Point Patterns

SERGIO REY

Geographic Information Analysis
School of Geographical Sciences and Urban Planning
Arizona State University

Outline

Outline

Point Pattern Analysis Objectives

Goals

- Pattern detection
- Assessing the presence of *clustering*
- Identification of individual *clusters*

General Approaches

- Estimate intensity of the process
- Formulating an idealized model and investigating deviations from expectations
- Formulating a stochastic model and fitting it to the data

Point Pattern Analysis Objectives

Goals

- Pattern detection
- Assessing the presence of *clustering*
- Identification of individual *clusters*

General Approaches

- Estimate intensity of the process
- Formulating an idealized model and investigating deviations from expectations
- Formulating a stochastic model and fitting it to the data

Point Pattern Analysis Objectives

Goals

- Pattern detection
- Assessing the presence of *clustering*
- Identification of individual *clusters*

General Approaches

- Estimate intensity of the process
- Formulating an idealized model and investigating deviations from expectations
- Formulating a stochastic model and fitting it to the data

Point Pattern Analysis Objectives

Goals

- Pattern detection
- Assessing the presence of *clustering*
- Identification of individual *clusters*

General Approaches

- Estimate intensity of the process
- Formulating an idealized model and investigating deviations from expectations
- Formulating a stochastic model and fitting it to the data

Point Pattern Analysis Objectives

Goals

- Pattern detection
- Assessing the presence of *clustering*
- Identification of individual *clusters*

General Approaches

- Estimate intensity of the process
- Formulating an idealized model and investigating deviations from expectations
- Formulating a stochastic model and fitting it to the data

Point Pattern Analysis Objectives

Goals

- Pattern detection
- Assessing the presence of *clustering*
- Identification of individual *clusters*

General Approaches

- Estimate intensity of the process
- Formulating an idealized model and investigating deviations from expectations
- Formulating a stochastic model and fitting it to the data

Point Pattern Analysis Objectives

Goals

- Pattern detection
- Assessing the presence of *clustering*
- Identification of individual *clusters*

General Approaches

- Estimate intensity of the process
- Formulating an idealized model and investigating deviations from expectations
- Formulating a stochastic model and fitting it to the data

Point Pattern Analysis Objectives

Goals

- Pattern detection
- Assessing the presence of *clustering*
- Identification of individual *clusters*

General Approaches

- Estimate intensity of the process
- Formulating an idealized model and investigating deviations from expectations
- Formulating a stochastic model and fitting it to the data

Outline

Point Pattern Analysis Definitions

Spatial Point Pattern

A set of events, irregularly distributed within a region A and presumed to have been generated by some form of stochastic mechanism.

Representation

$\{Y(A), A \subset \Re\}$, where $Y(A)$ is the number of events occurring in area A.

Events, points, locations

Event an occurrence of interest

Point any location in study area

Event location a particular point where an event occurs

Point Pattern Analysis Definitions

Spatial Point Pattern

A set of events, irregularly distributed within a region A and presumed to have been generated by some form of stochastic mechanism.

Representation

$\{Y(A), A \subset \Re\}$, where $Y(A)$ is the number of events occurring in area A.

Events, points, locations

Event an occurrence of interest

Point any location in study area

Event location a particular point where an event occurs

Point Pattern Analysis Definitions

Spatial Point Pattern

A set of events, irregularly distributed within a region A and presumed to have been generated by some form of stochastic mechanism.

Representation

$\{Y(A), A \subset \Re\}$, where $Y(A)$ is the number of events occurring in area A.

Events, points, locations

Event an occurrence of interest

Point any location in study area

Event location a particular point where an event occurs

Point Pattern Analysis Definitions

Spatial Point Pattern

A set of events, irregularly distributed within a region A and presumed to have been generated by some form of stochastic mechanism.

Representation

$\{Y(A), A \subset \Re\}$, where $Y(A)$ is the number of events occurring in area A.

Events, points, locations

Event an occurrence of interest

Point any location in study area

Event location a particular point where an event occurs

Point Pattern Analysis Definitions

Spatial Point Pattern

A set of events, irregularly distributed within a region A and presumed to have been generated by some form of stochastic mechanism.

Representation

$\{Y(A), A \subset \Re\}$, where $Y(A)$ is the number of events occurring in area A.

Events, points, locations

Event an occurrence of interest

Point any location in study area

Event location a particular point where an event occurs

Point Pattern Analysis Definitions

Spatial Point Pattern

A set of events, irregularly distributed within a region A and presumed to have been generated by some form of stochastic mechanism.

Representation

$\{Y(A), A \subset \Re\}$, where $Y(A)$ is the number of events occurring in area A.

Events, points, locations

Event an occurrence of interest

Point any location in study area

Event location a particular point where an event occurs

Point Pattern Analysis Definitions

Region: A

- Most often planar (two-dimensional Euclidean space)
- One dimensional applications also possible
- Three-dimensional increasingly popular (space + time)
- Point processes on networks (non-planar)

Point Pattern Analysis Definitions

Region: A

- Most often planar (two-dimensional Euclidean space)
- One dimensional applications also possible
- Three-dimensional increasingly popular (space + time)
- Point processes on networks (non-planar)

Point Pattern Analysis Definitions

Region: A

- Most often planar (two-dimensional Euclidean space)
- One dimensional applications also possible
- Three-dimensional increasingly popular (space + time)
- Point processes on networks (non-planar)

Point Pattern Analysis Definitions

Region: A

- Most often planar (two-dimensional Euclidean space)
- One dimensional applications also possible
- Three-dimensional increasingly popular (space + time)
- Point processes on networks (non-planar)

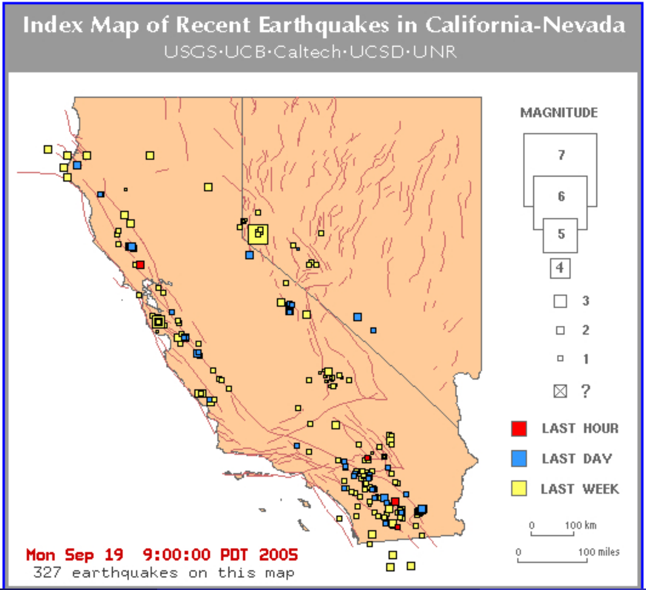
Point Pattern Analysis Definitions

Region: A

- Most often planar (two-dimensional Euclidean space)
- One dimensional applications also possible
- Three-dimensional increasingly popular (space + time)
- Point processes on networks (non-planar)

Space-Time Point Patterns

Space-Time Point Patterns



Point Patterns on Networks

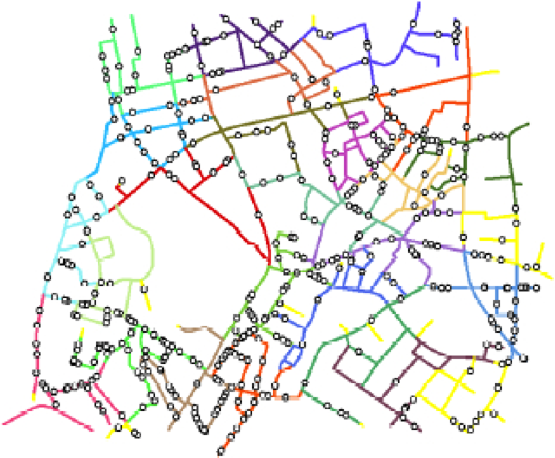


Figure 2: Retail stores assigned to the street network in Shibuya, Tokyo (cells are indicated by different colors)

Point Patterns

Unmarked Point Patterns

- Only location is recorded
- Attribute is binary (presence, absence)

Marked Point Patterns

- Location is recorded
- Non-binary stochastic attribute
- e.g., sales at a retail store, dbh of tree

Unmarked Point Patterns

- Only location is recorded
- Attribute is binary (presence, absence)

Marked Point Patterns

- Location is recorded
- Non-binary stochastic attribute
- e.g., sales at a retail store, dbh of tree

Point Patterns

Unmarked Point Patterns

- Only location is recorded
- Attribute is binary (presence, absence)

Marked Point Patterns

- Location is recorded
- Non-binary stochastic attribute
- e.g., sales at a retail store, dbh of tree

Point Patterns

Unmarked Point Patterns

- Only location is recorded
- Attribute is binary (presence, absence)

Marked Point Patterns

- Location is recorded
- Non-binary stochastic attribute
- e.g., sales at a retail store, dbh of tree

Point Patterns

Unmarked Point Patterns

- Only location is recorded
- Attribute is binary (presence, absence)

Marked Point Patterns

- Location is recorded
- Non-binary stochastic attribute
- e.g., sales at a retail store, dbh of tree

Point Patterns

Unmarked Point Patterns

- Only location is recorded
- Attribute is binary (presence, absence)

Marked Point Patterns

- Location is recorded
- Non-binary stochastic attribute
- e.g., sales at a retail store, dbh of tree

Point Patterns

Unmarked Point Patterns

- Only location is recorded
- Attribute is binary (presence, absence)

Marked Point Patterns

- Location is recorded
- Non-binary stochastic attribute
- e.g., sales at a retail store, dbh of tree

Realizations

Mapped Point Patterns

- *All* events are recorded and mapped
- Complete enumeration of events
- Full information on the realization from the process

Sampled Point Patterns

- *Sample* of events are recorded and mapped
- Complete enumeration of events impossible or intractable
- Partial information on the realization from the process
- Presence/“absence” data (ecology, forestry)

Realizations

Mapped Point Patterns

- *All* events are recorded and mapped
- Complete enumeration of events
- Full information on the realization from the process

Sampled Point Patterns

- *Sample* of events are recorded and mapped
- Complete enumeration of events impossible or intractable
- Partial information on the realization from the process
- Presence/“absence” data (ecology, forestry)

Realizations

Mapped Point Patterns

- *All* events are recorded and mapped
- Complete enumeration of events
- Full information on the realization from the process

Sampled Point Patterns

- *Sample* of events are recorded and mapped
- Complete enumeration of events impossible or intractable
- Partial information on the realization from the process
- Presence/“absence” data (ecology, forestry)

Realizations

Mapped Point Patterns

- *All* events are recorded and mapped
- Complete enumeration of events
- Full information on the realization from the process

Sampled Point Patterns

- *Sample* of events are recorded and mapped
- Complete enumeration of events impossible or intractable
- Partial information on the realization from the process
- Presence/"absence" data (ecology, forestry)

Realizations

Mapped Point Patterns

- *All* events are recorded and mapped
- Complete enumeration of events
- Full information on the realization from the process

Sampled Point Patterns

- *Sample* of events are recorded and mapped
- Complete enumeration of events impossible or intractable
- Partial information on the realization from the process
- Presence/“absence” data (ecology, forestry)

Realizations

Mapped Point Patterns

- *All* events are recorded and mapped
- Complete enumeration of events
- Full information on the realization from the process

Sampled Point Patterns

- *Sample* of events are recorded and mapped
- Complete enumeration of events impossible or intractable
- Partial information on the realization from the process
- Presence/“absence” data (ecology, forestry)

Realizations

Mapped Point Patterns

- *All* events are recorded and mapped
- Complete enumeration of events
- Full information on the realization from the process

Sampled Point Patterns

- *Sample* of events are recorded and mapped
- Complete enumeration of events impossible or intractable
- Partial information on the realization from the process
- Presence/“absence” data (ecology, forestry)

Realizations

Mapped Point Patterns

- *All* events are recorded and mapped
- Complete enumeration of events
- Full information on the realization from the process

Sampled Point Patterns

- *Sample* of events are recorded and mapped
- Complete enumeration of events impossible or intractable
- Partial information on the realization from the process
- Presence/"absence" data (ecology, forestry)

Realizations

Mapped Point Patterns

- *All* events are recorded and mapped
- Complete enumeration of events
- Full information on the realization from the process

Sampled Point Patterns

- *Sample* of events are recorded and mapped
- Complete enumeration of events impossible or intractable
- Partial information on the realization from the process
- Presence/“absence” data (ecology, forestry)

Outline

Research Questions

Location Only

are points randomly located or patterned

Location and Value

- marked point pattern
- is combination of location and value random or patterned

Both Cases

What is the Underlying Process?

Points on a Plane

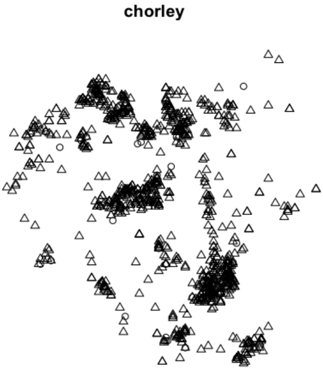
Classic Point Pattern Analysis

- points on an isotropic plane
- no effect of translation and rotation
- classic examples: tree seedlings, rocks, etc

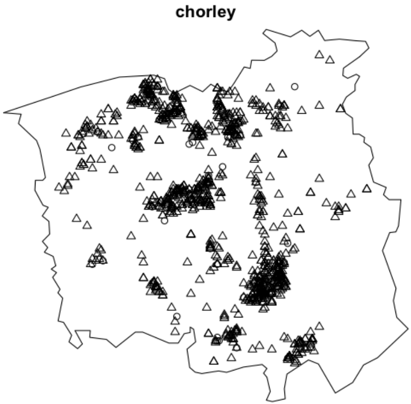
Distance

- no directional effects
- no translational effects
- straight line distance only

Events: Point Map



Points in Context



Intensity

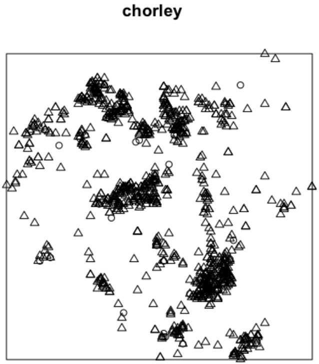
First Moment

- number of points N, area of study $|A|$
- intensity: $\lambda = N/|A|$
- area depends on bounds, often arbitrary

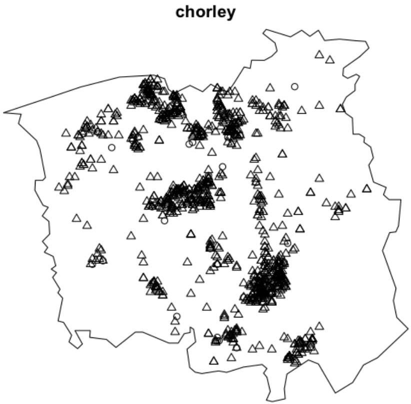
Artificial Boundaries

- bounding box (rectangle, square)
- other (city boundary)

Bounding Box



District Boundary



Convex Hull

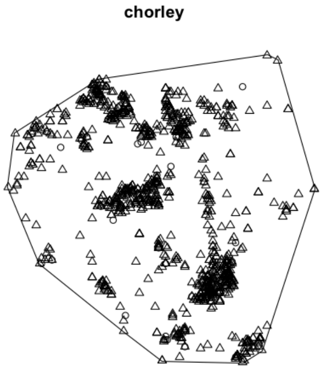
Tightest fit

various algorithms

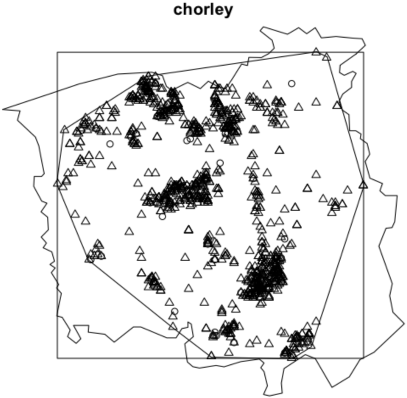
Rescaled Convex Hull (Ripley-Rasson)

- adjust to properly reflect spatial domain of point process
- use centroid of convex hull
- rescale by $1/[\sqrt{(1-m/N)}]$
- m: number of vertexes of convex hull

Convex Hull



Multiple Boundaries



Intensity Estimates

	Area km^2	Intensity $cases/km^2$
District Boundary	315.155	3.29
Bounding Box	310.951	3.33
Convex Hull	229.421	4.52

N=1036

Points on a Network

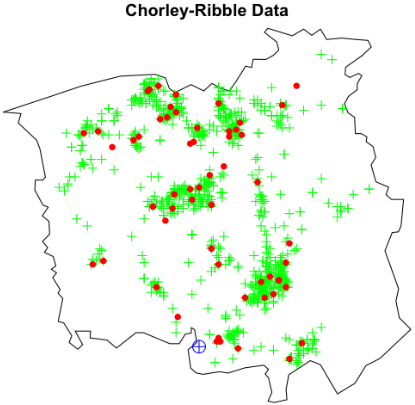
Realistic Location

- addresses
- remove impossible locations (lakes)

Network Distance

- shortest path along network
- Manhattan block distance
- distance vs. travel time or cost

Case-Control Design: Lancashire Cancer

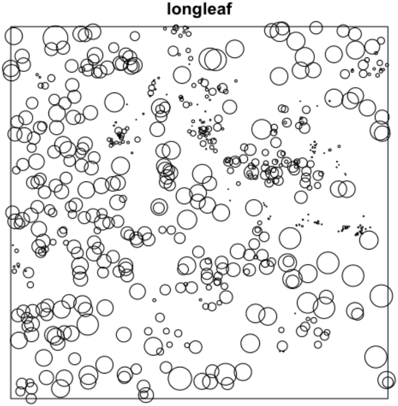


Marked Point Patterns

Both Location and Value

- Patterns in the Location
- Value Associated with Location

Marked Point Pattern: Longleaf Pine

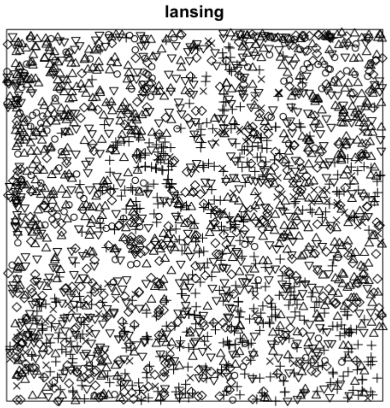


Multi-Type Patterns

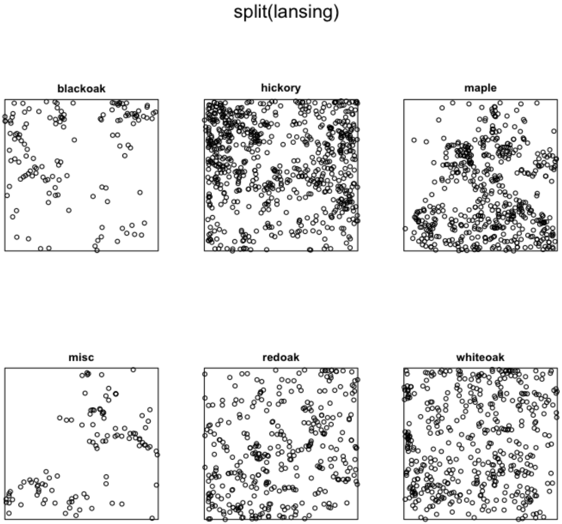
Multiple Categories

- Patterns in Single Category
- Association between Patterns in Other Categories
- Repulsion or Attraction

Multi-Type Pattern: Lansing Woods



Multi-Type Pattern: Lansing Woods



Areal Aggregation

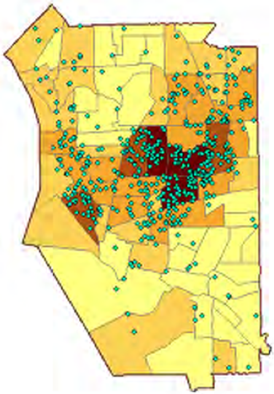
Event Counts

- points aggregated by areal unit
- spatially extensive variable

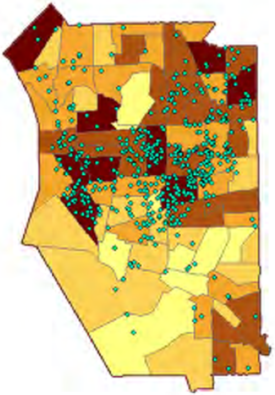
Rates

- events / population at risk
- non-homogeneous population at risk
- risk = probability of an event
- rate is an estimate of underlying risk

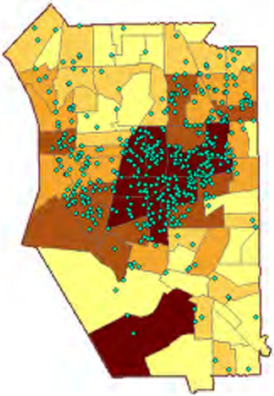
Homicide Counts by Census Tracts



Population Count by Census Tracts



Homicide Rate by Census Tracts



Maps as Outcomes of Processes

Processes **?**

Patterns **?**

Figure 4.1 Our current view of spatial statistical analysis. In this chapter and the next, we will be fleshing out this rather thin description.

Maps as Outcomes of Processes

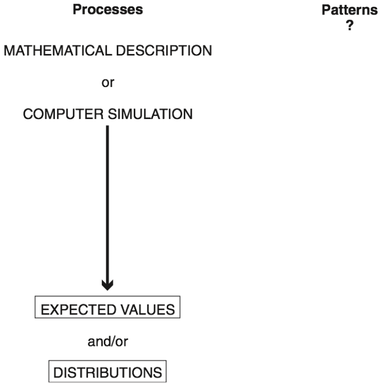


Figure 4.9 The developing framework for spatial statistical analysis. We now have a clearer picture of the meaning of a spatial process. Patterns will be tackled in the next chapter.

Outline

First Order Properties: Spatial Analysis

Mean value of the process in space

- Variation in mean value of the process in space
- Global, large scale spatial trend

First Order Property of Point Patterns, Intensity: λ

- Intensity: λ = number of events expected per unit area
- Estimation of λ
- Spatial variation of λ, $\lambda(s)$, s is a location

$$\lambda(s) = \lim_{ds \to 0} \left\{ \frac{E(Y(ds))}{ds} \right\} \tag{1}$$

Outline

Spatial Correlation Structure

- Deviations in values from process mean
- Local or small scale effects

Second Order Property of Point Patterns

- Relationship between number of events in pairs of areas
- Second order intensity $\gamma(s_i, s_j)$

$$\gamma(s_i, s_j) = \lim_{ds_i \to 0, ds_j \to 0} \left\{ \frac{E(Y(ds_i)Y(ds_j))}{ds_i ds_j} \right\} \tag{2}$$

Spatial Stationarity

First Order Stationarity

$$\lambda(s) = \lambda \forall s \in A \tag{3}$$

$$E(Y(A)) = \lambda \times A \tag{4}$$

Second Order Stationarity

$$\gamma(s_i, s_j) = \gamma(s_i - s_j) = \gamma(h) \tag{5}$$

- h is the vector difference between locations s_i and s_j
- h encompasses direction and distance (relative location)
- Second order intensity only depends on h for second order stationarity

Spatial Isotropy and Stationarity

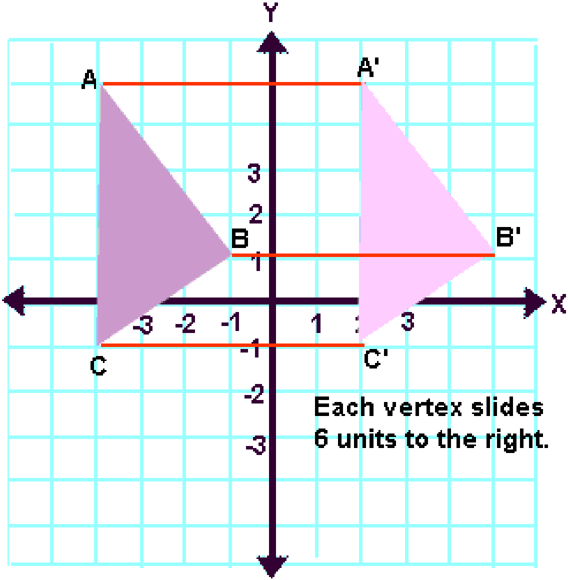
Isotropic Process

- When a stationary process is invariant to rotation about the origin.
- Relationship between two events depend only on the distance separating their locations and not on their orientation to each other.
- Depends only on distance, not direction

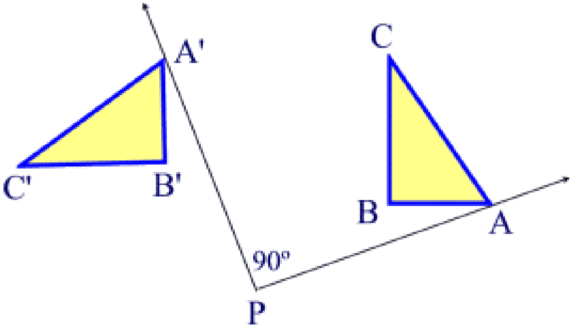
Usefulness

- Two pairs of events from a stationary process separated by same distance and relative direction should have same “relatedness”
- Two pairs of events from a stationary *and* isotropic process separated by the same distance (irrespective of direction) should have the same “relatedness”
- Both allow for replication and the ability to carry out estimation of the underlying DGP.

Invariant Under Translation



Invariant Under Rotation



Outline

Complete Spatial Randomness

CSR

- Standard of Reference
- Uniform: each location has equal probability
- Independent: location of points independent
- Homogeneous Planar Poisson Point Process

Poisson Point Process

Intensity

- number of points in region A : $N(A)$
- intensity: $\lambda = N(A)/|A|$
- implies: $\lambda|A|$ points randomly scattered in a region with area $|A|$
- e.g., 10×1 (points per km^2)

Poisson Distribution

$N(A) \sim Poi(\lambda|A|)$

Poisson Distribution

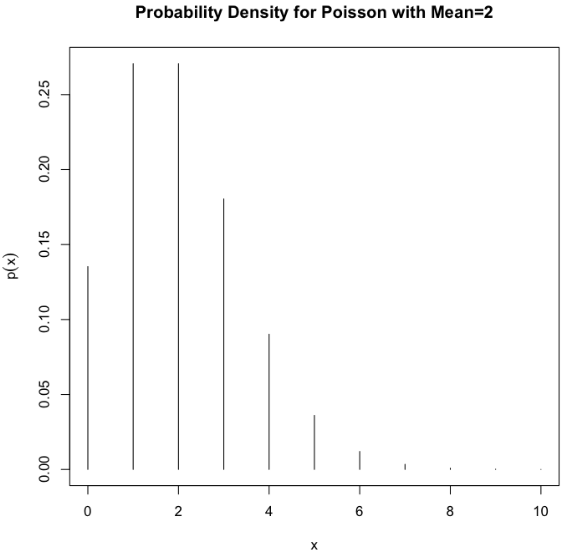
Single Parameter Distribution: $\lambda|A|$

- Generally, λ is the number of events in some well defined *interval*
  - Time: phone calls to operator in one hour
  - Time: accidents at an intersection per week
  - Space: trees in a quadrat
- Let x be a Poisson random variable
  - $E[x] = V[x] = \lambda|A|$

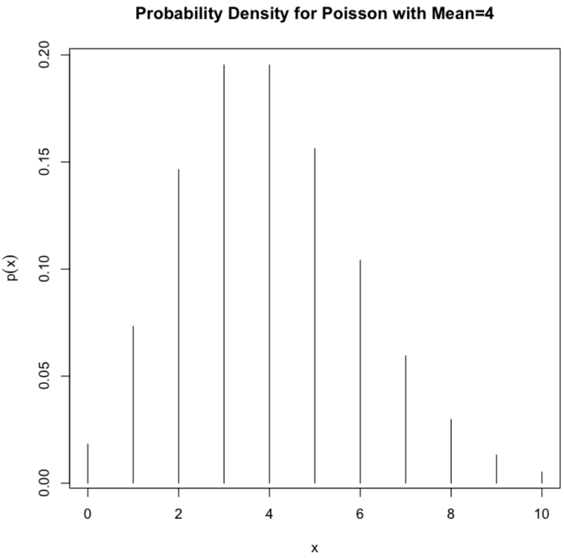
Poisson Distribution

$$P(x) = \frac{e^{-\lambda|A|}(\lambda|A|)^x}{x!} \tag{6}$$

Poisson Distribution $\lambda = 2$



Poisson Distribution $\lambda = 4$



In Space

Single Parameter

$$P[N(A) = x] = e^{-\lambda|A|}(\lambda|A|)^x / x! \tag{7}$$

Spatial Example

CSR with $\lambda = 5/km^2$

- Region = Circle
  - area = $|A| = \pi r^2$
  - $r = 0.1$ km then area ≈ 0.03 km^2
- Probability of Zero Points in Circle

$$
\begin{aligned}
P[N(A) = 0] &= e^{-\lambda|A|}(\lambda|A|)^x/x! && (8)\\
&\approx e^{-5\times 0.03}(5 \times 0.03)^0/0! && (9)\\
&\approx e^{-5\times 0.03} && (10)\\
&\approx 0.86 && (11)
\end{aligned}
$$

Complete Spatial Randomness (CSR)

Homogeneous spatial Poisson point process

1. The number of events occurring within a finite region A is a random variable following a Poisson distribution with mean $\lambda|A|$, with $|A|$ denoting area of A.
2. Given the total number of events N occurring within an area A, the locations of the N events represent an independent random sample of N locations where each location is equally likely to be chosen as an event.

- Criterion 2 is the general concept of CSR (uniform (random)) distribution in A.
- Criterion 1 pertains to the intensity λ.

Homogeneous Poisson process

Implications

1. The number of events in nonoverlapping regions in A are statistically independent.
2. For any region $R \subset A$:

$$\lim_{|R| \to 0} \frac{Pr[exactly\ one\ event\ in\ R]}{|R|} = \lambda > 0 \qquad (12)$$

3. $$\lim_{|R| \to 0} \frac{Pr[more\ than\ one\ event\ in\ R]}{|R|} = 0 \qquad (13)$$

Homogeneous Poisson process

Implications

- λ is the intensity of the spatial point pattern.
- For a Poisson random variable, Y:

$$E[Y] = \lambda = V[Y] \tag{14}$$

- Provides the motivation for some quadrat tests of CSR hypothesis.
  - If Y_R is the count in quadrat R
  - If $\widehat{E[Y]} < \widehat{V[Y]}$: overdispersion = spatial clustering
  - If $\widehat{E[Y]} > \widehat{V[Y]}$: underdispersion = spatial uniformity

Simulating CSR

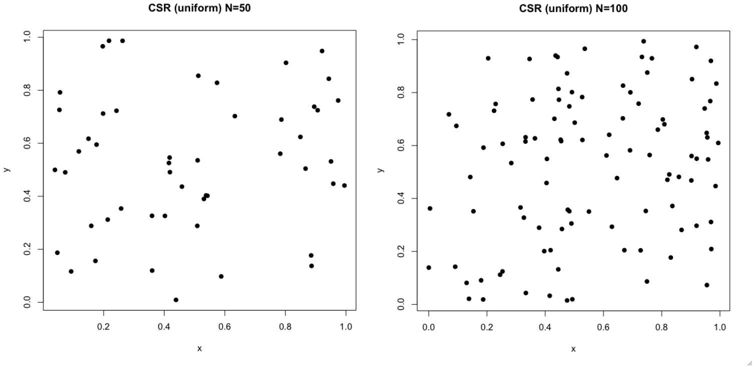
$N - conditioned$

- CSR= uniform distribution
- random uniform draws for x and y point coordinates
- N fixed

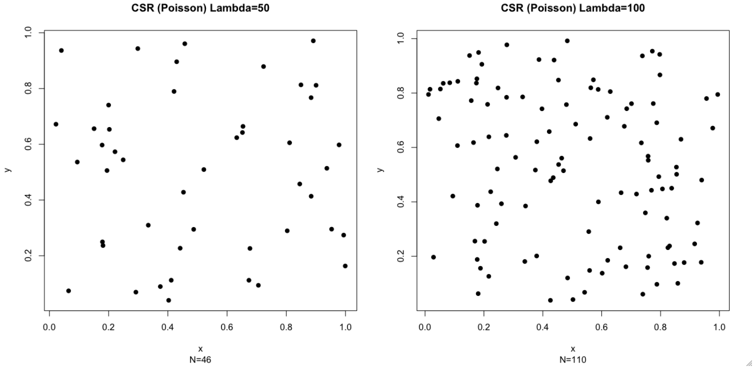
$\lambda - conditioned$

- CSR= Poisson distribution
- λ and $|A|$ given
- $N(A)$ random

CSR Uniform



CSR Poisson



CSR (Poisson) Lambda=50
CSR (Poisson) Lambda=100
y
x
N=46
N=110
0.0
0.2
0.4
0.6
0.8
1.0

Limitations of CSR

Stationary Poisson Process

- homogeneous
- translation invariant

Rare in practice

very few (any?) actual processes are CSR

Strawman

- purely a benchmark
- null hypothesis

Outline

Clustered Pattern

More Grouped than CSR

- some higher densities, aggregated
- many points at shorter distances

Overdispersion

- variance > mean
- greater variation in densities than CSR

Sources of Clustering

Contagion

- presence of events at x affects probability of event at y
- correlated point processes

Heterogeneity

- intensity $\lambda(s)$ varies with s
- heterogeneous Poisson point process

Contagious Distributions

Two stages

- point pattern for parents
- point pattern for offspring centered on parent locations
- parents may or may not be included

Examples

- Poisson cluster process (Neyman-Scott)
- Matern cluster process

Poisson Cluster Process

Parent Events

- Poisson process with intensity λ

Number of Offspring Events S

- identical distribution for each parent
- $E[S] = \mu$

Location of Offspring Events

- independent and identically distributed
- following a bivariate density h

Example

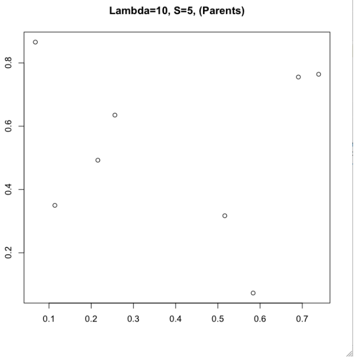
Parent Process Poisson

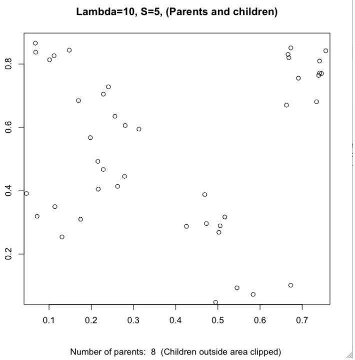
- homogeneous, intensity λ constant

Child Process

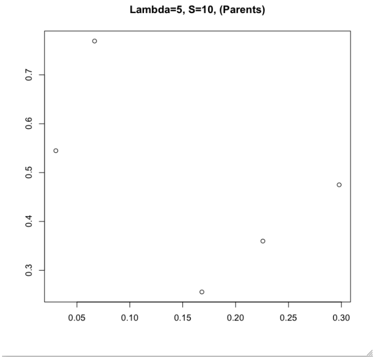
- uniform points in circle centered on parent
- fixed number of points in circle centered on parent
- points outside window eliminated

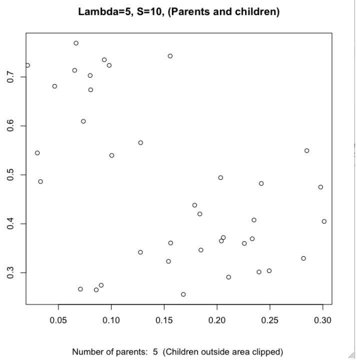
Neyman Scott $\lambda = 10$, $S = 5$





Neyman Scott $\lambda = 5, S = 10$





Implications

- Apparent clusters can occur solely due to heterogeneities in the intensity function $\lambda(s)$.
- Individual event locations still remain independent of one another.
- Process is not stationary due to intensity heterogeneity

HPP vs. IPP

HPP is a special case of IPP with a constant intensity

CSR vs. Constant Risk Hypotheses

CSR

- Intensity is spatially constant
- Population at risk assumed spatially uniform
- Useful null hypothesis if these conditions are met

Constant Risk Hypothesis

- Population density variable
- Individual risk constant
- Expected number of events should vary with population density
- Clusters due to deviation from CSR
- Clusters due to deviation from CSR and Constant Risk

Inhomogeneous Poisson Process

Non-Stationary

- spatially varying intensity $\lambda(s)$
- mean is $\int_A \lambda(s)ds$
- an integral of the location-specific intensities over the region

Properties

- $N(A) \sim Poi(\int_A \lambda(s)ds)$
- $N(A) = n,\ n$ events independent sample with pdf proportional to $\lambda(s)$

Sources of Variability

Deterministic

- function for variability of $\lambda(s)$
- introduce covariates: $\lambda(s) = f(z)$

Stochastic

- doubly stochastic process
- distribution for $\lambda(s) \sim \Lambda(s)$

Examples

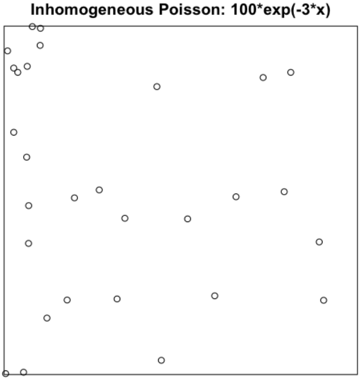
Intensity Varies with a Covariate

- trend surface
- $\lambda(s) = exp(\alpha + \beta s)$

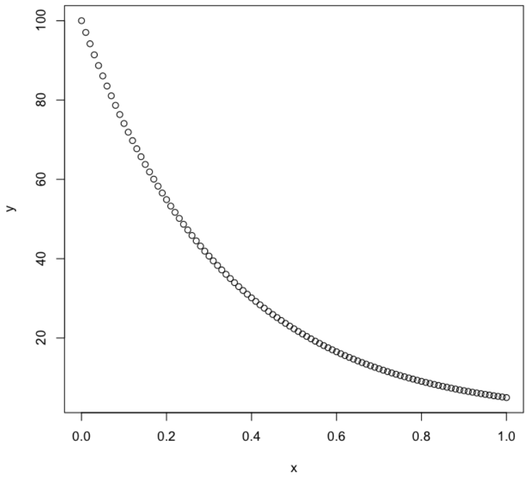
Intensity Varies with Distance to Focus

- $\lambda(s) = \lambda 0(s).f(||s - s_0||, \theta)$

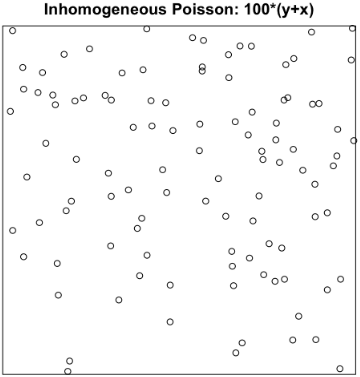
Inhomogeneous Poisson Process: $\lambda(x, y) = 100 * exp(-3x)$



Inhomogeneous Poisson Process: $\lambda(x, y) = 100 * exp(-3x)$



Inhomogeneous Poisson Process: $\lambda(x, y) = 100 * (x + y)$



Thinning

From Homogeneous to Heterogeneous

- remove points

Types

- p-thinning: constant probability
- $p(s)$-thinning: probability varies with s
- Π-thinning: thinning function is random

Simulation

Start with homogeneous Poisson

- $\lambda = max[\alpha(s)]$

Apply $p(s)$ Thinning

- keep points with probability $p(s)$
- $p(s) = \alpha(s)/\lambda$
- e.g., keep if generated uniform random number < $p(s)$

Cox Process

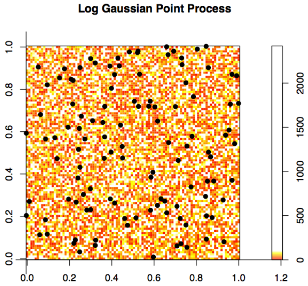
Doubly Stochastic Process

- $\Lambda(s)$ is stochastic process over A
- events inhomogeneous Poisson process with $\lambda(s) = \Lambda(s)$ (a realization)

Log-Gaussian Process

- $\Lambda(s) = exp[Z(s)]$ with $Z(s) \sim N(\mu, \sigma^w)$
- $E[\lambda] = exp(\mu + 0.5\sigma^2)$

Cox process



$\ln Z \sim N(4.1, 1)$ $E[\lambda] \approx 100$ $\lambda = 113$

Identification

Inverse problem

- identify process from pattern

True Contagion - Apparent Contagion

- impossible to distinguish contagious process from heterogeneous process

Identification

Bartlett Equivalence

- Cox process (heterogeneity) and Poisson Cluster process (contagion) yield equivalent patterns

Identification Strategies

- repeated observation, covariates
- heterogeneous in same location, contagious not

Outline

Regular Pattern

Less Grouped than CSR

- fewer high densities, empty space
- dispersed
- repulsion, competition

Underdispersion

- variance < mean
- less variation in densities than CSR

Inhibition Process

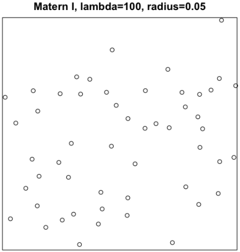
Minimum Permissible Distance

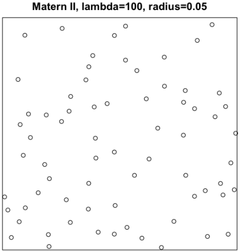
- no two points closer than δ
- packing intensity $\tau = \lambda\pi\delta^2/4$

Matern Process

- I: thinned Poisson process using δ
- II: sequential inhibition process, generates points conditional on previous points and distance (denser than I)

Matern I and II $\lambda = 100$





Matern I and II $\lambda = 500$