Introduction to Point Processes

Reminder: Point Process Research Questions

- Is the spatial pattern of points random?
- Is there clustering in the events?
- Do events repel each other?

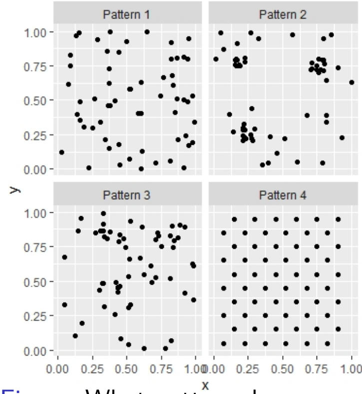


Figure: What pattern do you see?

Steps to Analyzing a Point Pattern

1. Descriptive statistics, such as the average intensity over a region
2. Tests for Complete Spatial Randomness (similar to Moran's I for areal data)
3. Model for the intensity λ (similar to spatial regression for areal data)

Point Processes

- It is difficult to assess the relationship in point patterns from visualizations, though they can be helpful (just like with areal data)

Point Processes

- It is difficult to assess the relationship in point patterns from visualizations, though they can be helpful (just like with areal data)
- We model the **intensity**, λ, of a spatial point pattern to characterize the pattern over a region W

Point Processes

- It is difficult to assess the relationship in point patterns from visualizations, though they can be helpful (just like with areal data)
- We model the **intensity**, λ, of a spatial point pattern to characterize the pattern over a region W
  - How many points do we expect in a given region? (λ)

Point Processes

- It is difficult to assess the relationship in point patterns from visualizations, though they can be helpful (just like with areal data)
- We model the **intensity**, λ, of a spatial point pattern to characterize the pattern over a region W
  - How many points do we expect in a given region? (λ)
  - How many points do we expect to be near a specific location? ($\lambda(s)$)

Point Processes

- It is difficult to assess the relationship in point patterns from visualizations, though they can be helpful (just like with areal data)
- We model the **intensity**, λ, of a spatial point pattern to characterize the pattern over a region W
  - How many points do we expect in a given region? (λ)
  - How many points do we expect to be near a specific location? ($\lambda(s)$)
- The intensity is not known and must be estimated

Descriptive Statistics

Intensity Estimate- Point Processes

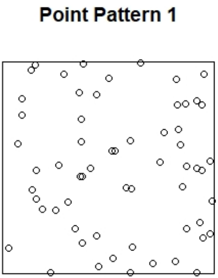
Basic estimator of the intensity, λ:

$$\hat{\lambda} = \frac{n}{a}$$

where n is the number of points $(s_1, ...s_n)$ and a is the area of the window W.

Descriptive Statistics

$$\hat{\lambda} = \frac{n}{a}$$



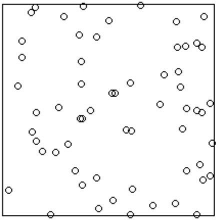
R output:

- Planar point pattern: 60 points
- Window: rectangle = [0, 1] x [0, 1] units
- Window area = 1 square unit
- **Average intensity ($\hat{\lambda}$):**

Descriptive Statistics

$$\hat{\lambda} = \frac{n}{a}$$

Point Pattern 1



R output:

- Planar point pattern: 60 points
- Window: rectangle = [0, 1] x [0, 1] units
- Window area = 1 square unit
- **Average intensity ($\hat{\lambda}$): 60 points per square unit**

Descriptive Statistics

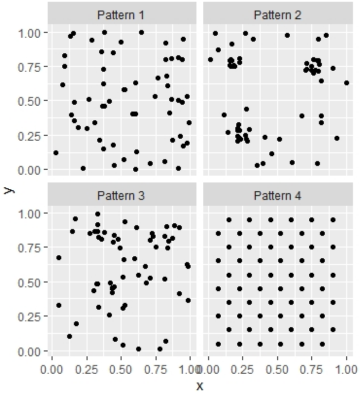


Figure: n = 60 and $W = [0, 1] \times [0, 1]$ for all point patterns

What is the average intensity for each pattern?

Descriptive Statistics

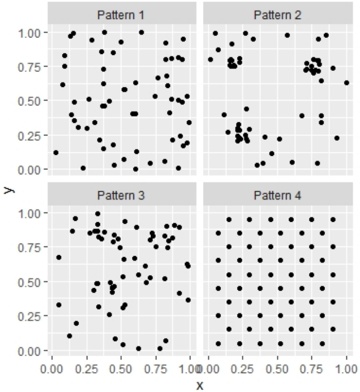


Figure: n = 60 and $W = [0, 1] \times [0, 1]$ for all point patterns

What is the average intensity for each pattern?

$$\hat{\lambda} = \frac{n}{a}$$

Descriptive Statistics- Crime Data

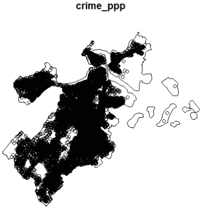


Figure: Crime Data Point Pattern

R output:

- Planar point pattern: 2511520 points
- **Average intensity: 0.009470109 points per square unit**
- Window: polygonal boundary
- Window area = 265205000 square units

Complete Spatial Randomness

Randomness

Is the point process random?

- **Deterministic**: A process is deterministic if the process contains no elements of randomness, can be predicted with certainty
- **Stochastic**: A process is stochastic if the process contains elements of randomness (non-technical definition)

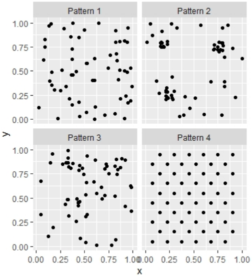


Figure: Which pattern is stochastic? Which is deterministic?

Spatial Randomness

- **Complete Spatial Randomness (CSR):** the process is distributed independently at random and uniformly over the window W

Spatial Randomness

- **Complete Spatial Randomness (CSR):** the process is distributed independently at random and uniformly over the window W
  - No region where events are more or less likely to happen

Spatial Randomness

- **Complete Spatial Randomness (CSR):** the process is distributed independently at random and uniformly over the window W
  - No region where events are more or less likely to happen
  - The presence of a nearby event doesn't mean a given location is more or less likely to have an event

Spatial Randomness

- **Complete Spatial Randomness (CSR):** the process is distributed independently at random and uniformly over the window W
  - No region where events are more or less likely to happen
  - The presence of a nearby event doesn't mean a given location is more or less likely to have an event
- **Clustering:** there is attraction between points

Spatial Randomness

- **Complete Spatial Randomness (CSR):** the process is distributed independently at random and uniformly over the window W
    - No region where events are more or less likely to happen
    - The presence of a nearby event doesn't mean a given location is more or less likely to have an event
- **Clustering:** there is attraction between points
- **Inhibition:** there is competition among points (leads to regular patterns)

Spatial Randomness

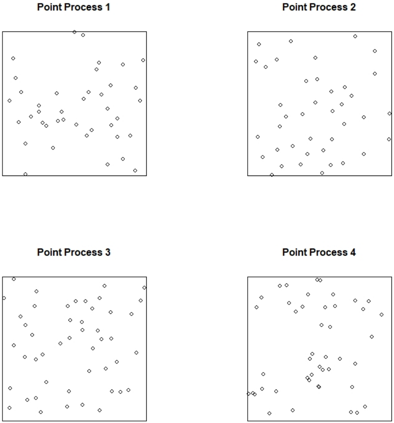


Figure: Which pattern exhibits CSR? Inhibition? Clustering?

Spatial Randomness

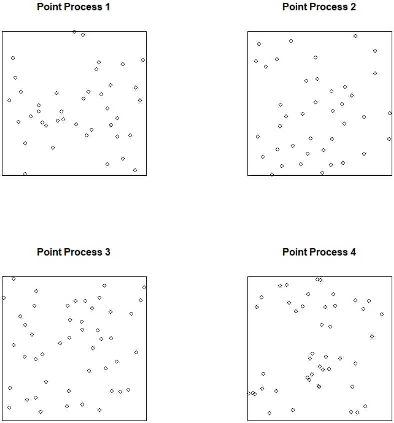


Figure: Which pattern exhibits CSR? Inhibition? Clustering?

1: CSR, 2 & 3: Inhibition, 4: Clustering
It is really difficult to tell from images!!

Nearest Neighbors

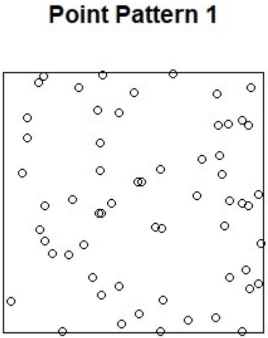
Testing for CSR

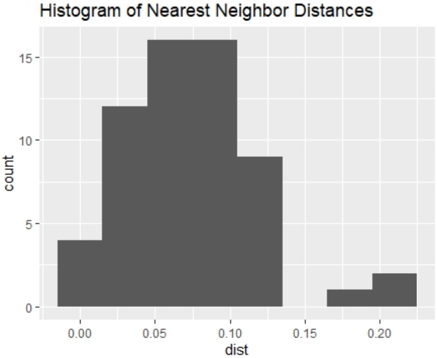
Nearest Neighbors

- Nearest neighbor of a location is the event that is closest (smallest distance) to this location
- Formally, nearest neighbor of event i is event j with shortest distance d from i

$$d_{ij} \leq d_{ik} \forall k$$

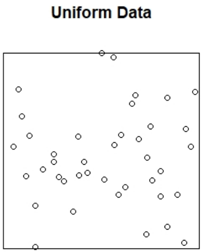
Nearest Neighbor Example

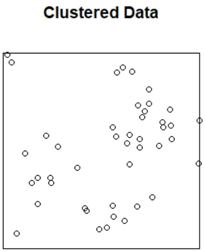




Nearest Neighbor Example

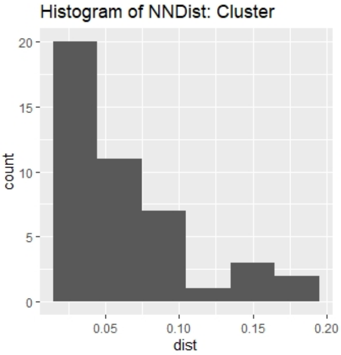
- What would you expect the distribution of nearest neighbor distances to look like for clustering vs CSR?





Nearest Neighbor Example

- Smaller distances for clustering vs uniform data



G Function

Testing for CSR

G Function:

- Summary of the distribution of the distances from an arbitrary event to its nearest event

Testing for CSR

G Function:

- Summary of the distribution of the distances from an arbitrary event to its nearest event
- Estimated from the data

Testing for CSR

G Function:

- Summary of the distribution of the distances from an arbitrary event to its nearest event
- Estimated from the data
- **Cumulative distribution function:** proportion of events that have a nearest neighbor at a distance less than r

Testing for CSR

G Function:

- Summary of the distribution of the distances from an arbitrary event to its nearest event
- Estimated from the data
- **Cumulative distribution function:** proportion of events that have a nearest neighbor at a distance less than r
- When distance r is very small, proportion of events with nearest neighbor at distance less than r is:

Testing for CSR

G Function:

- Summary of the distribution of the distances from an arbitrary event to its nearest event
- Estimated from the data
- **Cumulative distribution function:** proportion of events that have a nearest neighbor at a distance less than r
- When distance r is very small, proportion of events with nearest neighbor at distance less than r is: **small**.

Testing for CSR

G Function:

- Summary of the distribution of the distances from an arbitrary event to its nearest event
- Estimated from the data
- **Cumulative distribution function:** proportion of events that have a nearest neighbor at a distance less than r
- When distance r is very small, proportion of events with nearest neighbor at distance less than r is: **small**.
- When distance r is very large, proportion of events with nearest neighbor at distance less than r is:

Testing for CSR

G Function:

- Summary of the distribution of the distances from an arbitrary event to its nearest event
- Estimated from the data
- **Cumulative distribution function:** proportion of events that have a nearest neighbor at a distance less than r
- When distance r is very small, proportion of events with nearest neighbor at distance less than r is: **small**.
- When distance r is very large, proportion of events with nearest neighbor at distance less than r is: **large**.

G function

$$\hat{G}(r) = \frac{\#\{d_i : d_i \leq r, \forall i\}}{n}$$

- $d_i = min_j\{d_{ij}, \forall j \neq i\}$ (**distance to nearest event**)
- numerator is the number of elements in the set of n.n. distances that are lower than or equal to r
- n is number of points

G Function

Compare the data to the **G function under CSR**:

$$G(r) = 1 - \exp(-\lambda \pi r^2)$$

where λ is the mean number of events per unit area.

Procedure:

G Function

Compare the data to the **G function under CSR**:

$$G(r) = 1 - \exp(-\lambda \pi r^2)$$

where λ is the mean number of events per unit area.

Procedure:

1. Calculate $G(r)$ under CSR

G Function

Compare the data to the **G function under CSR**:

$$G(r) = 1 - \exp(-\lambda \pi r^2)$$

where λ is the mean number of events per unit area.

Procedure:

1. Calculate $G(r)$ under CSR
2. Calculate $\hat{G}(r)$ for our dataset

G Function

Compare the data to the **G function under CSR**:

$$G(r) = 1 - \exp(-\lambda \pi r^2)$$

where λ is the mean number of events per unit area.

Procedure:

1. Calculate $G(r)$ under CSR
2. Calculate $\hat{G}(r)$ for our dataset
3. Compare $G(r)$ under CSR and $\hat{G}(r)$ for our dataset

G Function

Compare the data to the **G function under CSR**:

$$G(r) = 1 - \exp(-\lambda \pi r^2)$$

where λ is the mean number of events per unit area.

Procedure:

1. Calculate $G(r)$ under CSR
2. Calculate $\hat{G}(r)$ for our dataset
3. Compare $G(r)$ under CSR and $\hat{G}(r)$ for our dataset
4. **More formal test:** Can also repeatedly simulate datasets under CSR for your window and with the same intensity $\hat{\lambda}$ and compare to the distribution of those simulations
   - Similar to permutation tests in areal data

Idea of Assessing CSR with the G function

- We have one realization of what the G function (distribution of nearest neighbor distances) looks like for our data

Idea of Assessing CSR with the G function

- We have one realization of what the G function (distribution of nearest neighbor distances) looks like for our data
- We can simulate many many data sets that are completely spatially random and calculate the G function for each

Idea of Assessing CSR with the G function

- We have one realization of what the G function (distribution of nearest neighbor distances) looks like for our data
- We can simulate many many data sets that are completely spatially random and calculate the G function for each
- Compare these simulated datasets to our dataset

Idea of Assessing CSR with the G function

- We have one realization of what the G function (distribution of nearest neighbor distances) looks like for our data
- We can simulate many many data sets that are completely spatially random and calculate the G function for each
- Compare these simulated datasets to our dataset
- Does the G function for the data look different than for the simulated (spatially random) datasets?

Application of G function

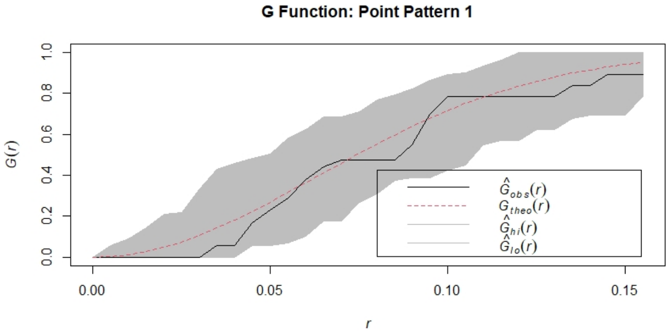


Figure: Is there evidence against CSR?

- Red line: $G(r)$ under Complete Spatial Randomness
- Black line: $\hat{G}(r)$ estimated from the data
- Grey shaded area: bounds from the "simulation envelope" from many many CSR simulations

Application of G function

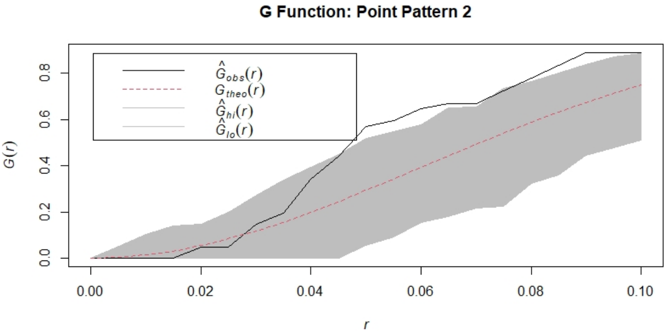


Figure: Is there evidence against CSR?

Application of G function

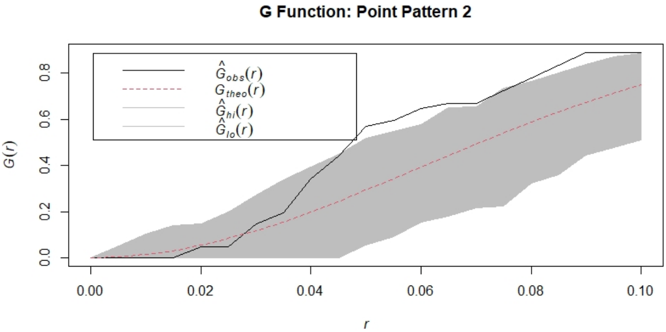


Figure: Is there evidence against CSR?

Intuition: The points are closer together than under CSR (indicates clustering)

Another Application of the G function

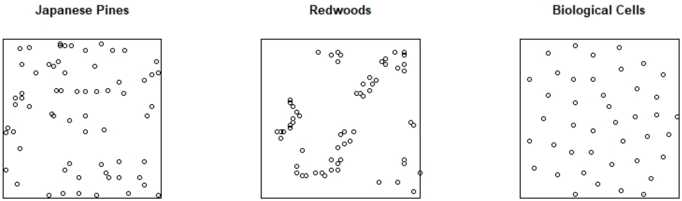


Figure: Which dataset looks like CSR? Inhibition? Clustering?

Another Application of the G function

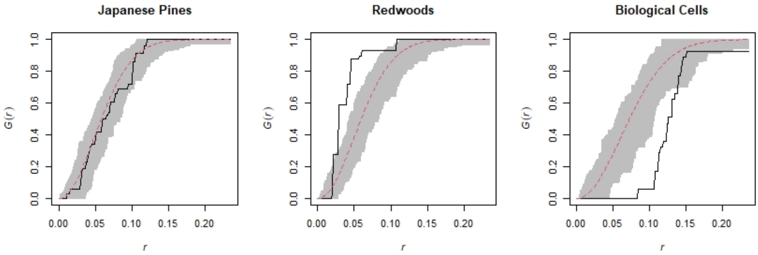


Figure: Which G function gives evidence for CSR? Inhibition? Clustering? (if any)

Another Application of the G function

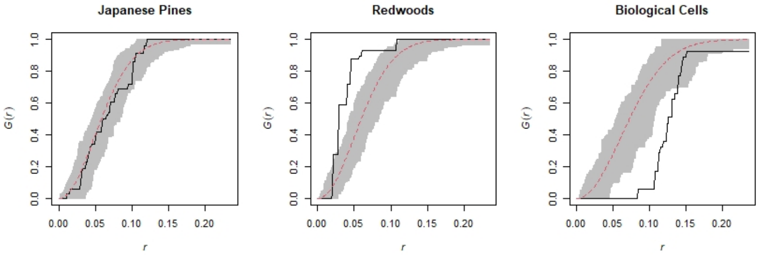


Figure: Which G function gives evidence for CSR? Inhibition? Clustering? (if any)

What is the intuition behind this plot?

Another Application of the G function

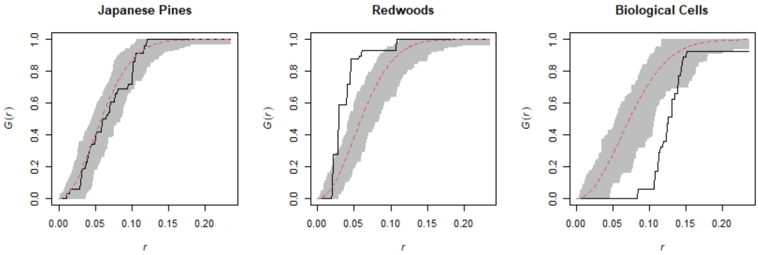


Figure: Which G function gives evidence for CSR? Inhibition? Clustering? (if any)

What is the intuition behind this plot?

With the redwood data (clustered), there are more points than expected under CSR at many radii. **Events are closer than expected.**
With the cell data (inhibition), there are fewer points than expected under CSR at most radii. **Events are further than expected.**

F Function

Testing for CSR

F Function:

Testing for CSR

F Function:

- Distribution of the distances from an arbitrary **point** (not event) to its nearest event

Testing for CSR

F Function:

- Distribution of the distances from an arbitrary **point** (not event) to its nearest event
- Also called the **Empty Space Function**
  - Measure of average space left between events
  - More empty space = longer distances between point and nearest event

Testing for CSR

F Function:

- Distribution of the distances from an arbitrary **point** (not event) to its nearest event
- Also called the **Empty Space Function**
  - Measure of average space left between events
  - More empty space = longer distances between point and nearest event
- Estimated from the data

F Function from the data

$$\hat{F}(r) = \frac{\#\{d_{ik} : d_{ik} \leq r, \forall i\}}{n}$$

- d_{ik} distance from point i to nearest neighbor event k
- Compare the data to the F function under CSR:

$$F(r) = 1 - \exp(-\lambda \pi r^2)$$

Application of the F function

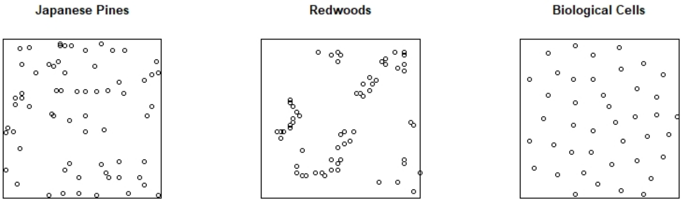


Figure: Reminder of the datasets

Empty Space Illustration

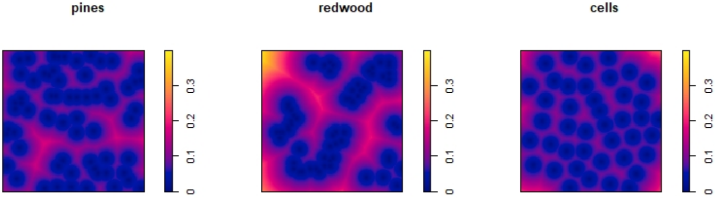


Figure: Comparison of empty space between datasets

- Blue: where points are located
- Pink/yellow: empty space

Application of the F function

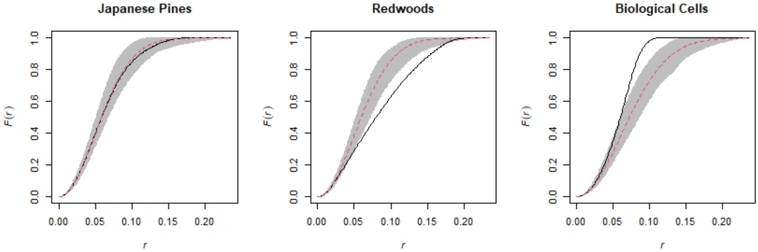


Figure: Which F function gives evidence for CSR? Inhibition? Clustering? (if any)

Application of the F function

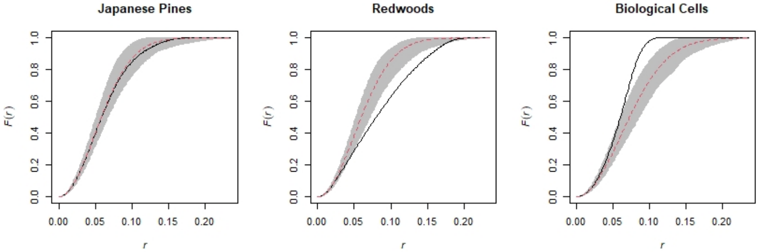


Figure: Which F function gives evidence for CSR? Inhibition? Clustering? (if any)

- Interpretation is opposite of the G function
- **Inhibition:** Empty spaces are closer to events than expected under CSR
- **Clustering:** Empty spaces are further than events than expected under CSR