the probability of sampling a particular point S proportional to $\lambda(s)$.

Point Pattern Data: Inhomogeneous Poisson Process

- For an inhomogeneous Poisson process, $\lambda(s)$ is closely related to the density of events over the domain.
- Density estimators provide an estimate of intensity (e.g. kernel density).

Point Pattern Data: Inhomogeneous Poisson Process

- We can estimate the intensity function in different ways: parametrically by defining a specific function or non-parametrically using kernel smoothing.
- A parametric form is a function of x and y.
- A non-parametric form is:

$$\hat{\lambda}(s) = \frac{1}{h^2} \sum_i \kappa(\frac{||s - s_i||}{h})/q(||s||)$$

Where $\kappa(s)$ is a kernel function and $q||s||$ is a boundary correction. The distance h is our bandwidth for smoothing

Point Pattern Data: Inhomogeneous Poisson Process

- There are various kernel functions, but a quadratic function is often used:

$$\kappa(s) = \frac{3}{\pi}(1 - ||s||^2)^2$$

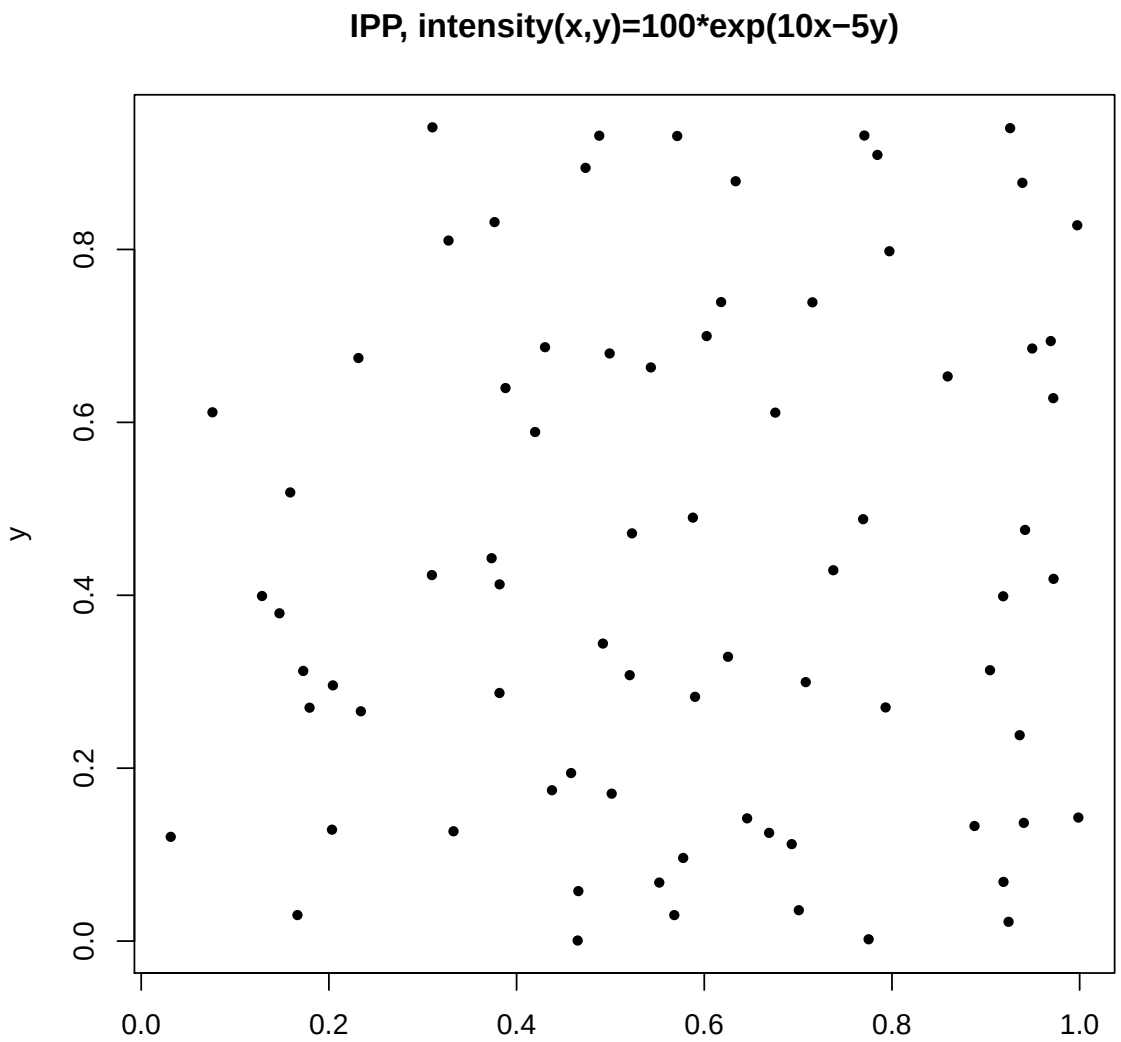
Point Pattern Data: Inhomogeneous Poisson Process

Parametric Intensity

- Example of varying intensity function $\lambda(s)$ could be that intensity varies with location due to environmental heterogeneity
- Example if D i a square unit and N(D)=100
- $\lambda(x, y) = 100 * \exp(10x - 5y)$
- $\lambda(x, y) = 100 * \exp(-10x + 5y)$

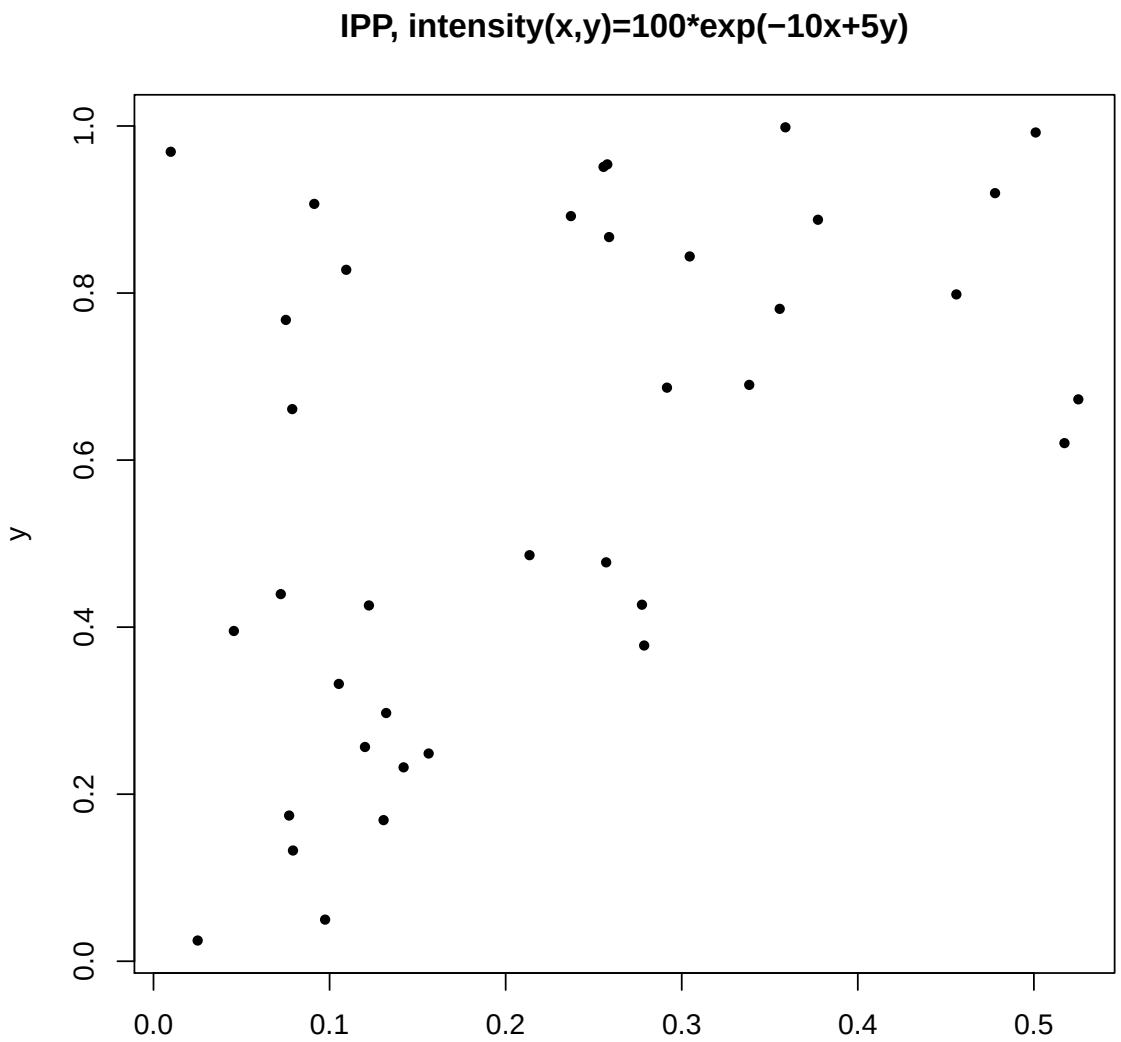
Point Pattern Data: Inhomogeneous Poisson Process

Parametric Intensity



Point Pattern Data: Inhomogeneous Poisson Process

Parametric Intensity



Point Pattern Data: Inhomogeneous Poisson Process

Parametric Intensity

- Or we might see that cases of respiratory disease differ with respect to distance from a point source of environmental pollutions_0

$$\lambda(s) = \lambda_0(s) f(||s - s_0||, \theta)$$

- Where $\lambda_0(s)$ models the variation in population density
- f(u,θ) models how the impact of the source varies with distance u (s-s_0) and angle θ

Point Pattern Data: Inhomogeneous Poisson Process

Non-Parametric Intensity Inhomogeneous Poisson Process intensity function

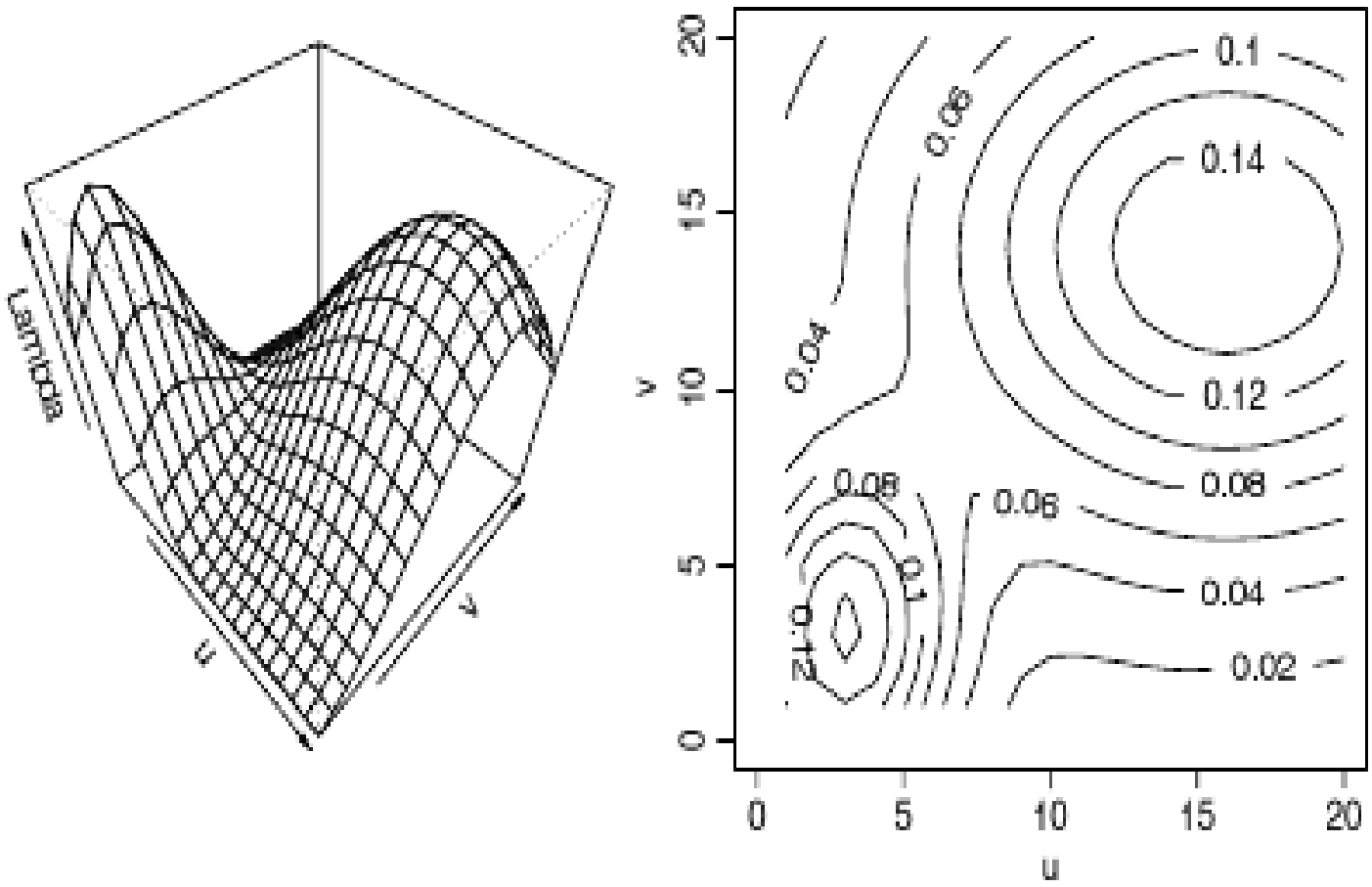


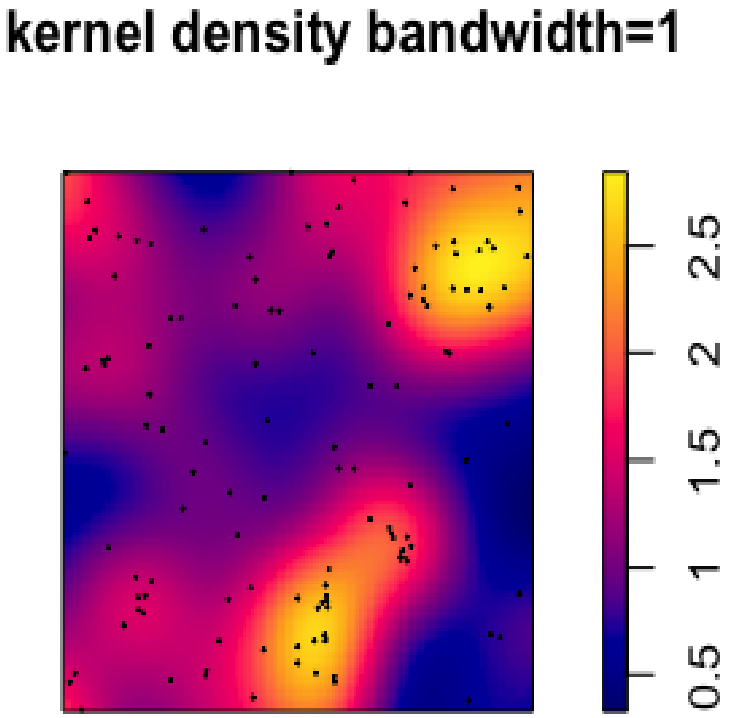
FIG. 5.5 Example intensity function, $\lambda(s)$, for a heterogeneous Poisson point process defined for

Point Pattern Data: Inhomogeneous Poisson Process

- 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
- In order to do kernel estimation we need to choose:
  - the kernel type (e.g. Gaussian)
  - the bandwidth (e.g. radius of the area where smoothing is applied). Smaller bandwidths produce more localized densities.

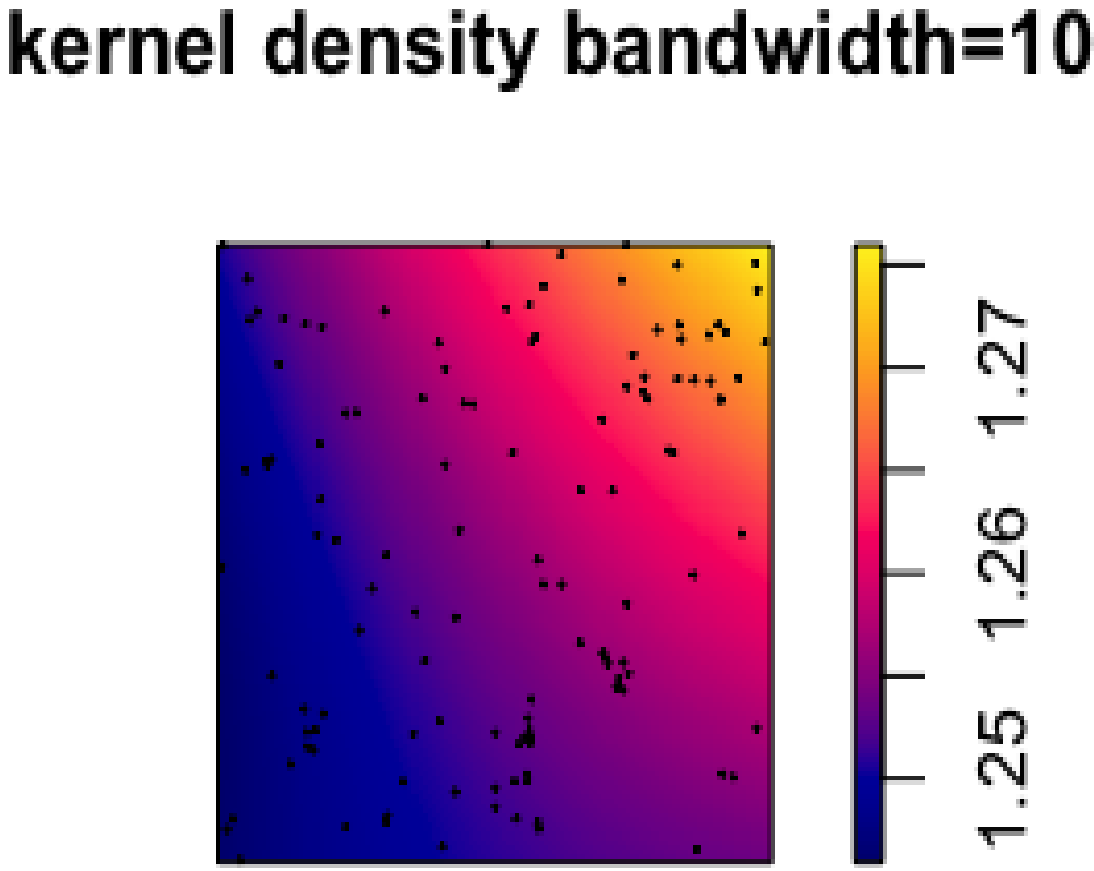
Point Pattern Data: Inhomogeneous Poisson Process

Several R packages produce kernel estimation: `density()` in base R, `ksmooth()` in spatstat, `kernel2d()` in splancs.
Example: finpines data
plot(density(finpines,1))
points(finpines,pch=19,cex=0.1)



Point Pattern Data: Inhomogeneous Poisson Process

```
plot(density(finpines,10))
points(finpines,pch=19,cex=0.1)
```



Point Pattern Data: Inhomogeneous Poisson Process

- First order properties describe the mean of the process (mean number of events per unit area)
- Second order properties are described by the inter-relationships between events (e.g. variance/covariance between locations)
- The second order intensity function can be written (for two locations s and u):

$$\lambda(s,u) = \lim_{|ds|,|du| \to 0} \frac{E[N(ds)N(du)]}{|ds||du|}$$

- Under IPP there are heterogeneities in the intensity function and individual event locations remain independent of one another.
- 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

Point Pattern Data

- Recall the K function for distance h (or radius r):

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

- 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?

Point Pattern Data

- 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 jth largest among the u_i, $i = 1, ..., s$
  - Then, under H_0, $P\{u_1 = u_{(j)}\} = s^{-1}, j = 1, ..., s$ and rejection of H_0 on the basis that u_1 ranks kth largest or higher gives an exact one sided test of size k/s

Point Pattern Data

- 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 homogeneous or inhomogeneous Poisson process, cluster process, regular process.

Point Pattern Data: Poisson Cluster Process

Poisson cluster process

- A spatial point process where each event belongs to a cluster
- There is a parent event that produces a random number of offspring
- Parent events are usually a realization of an Poisson process with intensity $\lambda(s)$
- We have i parents, and each parent produces a random number of offspring, O_i
- The O_i are distributed within h_i of the parent and follow a bivariate probability distribution

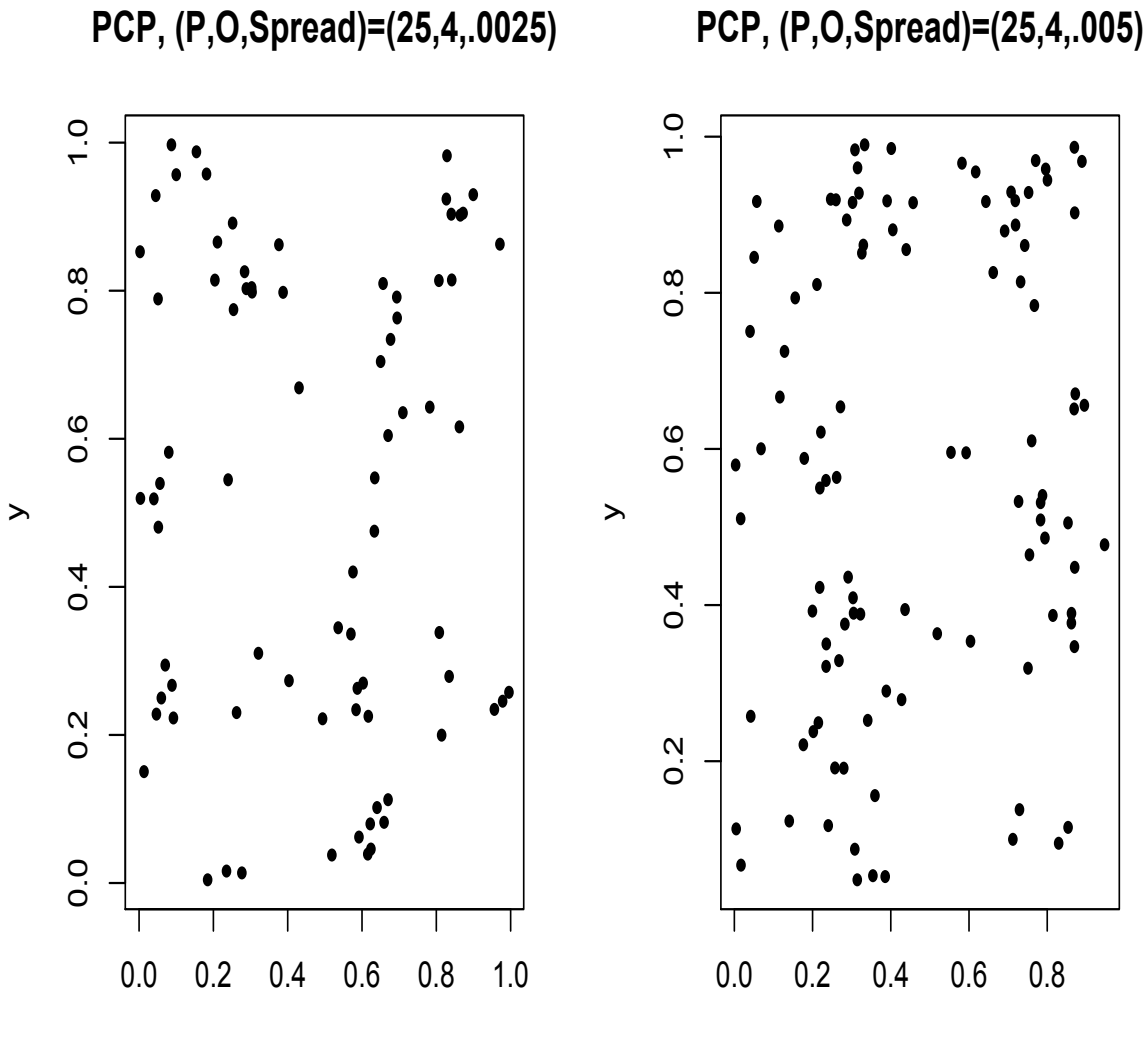
Point Pattern Data: Poisson Cluster Process

Poisson cluster process

- Can have homogeneous cluster processes where the intensity of the offspring around a parent is constant λ
- Or an inhomogeneous cluster process where the intensity of the offspring around a parent is not constant across domain $\lambda(s)$
- Parent events are usually a realization of an inhomogeneous Poisson process with intensity $\lambda(s)$ distribution

Point Pattern Data: Poisson Cluster Process

Poisson cluster process



Point Pattern Data

Poisson cluster process

- Left, we have a unit square as our D with intensity of parents $= 25$ and number of offspring $= 4$ and variation around parents $= 0.00025$
- Right, we have a unit square as our D with intensity of parents $= 25$ and number of offspring $= 4$ and variation around parents $= 0.005$

Point Pattern Data

Poisson cluster process

- Example of PCP: distribution of insect larvae or tree seeds
- Neyman Scott assumptions of homogeneous Poisson cluster process:
  - Parent events are realizations of a Poisson process with intensity ρ
  - Each parent i produces a random number of offspring S_i and the $S - i$ are iid
  - The positions of offspring wrt the parent are iid with bivariate pdf

Point Pattern Data

Poisson cluster process

- Isotropy in our pdf means it must be radially symmetric
- Example is the radially symmetric Gaussian distribution

$$h(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp(-\frac{x_1^2 + x_2^2}{2\sigma^2})$$

- where σ^2 are the offspring

Point Pattern Data

Poisson cluster process

- Neyman Scott assumptions of homogeneous Poisson cluster process
- Use ρ for intensity of parents, $E(s) = \mu$ is the expected number of offspring
- Then the overall intensity of a clustered process is $\lambda = \rho\mu$ (1st order intensity)
- Need second order properties of this process to derive $K(h)$ under PCP
- Second order intensity $\lambda_2 = \lambda^2 + \rho E[S(S-1)]h_2(s_i - s_j)$

Point Pattern Data

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

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

- Property 2) the events for an inhomogenous Poisson process (IPP) with intensity function $\lambda(x)$

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

Point Pattern Data

Cox processes

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

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

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

Point Pattern Data

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

Point Pattern Data

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 μ 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 μ, variance σ^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

Point Pattern Data

Fitting point process models

- Given our set of observed point events $\{x_1, ...x_n\}$ in region D 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 a parametric form of the intensity of an inhomogeneous Poisson process
  - Example: fitting the parameters ρ, μ and σ^2 of a clustered process
  - $\lambda(x, y) = \exp(\theta_0 + \theta_1 x + \theta_2 y)$
- We use familiar fitting methods: Least squares, Maximum Likelihood and non-parametric methods.

Point Pattern Data

Fitting point process models: Inhomogeneous PP

- In `spatstat` we require the intensity function $\lambda(x, y)$ to be loglinear in the θ parameters
  - $\log(\lambda(x, y)) = \theta S(x, y)$ where $S(x, y)$ is a function of the location referenced by x,y coordinates
- In practice $S(x, y)$ can be a function of the spatial coordinates, an observed covariate, or both.

Point Pattern Data

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

Point Pattern Data

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

Point Pattern Data

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`

Fitting point process models: Maximum Likelihood

- To fit inhomogeneous Poisson and pairwise interaction processes we need to rely on likelihood methods
- Recall for the inhomogeneous Poisson process:
  - $N(D)$ is Poisson with mean $\int_D \lambda(x)dx$
  - Conditional on $N(D) = n$, the n events in A form an independent random sample from D 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 $\lambda(u, \mathbf{x})du$

Point Pattern Data

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

Point Pattern Data

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"

Point Pattern Data

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

Point Pattern Data

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 $\exp(\alpha + \beta x + \gamma y)$

Point Pattern Data: Density Based Clustering

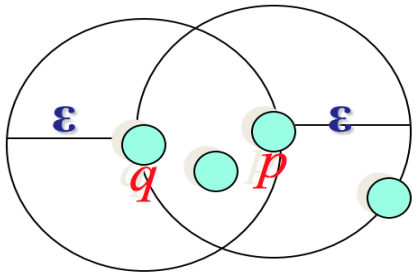
- Density-based spatial clustering is a data clustering algorithm that falls under "machine learning" techniques.
- It was popularly implemented in 1996 by Ester et al as "DBSCAN" (Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96): A density-based algorithm for discovering clusters in large spatial databases with noise).
- Density-based approaches like DBSCAN model clusters as high-density clumps of points.

Point Pattern Data: Density Based Clustering

- Density based clustering is different than other clustering algorithms such as k-means clustering:
  - In k-means clustering, each cluster is represented by a centroid, and points are assigned to whichever centroid they are closest to.
  - In DBSCAN, there are no centroids, and clusters are formed by linking nearby points to one another.
  - In k-means you have to choose the number of clusters before running the algorithm.
  - In DBSCAN you choose the minimum number of points that must be in a cluster of radius ϵ to consider it a cluster.
  - Clusters from density-based algorithms are arbitrary in shape.

Point Pattern Data: Density Based Clustering

The ϵ defines the neighbourhood, where points within ϵ radius from another point are considered part of a cluster as long as it fulfills that there are a certain number of MinPts.



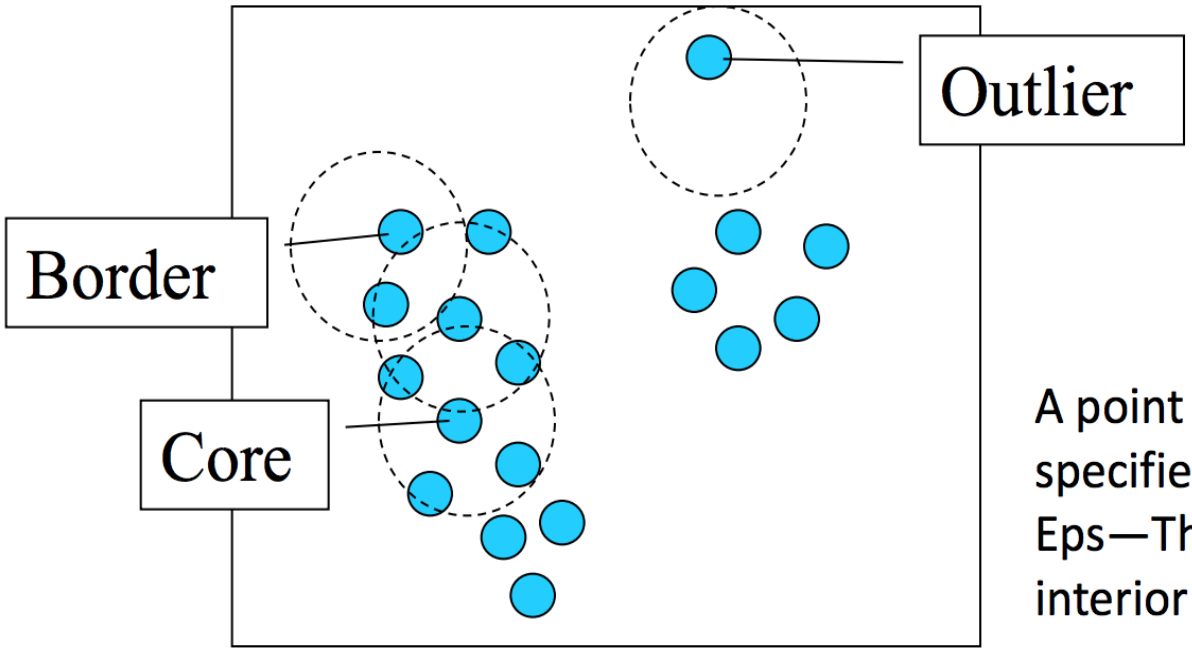
ε-Neighborhood of p

ε-Neighborhood of q

Density of p is "high" (MinPts = 4)

Density of q is "low" (MinPts = 4)

Point Pattern Data: Density Based Clustering



Given ε and *MinPts*, categorize the objects into three exclusive groups.

A point is a core point if it has more than a specified number of points (MinPts) within Eps—These are points that are at the interior of a cluster.

$\varepsilon = 1\text{unit}, \text{MinPts} = 5$

A border point has fewer than MinPts within Eps, but is in the neighborhood of a core point.

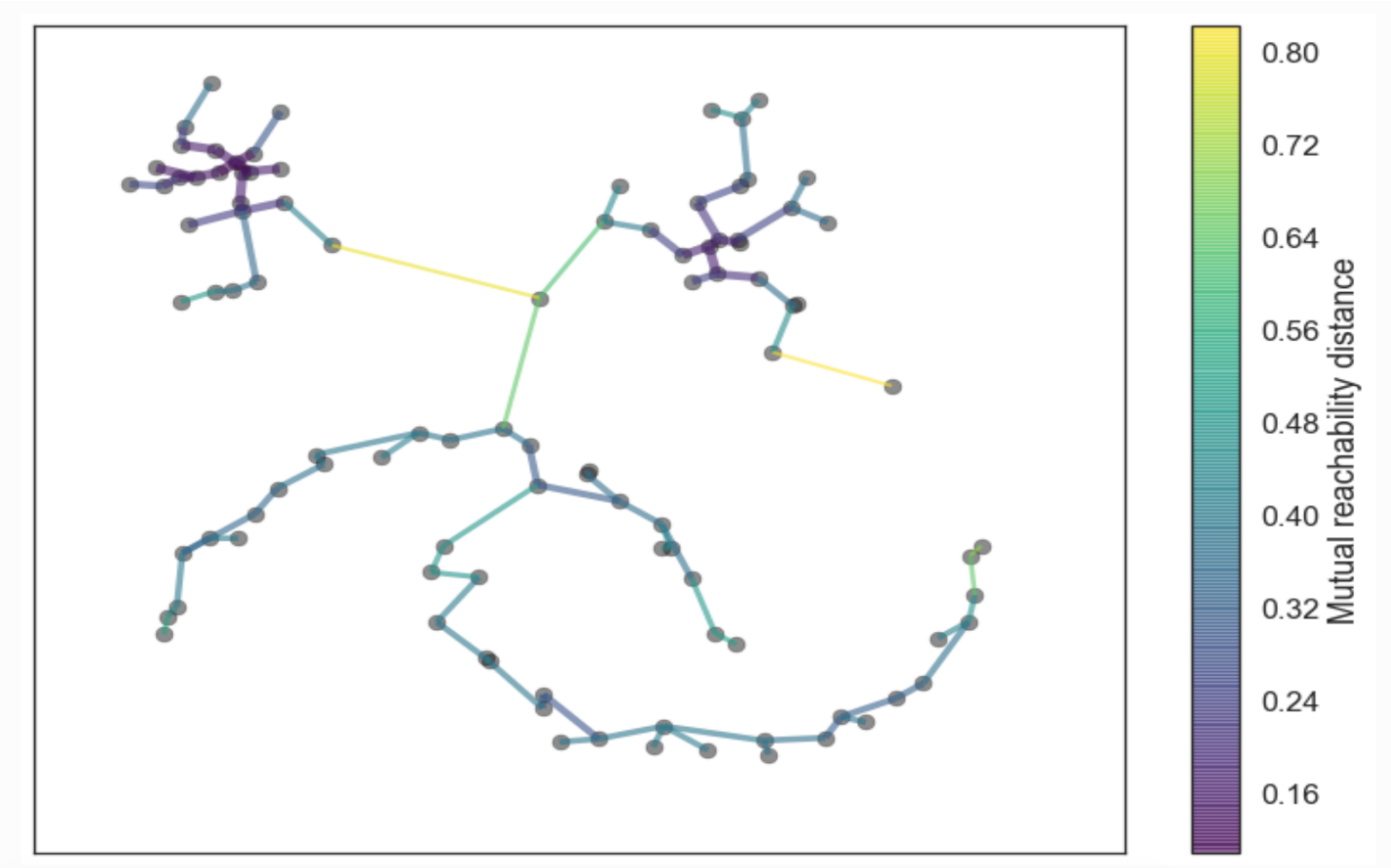
A noise point is any point that is not a core point nor a border point.

Point Pattern Data: Density Based Clustering

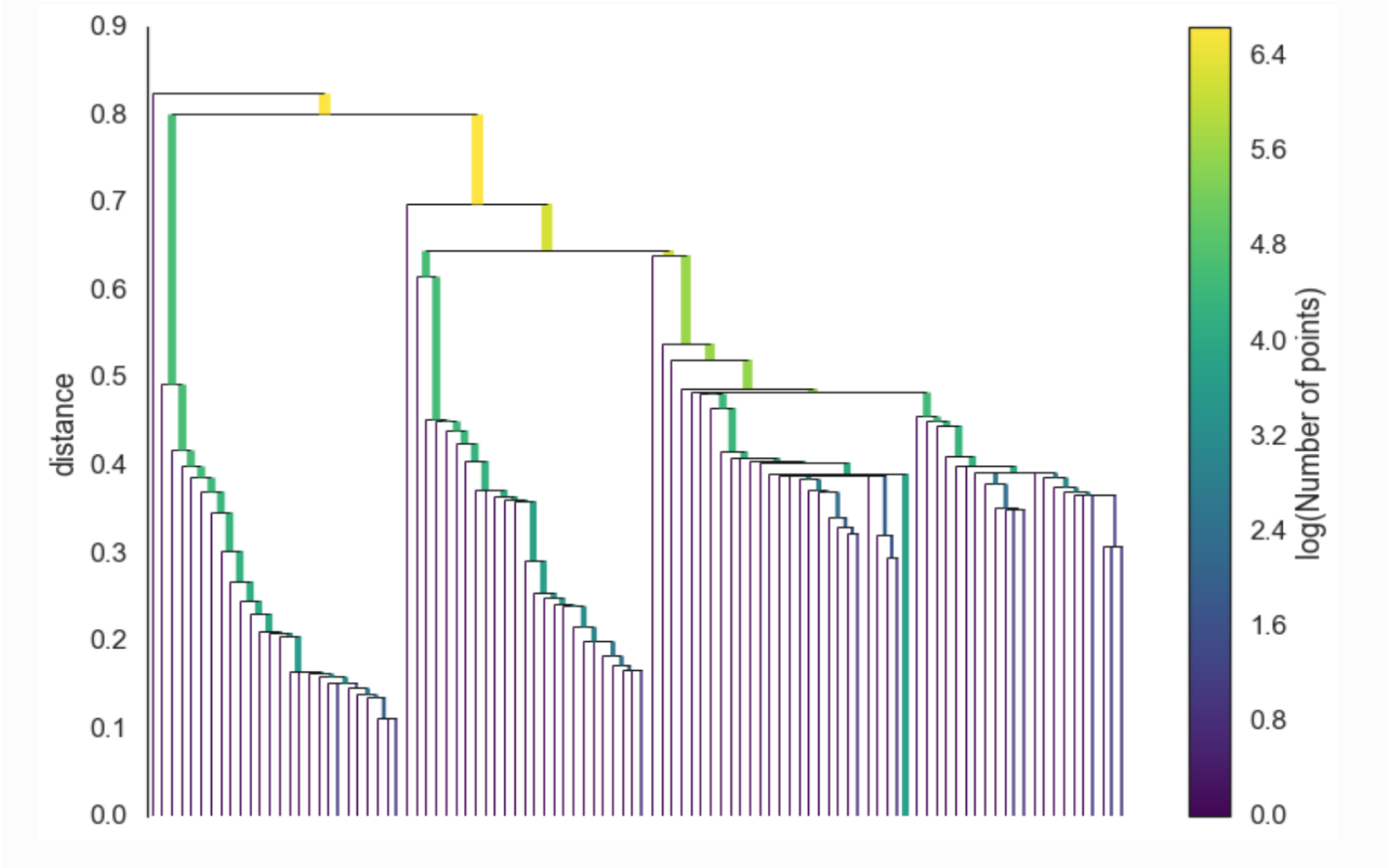
Here is a good visualization of DBSCAN results:
https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/
A better version in Hierarchical DBSCAN, where we do not need to set the ϵ neighbourhood. With HDBSCAN we do set the minimum number of points that we wish to have in a cluster (2 to n). A larger MinPts means larger clusters.

- HDBSCAN uses the concept of mutual reachability, where we look at distances that connect all of the points
- A tree (hierarchy) is formed based on these distances

Point Pattern Data: Density Based Clustering



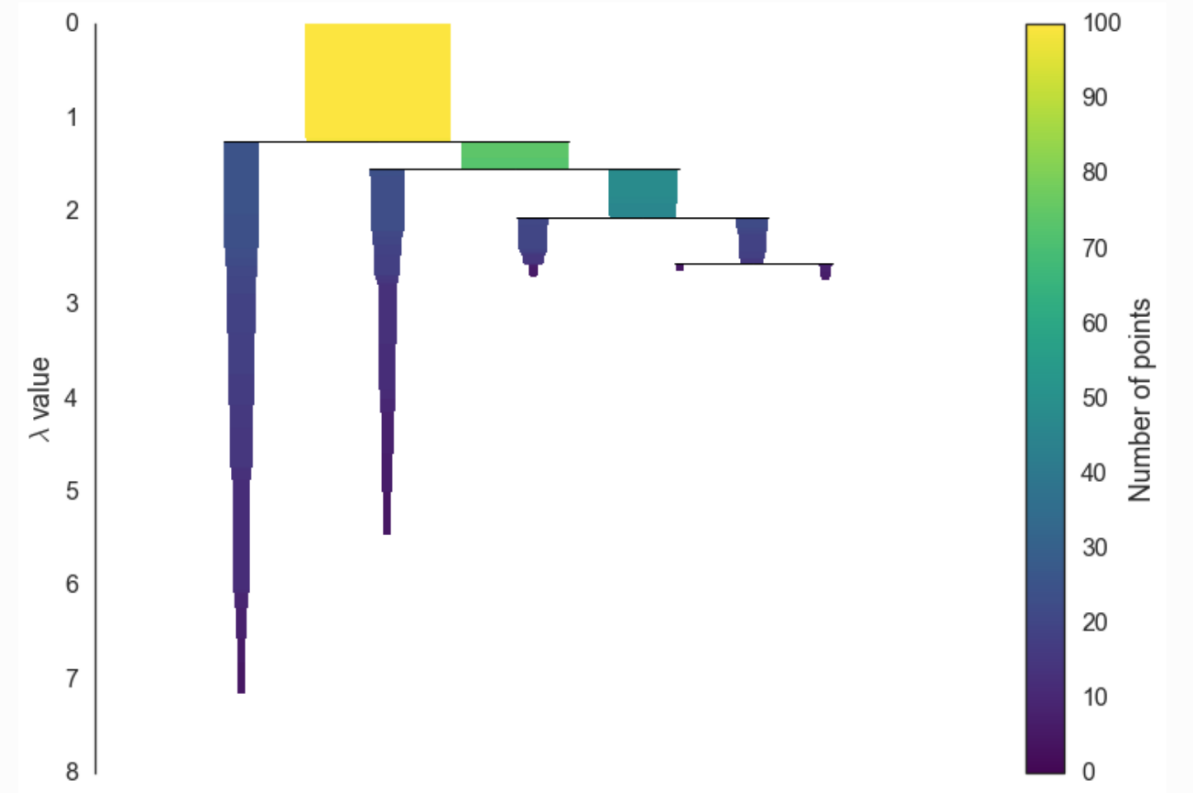
Point Pattern Data: Density Based Clustering



If we cut the tree at a given distance (ϵ) that is giving us the same estimate as DBSCAN

Point Pattern Data: Density Based Clustering

The tree has to be condensed somehow and provide us the main clusters. Using the minimum cluster size the algorithm walks through the hierarchy and at each split asks if one of the new clusters created by the split has fewer points than the minimum cluster size.



Point Pattern Data: Density Based Clustering

In this example with min number of points of 5, there are 3 clusters chosen