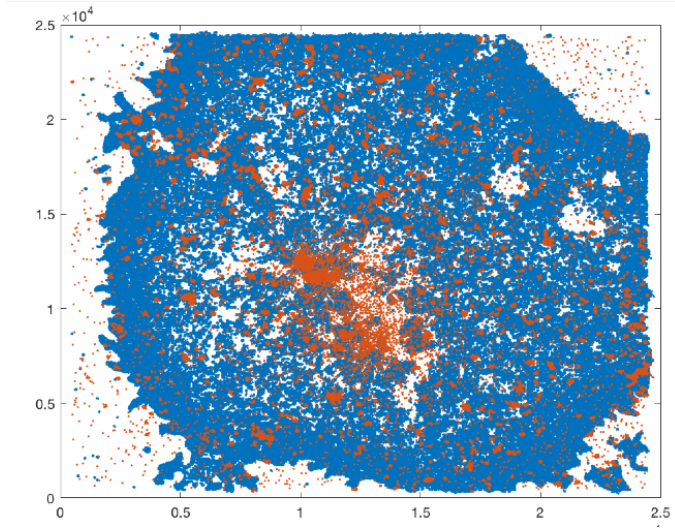
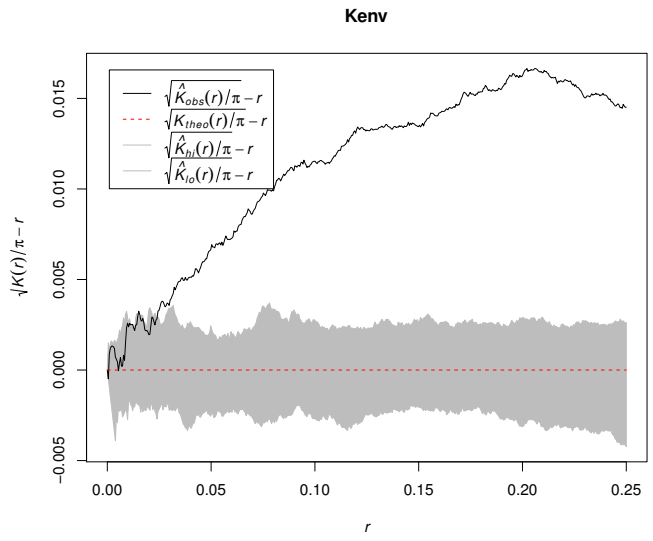
Spatial Statistics for Bioimage Analysis

Ed Cohen





Spatial Statistics for Bioimage Analysis

What's going on?

- 12 lectures.
- 2 problems classes, Tuesday and Thursday.
- Office Hour 2-3pm Wednesday, Huxley 536.
- Exam (Friday).

Who am I?

- Dr Ed Cohen, Statistics, Department of Mathematics
- Contact: e.cohen@imperial.ac.uk
- Research: Statistical Signal and Image Processing

What is this module about?

- Studying simple models for the type of spatial data found in fluorescence microscopy, as well as several other applications
- Developing simple statistical tools for analysing these type of data

Overview: Spatial Statistics for Bioimage Analysis

Things we'll look at this module:

- An intro to bioimaging and microscopy
- Spatial point patterns
- Spatial point processes
- Hypothesis testing for spatial data
- Model fitting
- Multivariate spatial data

The notes are self-contained, however, should you desire extra reading, I recommend

Diggle, P.J., Statistical analysis of spatial and spatio-temporal point patterns, CRC Press.

Cressie, N., Statistics for Spatial Data, Wiley.

Bioimaging

An intro to fluorescence microscopy

See other slides

Spatial data and spatial point patterns

Introduction to spatial data

- Spatial data is classed as any data that represents observations over a spatial domain.
- Typically, this spatial domain will be $\mathbb{R}^2$, but is often $\mathbb{R}^3$, and may even be in more general spaces (beyond the scope of this module).
- Spatial data can be broadly categorised into two types: (i) data sampled over a continuous domain, (ii) event data.

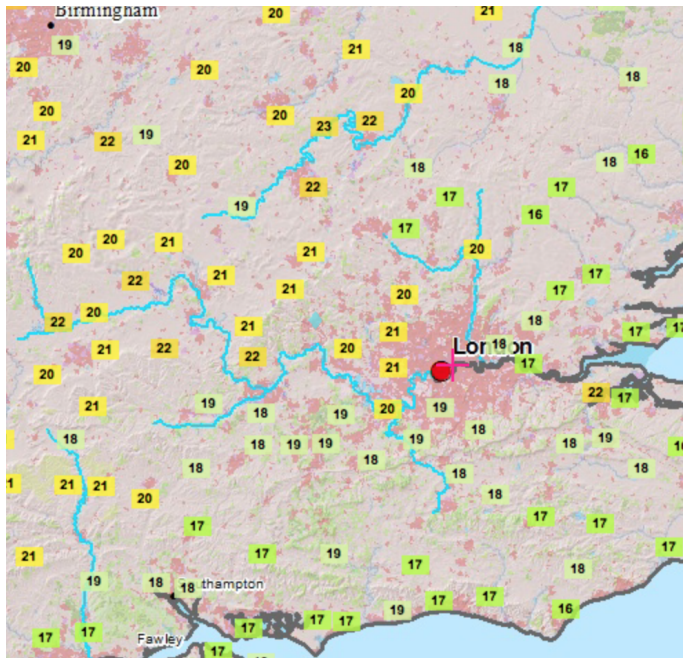
(i) Continuous domain data

Data sampled over a continuous domain are those that observe a continuous process. They could be sampled anywhere in the domain (hence the term *continuous*). Examples include

- Temperature recorded at weather stations, i.e.,

$$\{T(\boldsymbol{s}_1), T(\boldsymbol{s}_2), ...\},$$

where $\boldsymbol{s}_1, \boldsymbol{s}_2, ... \in \mathbb{R}^2$ are the grid coordinates of the weather stations.

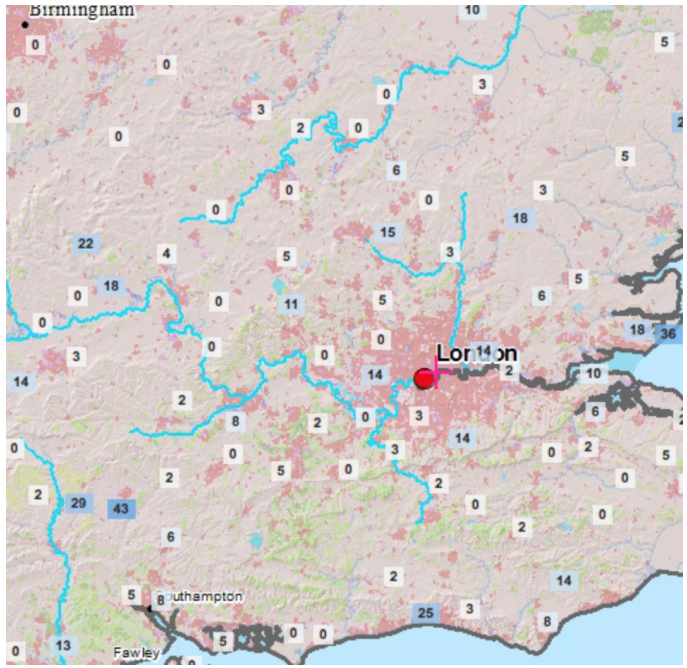


(i) Continuous domain data

- Wind velocity at weather stations, i.e.

$$\{\boldsymbol{V}(\boldsymbol{s}_1), \boldsymbol{V}(\boldsymbol{s}_2), ...\},$$

where $\boldsymbol{s}_1, \boldsymbol{s}_2, ... \in \mathbb{R}^2$ are the grid coordinates of the weather stations.

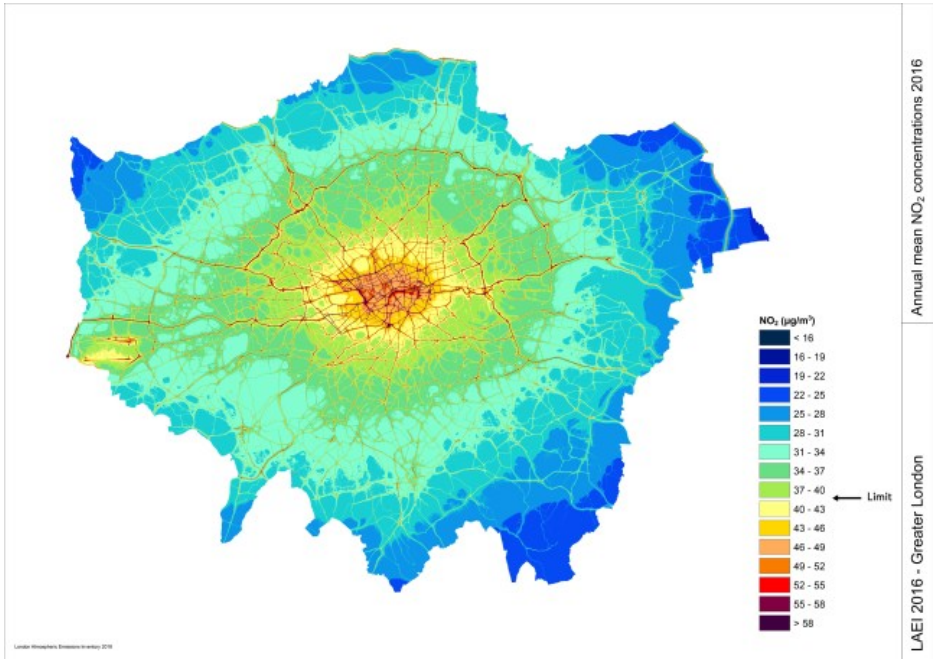


(i) Continuous domain data

- Air pollution recorded at pollution sensors, i.e.

$$\{P(\boldsymbol{s}_1), P(\boldsymbol{s}_2), ...\},$$

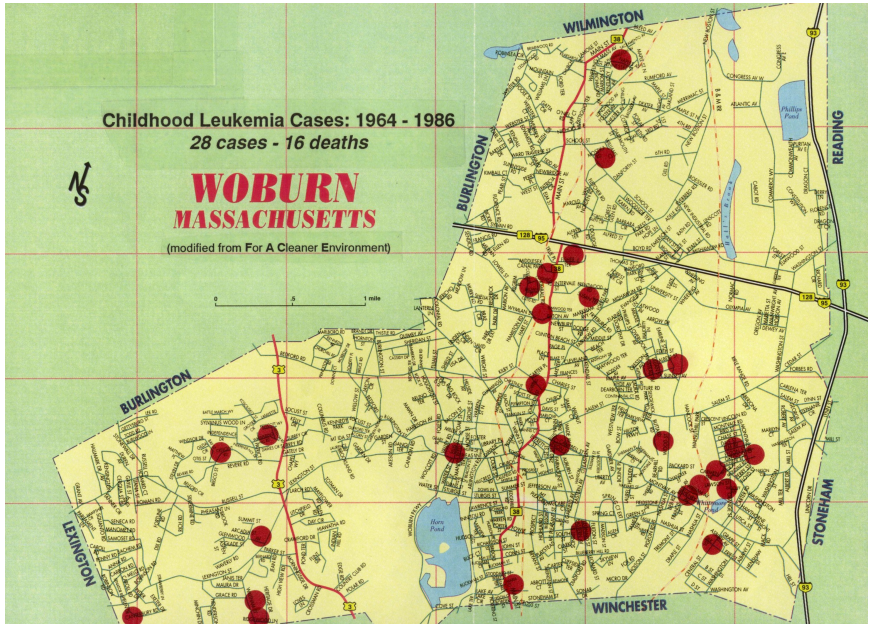
where $\boldsymbol{s}_1, \boldsymbol{s}_2, ... \in \mathbb{R}^2$ are the grid coordinates of the pollution sensors.



(ii) Event data

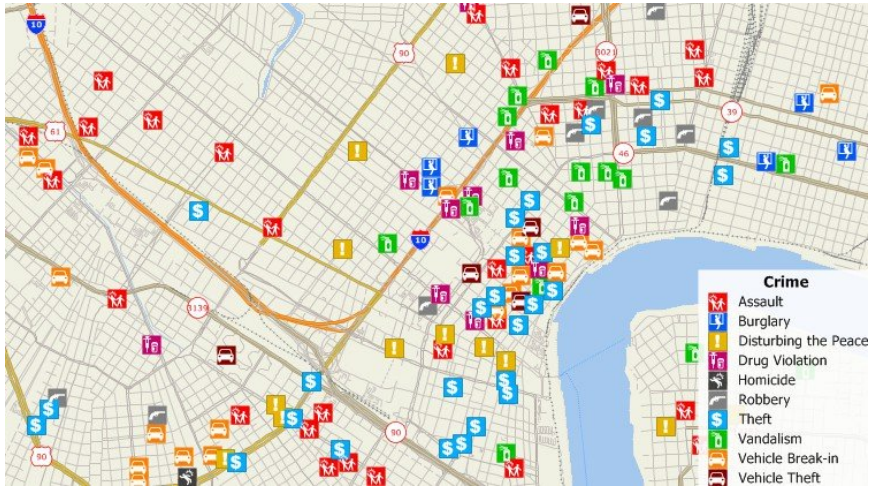
Event data is the location of specific events in $\mathbb{R}^2$ or $\mathbb{R}^3$. They can be represented as a list $\{\boldsymbol{s}_1, \boldsymbol{s}_2, ...\}$. Examples of event data include

- Location of Leukaemia sufferers 1964-1968 in Woburn Massachusetts, US.



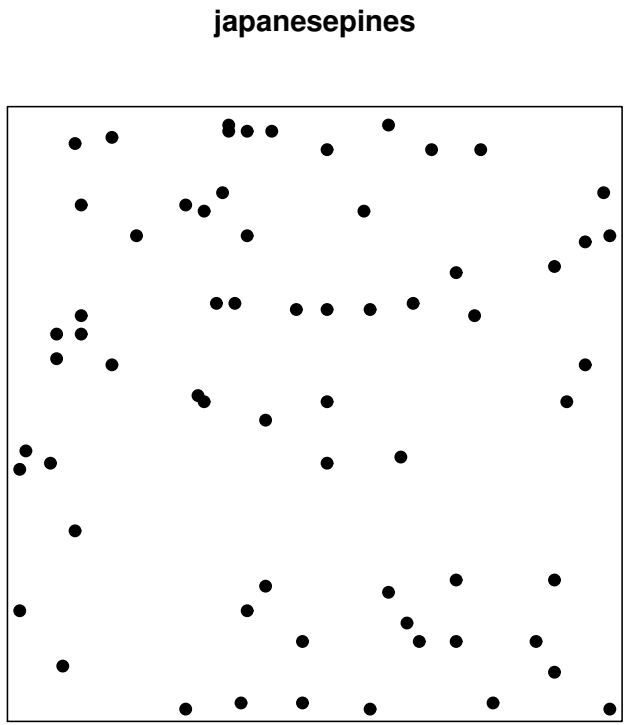
(ii) Event data

- Location of crimes in a city



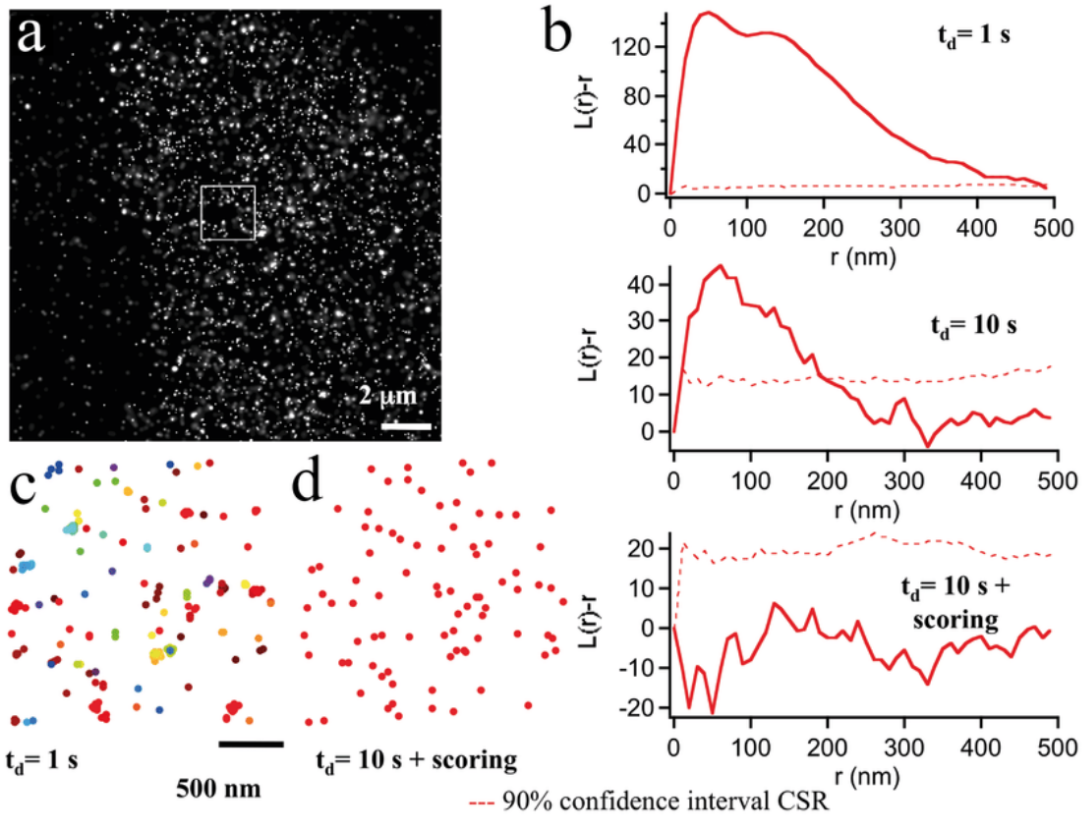
(ii) Event data

- Location of Japanese pine trees in a region of forest.



(ii) Event data

- Location of SrcN15-mEos2 on the plasma membrane of HeLa cells.



From: P Annibale, Investigating the Impact of Single Molecule Fluorescence Dynamics on Photo Activated Localization Microscopy Experiments, 2012

Spatial Point Patterns

- It is event data that will be the focus of this module. When the event locations are random, this is known as a *spatial point pattern*.
- Broadly speaking, "a spatial point pattern is a set of locations, irregularly distributed within a designated region and presumed to have been generated by some form of stochastic mechanism (random process)." [Diggle, p. xxix].
- Any such data-set is termed a spatial point pattern. The locations are referred to as *events*. This is to distinguish them from arbitrary points of the region of space.

Question:

Point patterns arise in many different contexts. Just a few examples have already been given. Can you think of any more?

Spatial Point Patterns

- Here are two spatial point patterns in a square region of interest (ROI).

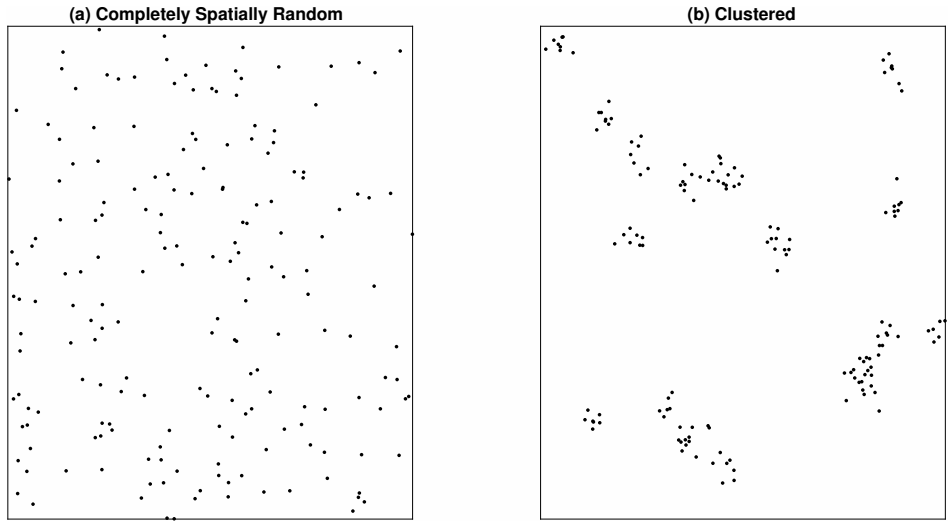


Figure: Example point patterns

- The two patterns appear to be very different. The left pattern appears shows no clear structure and might be regarded as "completely random". The right pattern, on the other hand, shows clear clustering.

Here is a further different type of pattern. This shows the location of cell nuclei.

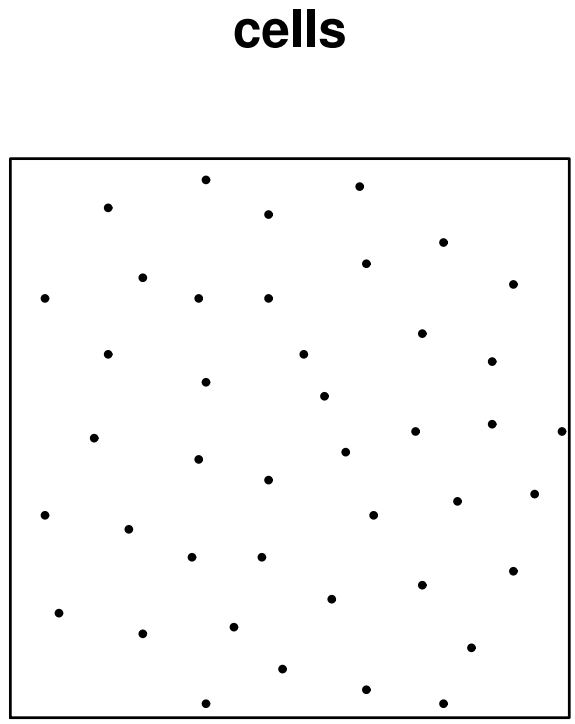


Figure: Position of cell nuclei in a widefield microscopy image. An example of a regular point pattern.

They are distributed more or less regularly over the ROI. This would be highly improbable, unless there is some underlying mechanism which promotes an even distribution.

Question

Can you think of data that might be appear completely random, data which you would expect to be clustered, and data which you might expect to be regular?

Objectives of statistical analysis

Spatial statistics is about analysing spatial data to make inference on some underlying mechanism.

Point processes

Poisson distribution

Before we move forward, it will be necessary for us to define the Poisson distribution.

Definition

A discrete random variable X is said to have a Poisson distribution with parameter $\mu > 0$, $X \sim \text{Poisson}(\mu)$, if for $k = 0, 1, 2, ...$

$$P(X = k) = \frac{\mu^k \exp(-\mu)}{k!}.$$

If $X \sim \text{Poisson}(\mu)$, then $E(X) = \mu$ and $\text{Var}(X) = \mu$.

NOTE: $0! = 1$ by convention.

Poisson distribution

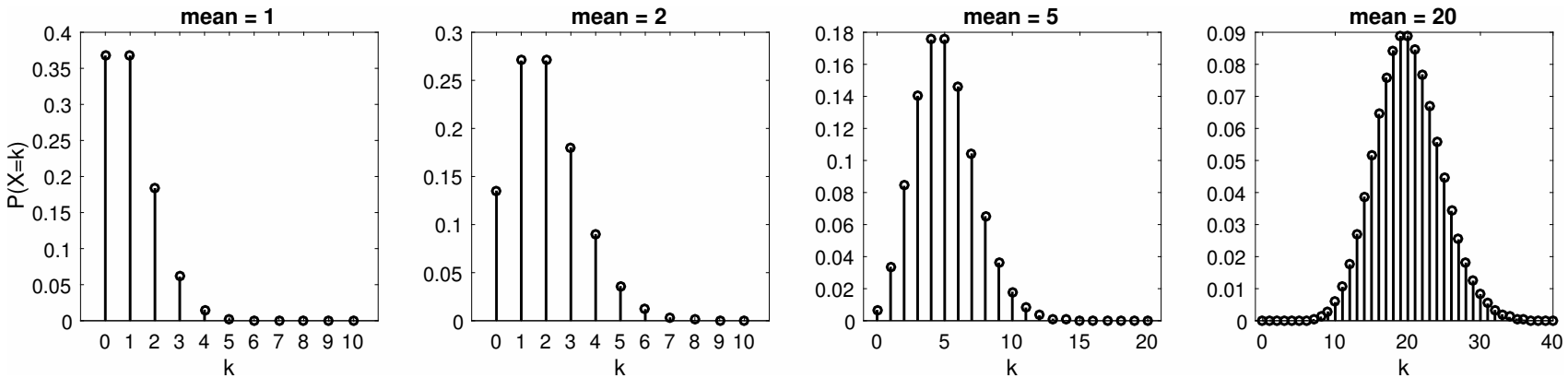


Figure: Probability mass function for the Poisson distributed when $\mu = 1$, $\mu = 2$, $\mu = 5$, $\mu = 20$

Poisson distribution

The Poisson distribution is a hugely important distribution in statistics as it models count data - in particular, it models the count of events that happen independently. For example:

- The number of letters you receive each day.
- The number of emails you receive each day.
- The number of people who join a post office queue between 9:30am and 10:00am.
- The number of phone calls a call centre receives each day.

Question: What other examples can you think of?

Simulating Poisson Data

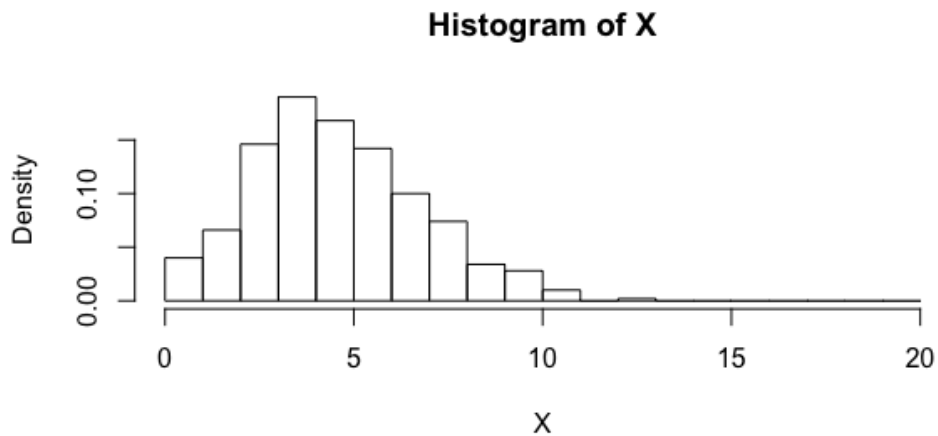
To simulate n Poisson random variables with expected value μ in R we use the function `rpois(n,mu)`.
For example:

```
X = rpois(10,5)
```

which gives me

```
 3 5 7 3 6 5 8 6 6 3
```

```
X = rpois(500,5)
hist(X,freq=F,breaks = seq(0,20,1)
```



Question

The number of emails I get each day is Poisson distributed with expected value 10. What is the probability I get no emails on any given day?

(a) 0.1

(b) 0

(c) $\exp(-10)$

(d) $10\exp(-10)$

(e) $\exp(-240)$

Solution

Question

The number of buses that arrive at my local bus stop between 08:00am and 09:00am is Poisson distributed with expected value 4. What is the probability at least one bus turns up?

(a) 0

(b) 1

(c) $\exp(-4)$

(d) $1 - \exp(-4)$

(e) None of the above

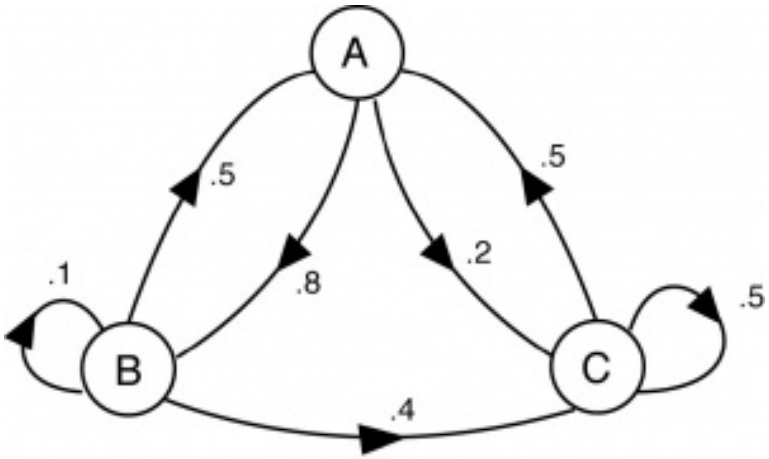
Solution

Stochastic processes

- A stochastic process is a random process that evolves in either time, space, or time and space.
- A key concept with stochastic processes is the difference between the process and the realization.
  - The process is the stochastic mechanism that generates the data.
  - The realization is the observed data for any one run of the stochastic process.
  - Run the same stochastic process again and you get a new realization.
- Often, we will only see one realization of the random process, from which we must try and understand the mechanics of the process itself.

Examples

Stochastic process (Markov chain)



Markov graph of transiton probabilites between states A, B and C

Realisations (starting in A)

$A, B, B, C, A, C, C, A, B, A, \ldots$

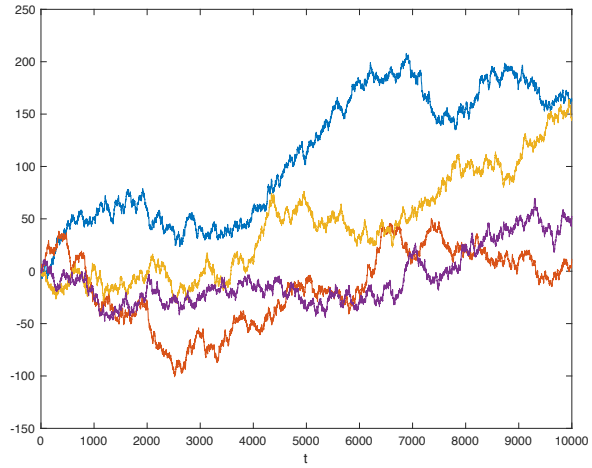
$A, C, C, A, B, C, C, C, A, C, \ldots$

$A, C, A, B, A, B, A, C, A, B, \ldots$

Stochastic process (Random walk)

$$X_t = X_{t-1} + \epsilon_t; \quad \epsilon_t \sim N(0, 1).$$

Realisations (Starting with $X_t = 0$.)



Point processes

- Point processes (also called *event processes*), are a type of stochastic process.
- They are used to model event data, in either the temporal domain, the spatial domain, or both.
- Temporal examples:
  - The times at which you receive an email
  - The times at which children are born in a maternity unit.
  - The times at which a bus arrives at a bus-stop
- Spatial examples:
  - where a particular species of trees grow in a forest
  - where protein molecules position themselves on a cellular membrane
  - where crimes happen in a city
- Spatio-temporal examples:
  - The time and location of earthquake events

STOCHASTIC MECHANISM

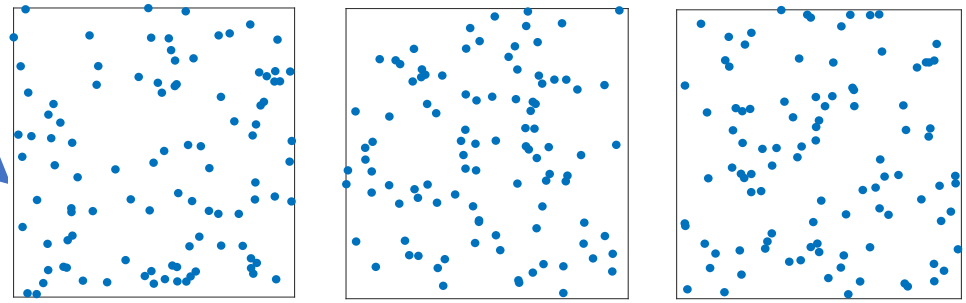
REALIZATIONS

ROLLING TWO DICE

BIVARIATE DISCRETE UNIFORM DISTRIBUTION ON {1,2,3,4,5,6} x {1,2,3,4,5,6}

CLATHARIN COATED PITS ON CELL MEMBRANCE

POISSON PROCESS WITH INTENSITY λ



STOCHASTIC MECHANISM

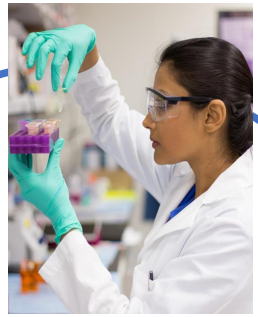
REALIZATIONS

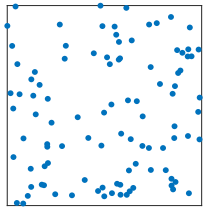
ROLLING TWO DICE

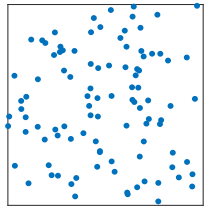
BIVARIATE DISCRETE
UNIFORM DISTRIBUTION ON
{1,2,3,4,5,6} x {1,2,3,4,5,6}

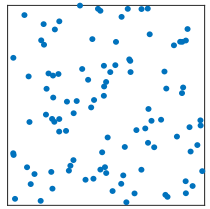
CLATHARIN COATED
PITS ON CELL MEMBRANCE

POISSON PROCESS
WITH INTENSITY λ









Point process in 1D

We will begin our exploration into point processes with temporal point processes (1D point processes).

- We will represent an event process as $N(A)$, where $A \subset \mathbb{R}$. $N(A)$ is a random number describing the number of events that occur in A.
- We will also use the notation $N(t)$ to denote the number of events that have occurred in the interval $(0, t]$. I.e., for an interval $(a, b]$, $N\{(a, b]\} = N(b) - N(a)$.

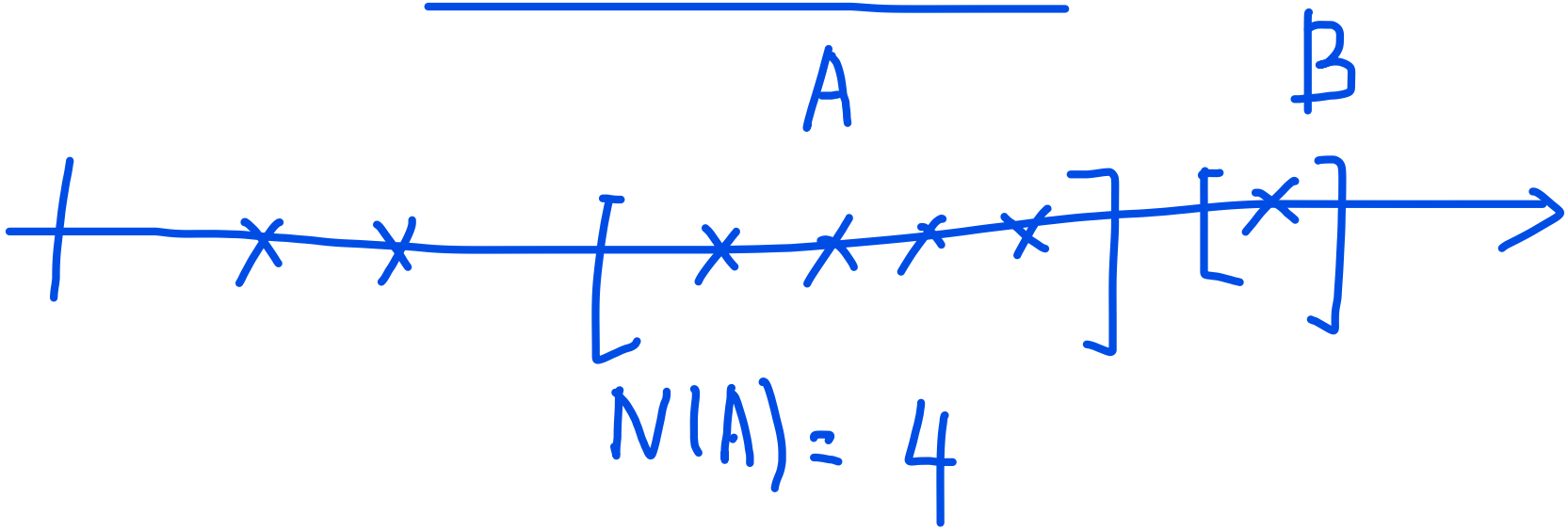


Figure: point process in 1D

Simple processes

We will solely be dealing with *simple* point processes

Definition (Simple)

N is called a simple point process if no two events can occur at exactly the same time,

$$P[N\{(t, t+\Delta t]\} > 1] = o(\Delta t).$$

The prob goes to 0 faster than del t goes to 0 s.t. $\lim_{\Delta t \to 0} P[\ ] / \Delta t = 0$

- We will make use of the incremental process $\mathrm{d}N(t) = N(t+\mathrm{d}t) - N(t)$.
- Essentially this tells us the number of events that occur in the interval $(t, t+\mathrm{d}t]$ and for simple processes is Bernoulli (i.e. takes a value of 0 or 1).

Intensity

Let us consider a key descriptor of a point process.

Definition (Intensity)

The intensity $\lambda(t)$ of a point process is the expected number of events per unit time, i.e.

$$\lambda(t) = \frac{E\{\mathrm{d}N(t)\}}{\mathrm{d}t} \equiv \lim_{\Delta t \to 0} \frac{E[N\{(t+\Delta t]\}]}{\Delta t}.$$

Poisson process in 1D

A Poisson process is the simplest type of event/point process.

Definition (Poisson process)

Event process N is called a Poisson process if the following two conditions hold:

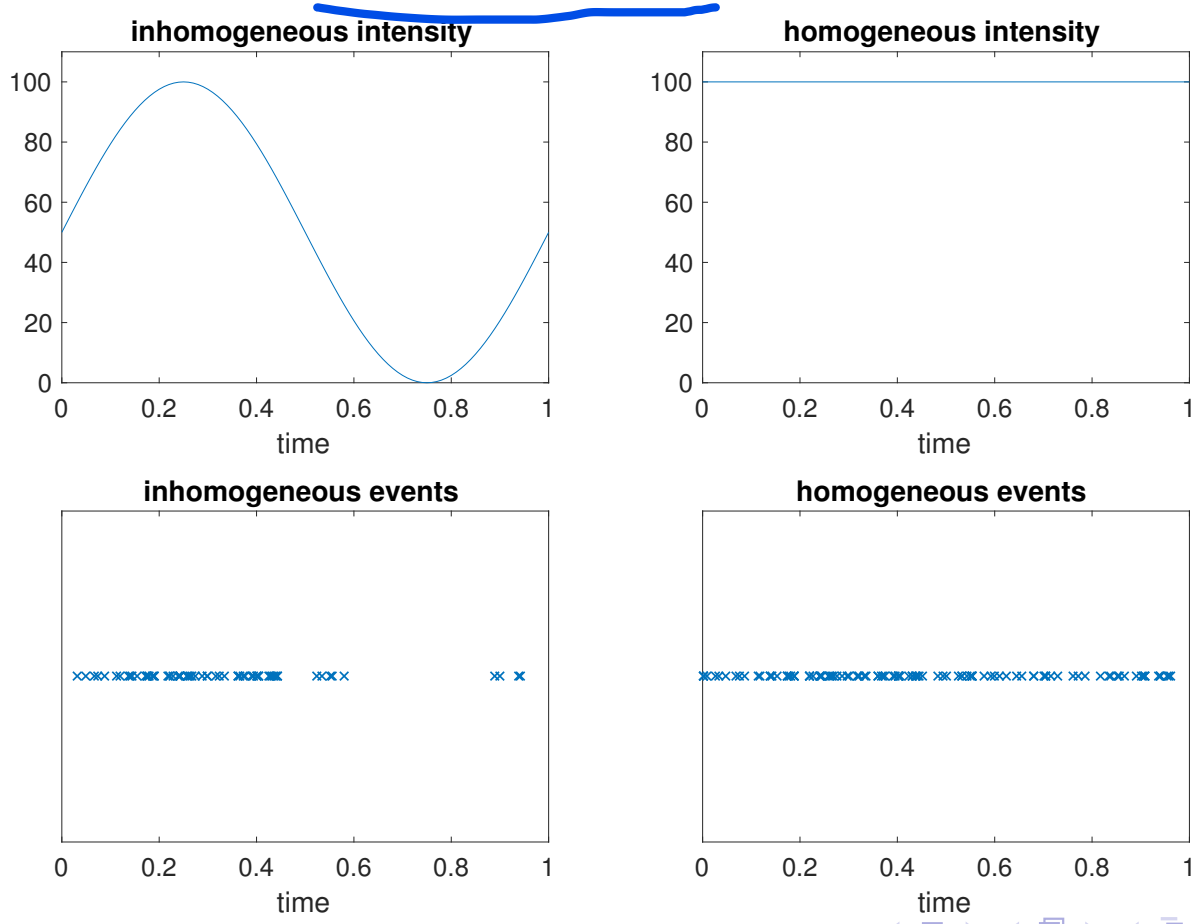
1. For any set $A \subset \mathbb{R}$, $N(A)$ is Poisson distributed with expected value $\mu(A) = \int_A \lambda(t)\mathrm{d}t$. We write $N(A) \sim \mathrm{Poisson}\{\mu(A)\}$.
2. For any disjoint pair of sets A and B, $N(A)$ and $N(B)$ are independent ~~events~~ random variables

Homogeneous process

Definition (Homogeneous)

A Poisson process N is called homogeneous is $\lambda(t) = \lambda$ for all t.
That is to say, it has a constant intensity for all time.

Processes that have a variable intensity are called inhomogeneous.



Properties of Poisson Processes

- A Poisson process has the key *memoryless* property.
- That is to say, the presence or absence of an event at time t has no bearing on whether there will be events anywhere else in the process. All events occur independently of the others.
- **Question:** Can you think of any real-life event processes that can legitimately be modelled as Poisson?

Question

People join a Post Office queue according to a homogeneous Poisson process at a constant intensity of 1 per minute. What is the probability that no people join the queue between 10:00:00am and 10:05:00am?

(a) $\exp(-5)$

(b) 1

(c) $\exp(-1)$

(d) $1 - \exp(-5)$

(e) None of the above

Solution

Question

Mass extinction events happen according to a Poisson process with intensity 1 per 10 million years. What is the probability that at least one will occur in the next 1 million years.

(a) $\exp(-10)$

(b) $1 - \exp(-10)$

(c) $1 - \exp(-0.1)$

(d) $\exp(-0.1)$

(e) None of the above

Solution

Question

The radioactive sample decays according to an inhomogeneous Poisson process at a intensity of $10\exp(-0.1t)$ $(t > 0)$ events per second. What is the $E[N\{(0, 10]\}]$?

(a) $100\{1 - \exp(-1)\}$

(b) $\exp(-1)$

(c) $10\exp(-1)$

(d) 10

(e) $1 - \exp(-1)$

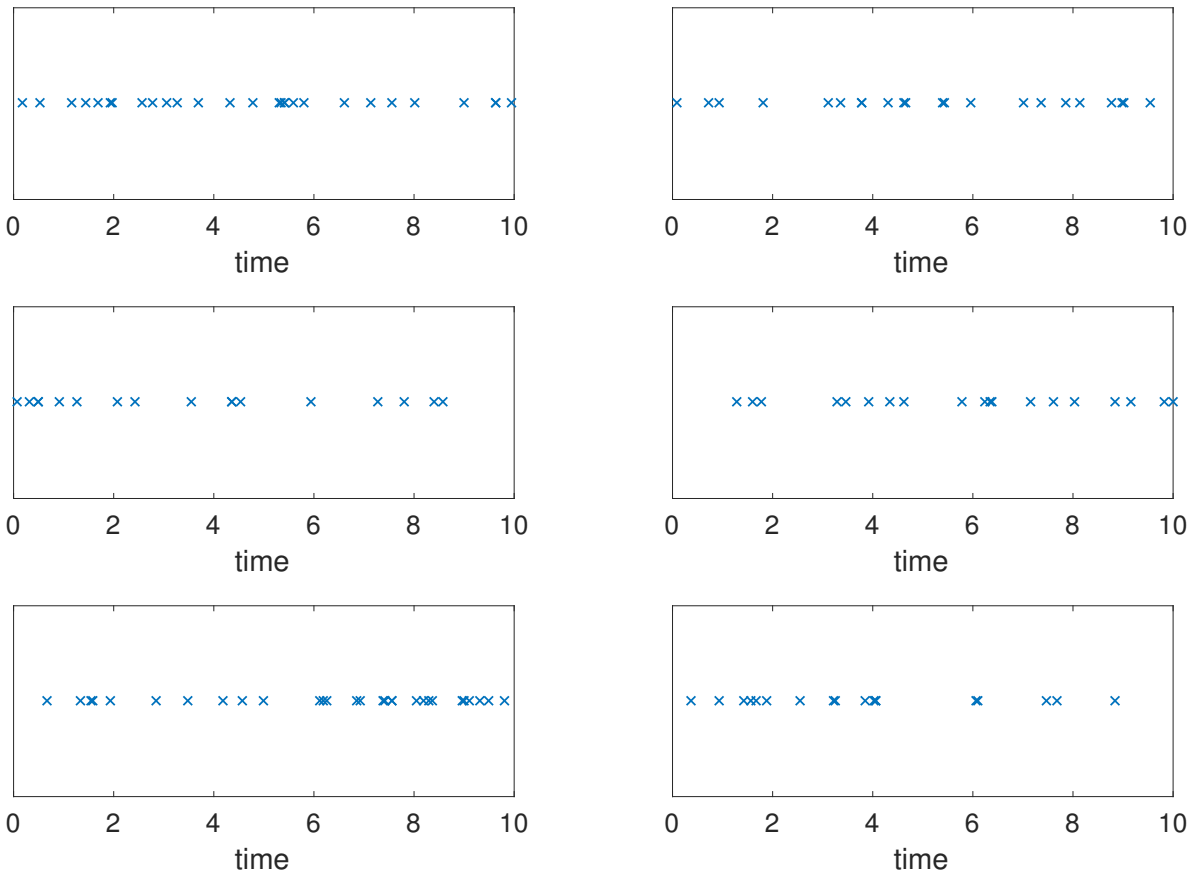
Solution

Simulating temporal Poisson processes

Suppose we wish to simulate a homogeneous Poisson process with intensity λ on some interval $(0, T]$. How might we go about doing so?

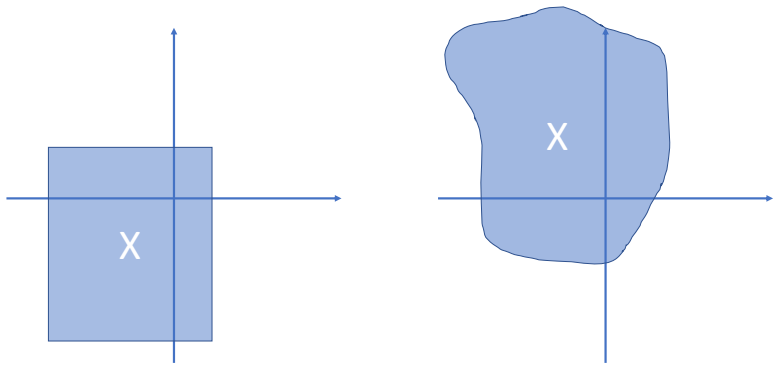
- We know that the number of events in $(0, T]$, i.e. $N(T) = N\{(0, T]\} \sim \text{Poisson}(\lambda T)$, so we first simulate a Poisson random variable with expected value λT. This will give us the total number of events in our realization.
  `n = rpois(1,lambda*T)`
- Then we wish to distribute these events in the region according to a uniform distribution.
  `E = runif(n,min=0,max=T)`
- It will probably be convenient to order these events in increasing order.
  `E = sort(E)`
- We could plot these with
  `plot(E,rep(0,length(E)),type="p",pch=19,cex=0.3)`

Realisations of a homogeneous Poisson process on (0,10] with intensity $\lambda = 2$



Spatial Point Processes

- The concept of a point process extends to higher dimensions (space).
- In this course, we will only be working with data in $\mathbb{R}^2$, so we will define it as such.
- We will in fact work with a region $X \subseteq \mathbb{R}^2$.
- It is often assumed the process exists on the entirety of $\mathbb{R}^2$, i.e. $X = \mathbb{R}^2$, however it could be some other region if we know the process is restricted to a certain domain.



Spatial Point Process

- We will again use the notation N to represent a point process, where for some set $A \subset X$, $N(A)$ is the random number of events in set A.

- We again start by defining the intensity of a spatial point process. In a slight switch of notation, we will let $\mathrm{d}\boldsymbol{s}$ be an infinitesimal *ball* (a disc in $\mathbb{R}^2$).

- We will solely be dealing with *simple* point processes

Definition (Simple)

N is called a simple point process if no two events can occur at exactly the same point in X.

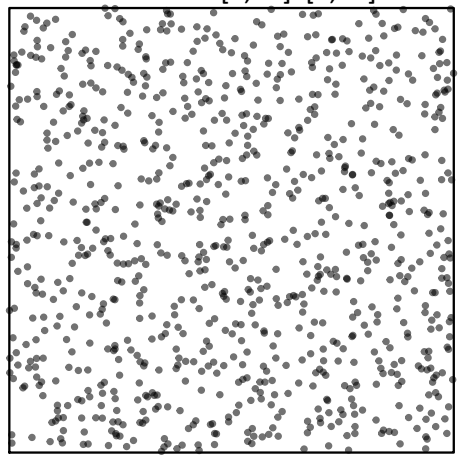
Intensity

Definition (Intensity)

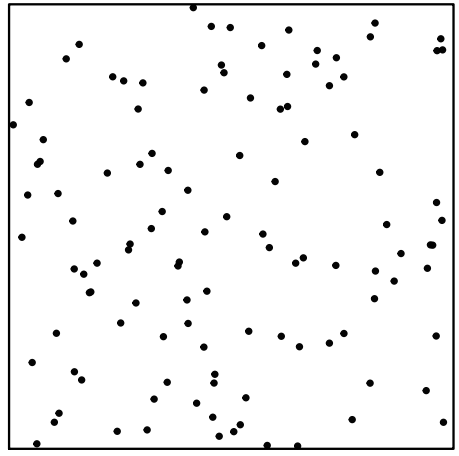
The intensity $\lambda(\boldsymbol{s})$ of a point process is the expected number of events per unit area, i.e.

$$\lambda(\boldsymbol{s}) = \lim_{|\mathrm{d}\boldsymbol{s}| \to 0} \frac{E\{N(\mathrm{d}\boldsymbol{s})\}}{|\mathrm{d}\boldsymbol{s}|}.$$

$\lambda = 10$ on [0,10]x[0,10]



$\lambda = 1$ on [0,10]x[0,10]



Intensity

A useful interpretation of the intensity is as follows. The number of events in the small ball $\mathrm{d}\boldsymbol{s}$ is a Bernoulli random variable (i.e. takes a value of either 0 or 1).

Therefore

$$E\{N(\mathrm{d}\boldsymbol{s})\} = 1 \cdot P\{N(\mathrm{d}\boldsymbol{s}) = 1\} + 0 \cdot P\{N(\mathrm{d}\boldsymbol{s}) = 0\} = P\{N(\mathrm{d}\boldsymbol{s}) = 1\}$$

and therefore

$$P\{N(\mathrm{d}\boldsymbol{s}) = 1\} = \lambda(\boldsymbol{s})|\mathrm{d}\boldsymbol{s}|.$$

Poisson Process

Definition (Poisson process)

Event process N is called a Poisson process if the following two conditions hold:

1. For any set $A \subset X$, $N(A)$ is Poisson distributed with expected value $\mu(A) = \int_A \lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$. We write $N(A) \sim \text{Poisson}(\mu(A))$.

2. For any disjoint pair of sets A and B, $N(A)$ and $N(B)$ are independent events.

Figure: Properties of a Poisson process

Homogeneous

Definition (Homogeneous)

A Poisson process N is called homogeneous is $\lambda(\boldsymbol{s}) = \lambda$ for all $\boldsymbol{s} \in X$. That is to say, it has a constant intensity over all space.

Processes that have a variable intensity are called inhomogeneous.

Homogeneous Poisson processes can be called *completely spatially random* (CSR). In other words, events are equally likely to occur anywhere, irrespective of where any other events are.

Question

Let N be a homogeneous Poisson process with intensity $\lambda = 2$, and let S be the unit disc ($r = 1$) centred at 0. Which of the following statements in FALSE?

(a) $E\{N(S)\} = 2\pi$.

(b) The number of events in S is random.

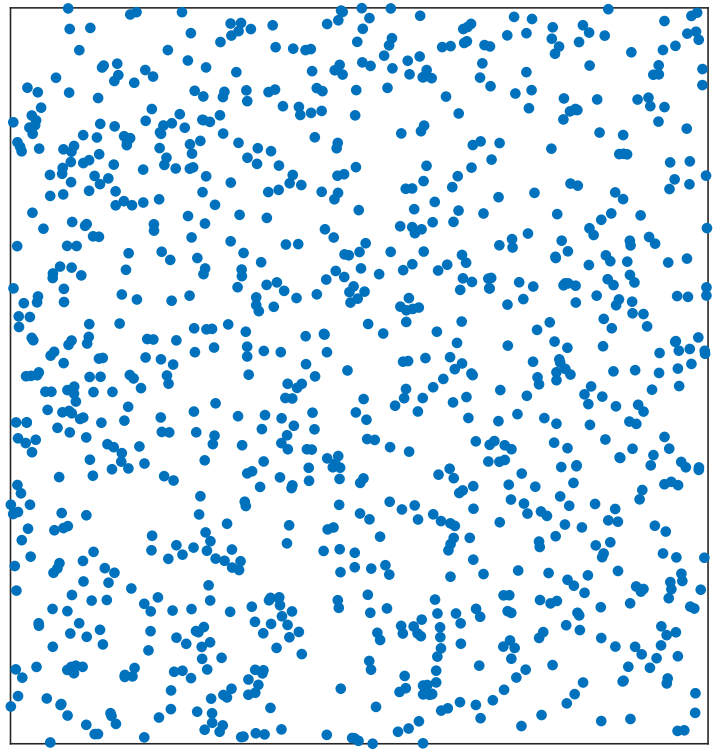
(c) $P\{N(S) = 0\} = \exp(-2\pi)$.

(d) $P\{N(S) = 1\} = \pi \exp(-2\pi)$.

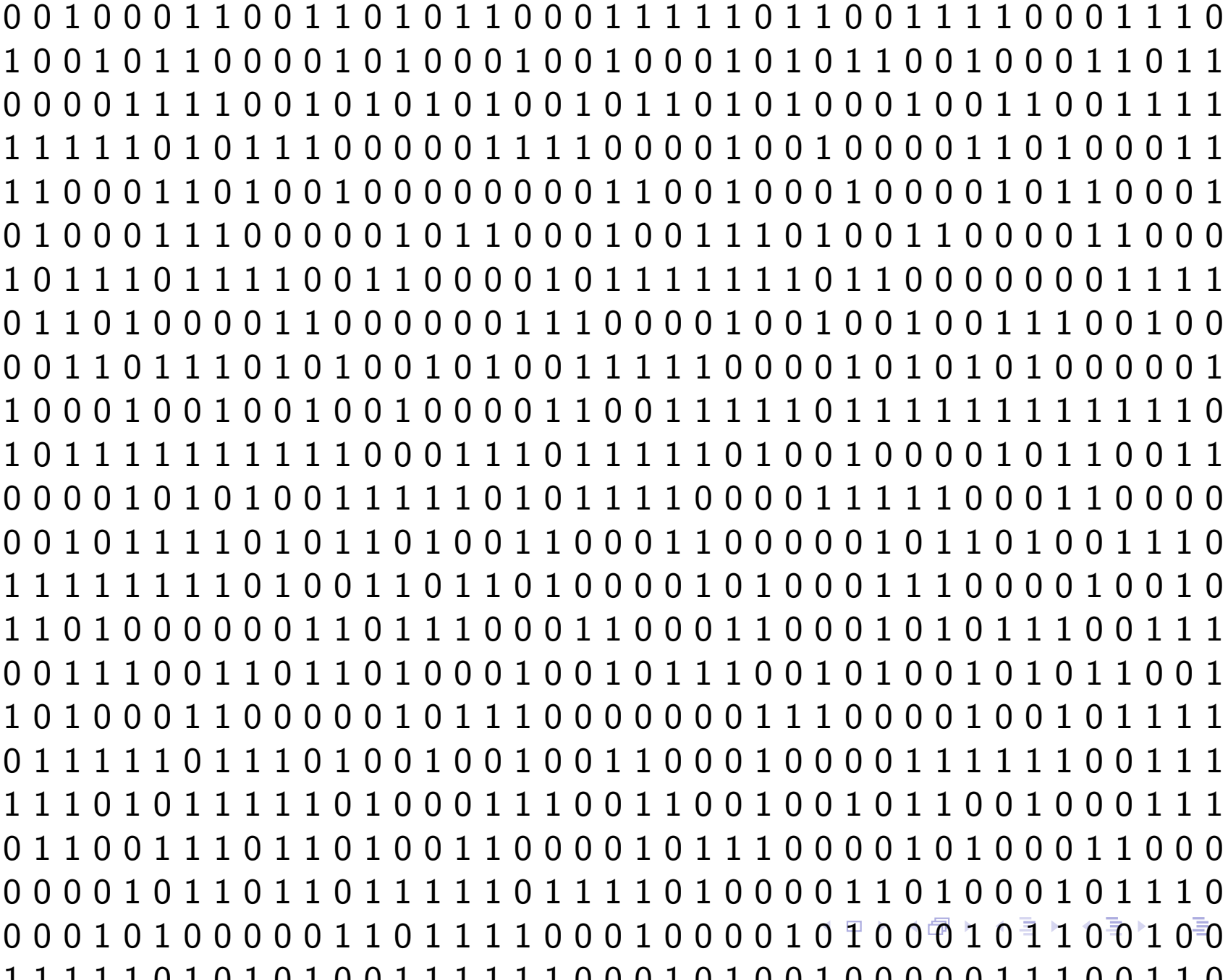
(e) $P\{N(S) > 0\} = 1 - \exp(-2\pi)$.

Solution

Complete spatial randomness



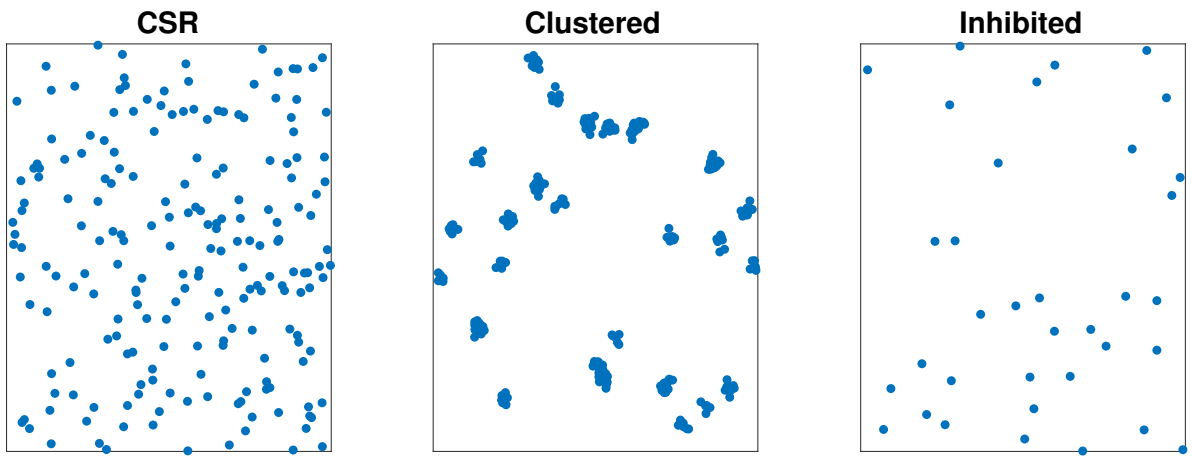
Complete spatial randomness



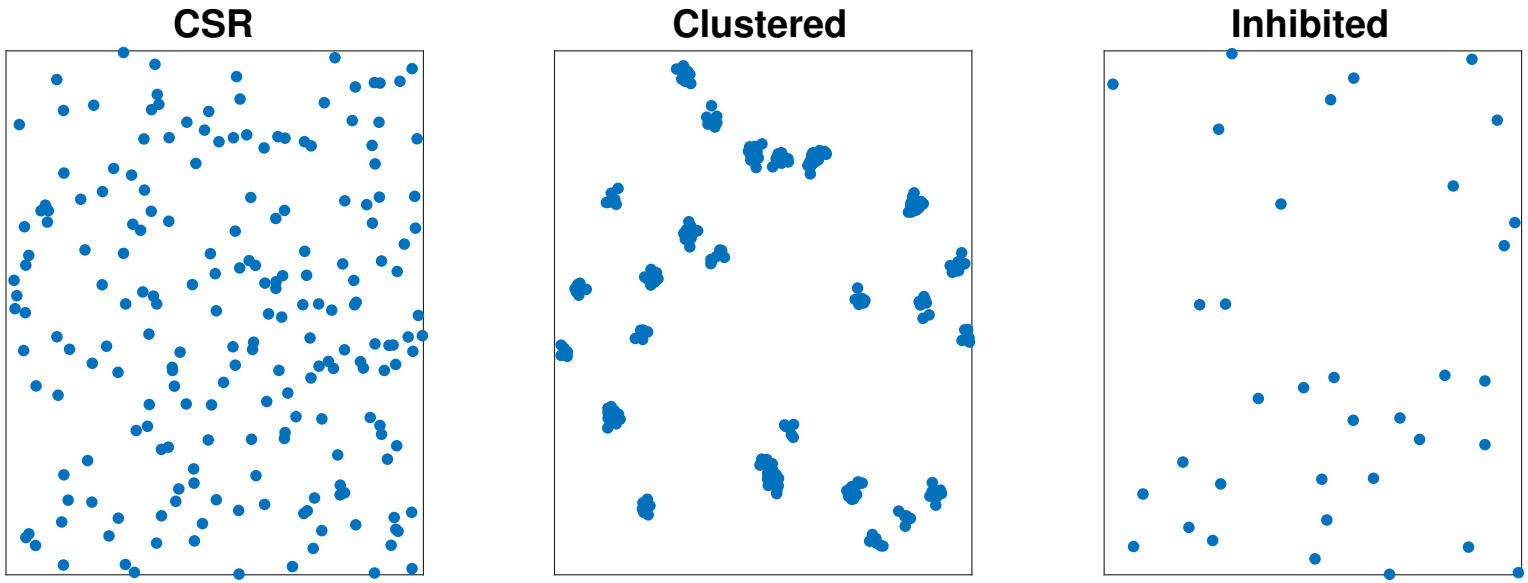
Different types of process

It again forms the dividing live between two other broad classes of process:

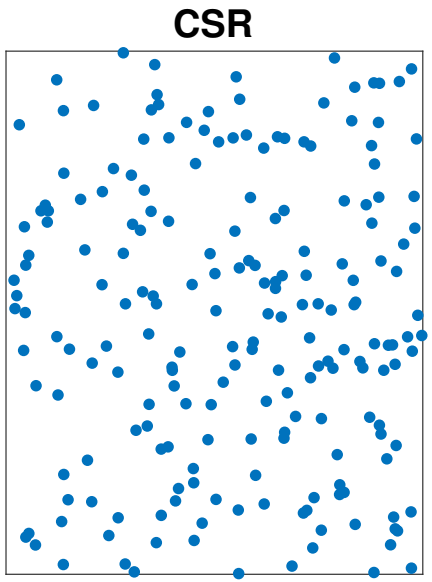
- Cluster processes tend to have events occur in groupings. What real-life spatial point processes exhibit clustering?
- Regular processes tend to have events that are well separated in space. What real-life spatial point processes exhibit regularity?

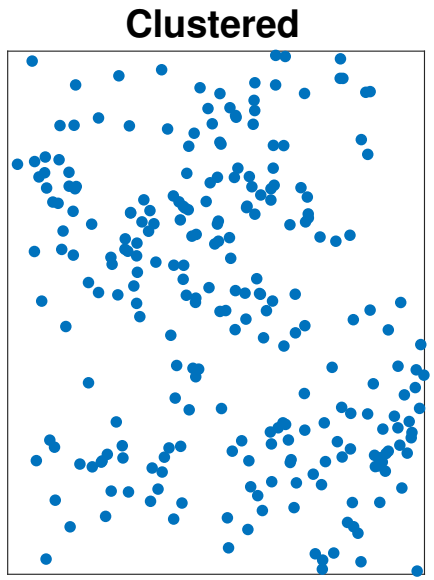


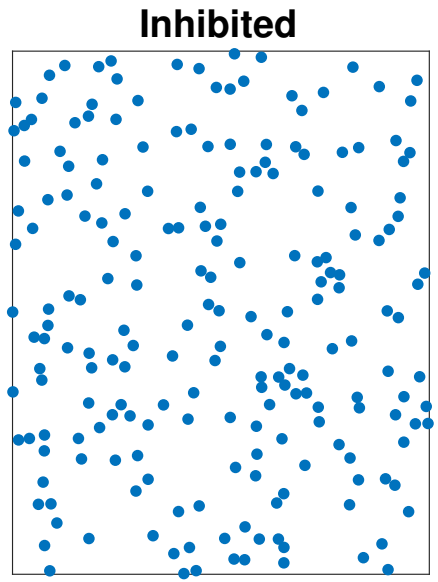
Obvious



Not obvious







Simulating spatial Poisson processes

- Suppose we wish to simulate a homogeneous Poisson process with intensity λ on some rectangular/square ROI $A = (a, b) \times (c, d)$.
- How might we go about doing so?
- We know that the $N(A)$, the number of events in A is Poisson($\lambda|A|$), so we first simulate a Poisson random variable with expected value $\lambda|A|$. This will give us the total number of events in our realization.
- Then we wish to distributed these events uniformly in the region.
- You could write your own piece of code to do this, or you could just use the function `rpoisspp` in `spatstat`.
- E.g.

```
N = rpoispp(10,win=c(0,10,0,10))
plot(N,type="p",pch=19,cex=0.3)
```

Second-order intensity and covariance intensity

- Recall from module 2, that covariance describes the joint variability of two random variables X and Y, and is defined as $E(XY) - E(X)E(Y)$.
- The covariance intensity extends this notion to point processes to describe joint variability in the process between two points in space.
- We first describe the second order intensity....

Second-order intensity

Definition (Second order intensity)

The second order intensity of a spatial point process at points $\boldsymbol{s}$ and $\boldsymbol{u}$ is given as

$$\gamma(\boldsymbol{s}, \boldsymbol{u}) = \lim_{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}| \to 0} \frac{E\{N(\mathrm{d}\boldsymbol{s})N(\mathrm{d}\boldsymbol{u})\}}{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|}.$$

Covariance intensity

We can now define the covariance intensity.

Definition (Covariance intensity)

The covariance intensity of a spatial point process at points $\boldsymbol{s}$ and $\boldsymbol{u}$ is given as

$$c(\boldsymbol{s}, \boldsymbol{u}) = \gamma(\boldsymbol{s}, \boldsymbol{u}) - \lambda(\boldsymbol{s})\lambda(\boldsymbol{u}).$$

The covariance intensity can be interpreted as the covariance between whether there is an event or not at $\boldsymbol{s}$ and at $\boldsymbol{u}$.

Covariance intensity for homogeneous Poisson

- Let us consider the homogeneous Poisson process as an example.
- We have said that all events happen independently of each other. This means that for two disjoint sets $A, B \subset \mathbb{R}^2$, we have

$$E\{N(A)N(B)\} = E\{N(A)\}E\{N(B)\}.$$

Therefore, for $\boldsymbol{s} \neq \boldsymbol{u}$, we have

$$E\{N(\mathrm{d}\boldsymbol{s})N(\mathrm{d}\boldsymbol{u})\} = E\{N(\mathrm{d}\boldsymbol{s})\}E\{N(\mathrm{d}\boldsymbol{u})\},$$

and hence

$$\begin{aligned} c(\boldsymbol{s},\boldsymbol{u}) &= \lim_{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|\to 0} \frac{E\{N(\mathrm{d}\boldsymbol{s})N(\mathrm{d}\boldsymbol{u})\}}{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|} - \lim_{|\mathrm{d}\boldsymbol{s}|\to 0} \frac{E\{N(\mathrm{d}\boldsymbol{s})\}}{|\mathrm{d}\boldsymbol{s}|} \lim_{|\mathrm{d}\boldsymbol{u}|\to 0} \frac{E\{N(\mathrm{d}\boldsymbol{u})\}}{|\mathrm{d}\boldsymbol{u}|} \\ &= \lim_{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|\to 0} \frac{E\{N(\mathrm{d}\boldsymbol{s})\}E\{N(\mathrm{d}\boldsymbol{u})\}}{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|} - \lim_{|\mathrm{d}\boldsymbol{s}|\to 0} \frac{E\{N(\mathrm{d}\boldsymbol{s})\}}{|\mathrm{d}\boldsymbol{s}|} \lim_{|\mathrm{d}\boldsymbol{u}|\to 0} \frac{E\{N(\mathrm{d}\boldsymbol{u})\}}{|\mathrm{d}\boldsymbol{u}|} \\ &= \lambda(\boldsymbol{s})\lambda(\boldsymbol{u}) - \lambda(\boldsymbol{s})\lambda(\boldsymbol{u}) = 0. \end{aligned}$$

Covariance intensity for homogeneous Poisson

- This demonstrates that an homogeneous Poisson process has a covariance intensity of zero between any two distinct points, i.e. there is no covariance between $N(\mathrm{d}\boldsymbol{s})$ (indicating whether an event occurs at $\boldsymbol{s}$) and $N(\mathrm{d}\boldsymbol{u})$ (indicating whether an event occurs at $\boldsymbol{u}$).

- Suppose the process is not homogeneous Poisson, instead an event occuring at $\boldsymbol{s}$ means we are *more* likely to see an event at $\boldsymbol{u}$ than by chance, then the covariance intensity $c(\boldsymbol{s}, \boldsymbol{u})$ would be positive.

- Oppositely, suppose an event occuring at $\boldsymbol{s}$ means we are *less* likely to see an event at $\boldsymbol{u}$ than by chance, then the covariance intensity $c(\boldsymbol{s}, \boldsymbol{u})$ would be negative.

Pair correlation function

Recall: correlation is a normalised measure of covariance. For a pair of random variables X and Y, it is defined as

$$\rho = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}.$$

This is a measure between -1 and 1 that adjusts for variance. We say the random variables are uncorrelated when $\rho = 0$.

We can again extend this concept to spatial point processes, however, the formulation is slightly different.

Pair correlation function

Definition (pair correlation function)

The pair correlation function for a spatial point process at $\boldsymbol{s}, \boldsymbol{u} \in \mathbb{R}^2$ is

$$g(\boldsymbol{s}, \boldsymbol{u}) = \frac{\gamma(\boldsymbol{s}, \boldsymbol{u})}{\lambda(\boldsymbol{s})\lambda(\boldsymbol{u})}.$$

Note that in the case of uncorrelated data (i.e. CSR), the pair correlation function is 1 (not zero) for all $\boldsymbol{s} \neq \boldsymbol{u}$.

Stationarity and Isotropy

Two key properties of a point process are *stationarity* and *isotropy*.

Definition (Stationary)

A point process N is called stationary if $\lambda(\boldsymbol{s}) = \lambda$ for all $\boldsymbol{s} \in X$ and $\gamma(\boldsymbol{s}, \boldsymbol{u})$ is a function only of $\boldsymbol{s} - \boldsymbol{u}$.

Definition (Isotropic)

A stationary point process N is called isotropic if $\gamma(\boldsymbol{s}, \boldsymbol{u})$ is a function only of the radial distance $r = \|\boldsymbol{s} - \boldsymbol{u}\|$.

Recall back to module 2, where you were taught about the concept of stationarity in time series. The definitions given here are in the same spirit.

- Stationarity tells us that no matter where we are in the X, the second order structure of the process looks the same.
- Isotropy (not defined for time series) tells that it looks the same in all directions.

Stationarity and Isotropy

This now means we can simplify the functions we've just presented for the stationary setting.

- Second order intensity: $\gamma(\boldsymbol{s}, \boldsymbol{u}) = \gamma(r)$.
- Covariance intensity: $c(\boldsymbol{s}, \boldsymbol{u}) = c(r) = \gamma(r) - \lambda^2$.
- Pair correlation function: $g(\boldsymbol{s}, \boldsymbol{u}) = g(r) = \gamma(r)/\lambda^2$.

Figure: Second-order intensity, covariance intensity and pair correlation function for CSR, clustered and regular processes.

Ripley's K-function

Ripley's K-function

In this section, we will give a detailed treatment of what is quite possibly the most commonly used analysis tool for spatial point patterns.

Definition

Ripley's K-function Let N be a stationary isotropic process, and let $N_0(r)$ represent the random number of events within a distance r of an arbitrarily chosen event (not including that event). Ripley's K-function is defined as

$$\begin{aligned} K(r) &= \lambda^{-1} E\{\text{number of events within a distance } r \text{ of an arbitrary event}\} \\ &= \lambda^{-1} E\{N_0(r)\}. \end{aligned}$$

Note: this is a theoretical function of the actual process.

Figure: Ripley's K-function

Ripley's K-function for Poisson process

- Let us consider the example of the Poisson process with intensity λ. The number of additional events within a disc of radius r drawn around some arbitrary event is distributed Poisson$(\lambda\pi r^2)$.
- Therefore, the expected number of events within that disc is $\lambda\pi r^2$. This gives

$$K(r) = \lambda^{-1}\lambda\pi r^2 = \pi r^2.$$

Question

Let N be a homogeneous Poisson process with intensity $\lambda = 10$. What is

$$E\{\text{number of events within a distance 2 of an arbitrary event}\}?$$

(a) 50π.

(b) 4π.

(c) 0.4π.

(d) 40π.

(e) π.

Solution

Assessing CSR

- This forms a baseline with which we can assess the clustering or regularity behaviour of a point process.
- If $K(r) > \pi r^2$, then we expect more points within a radius r of an arbitrary event than we would under complete spatial randomness. This implies clustering.
- If $K(r) < \pi r^2$, we expect fewer events than under complete spatial randomness. This implies regularity.

$L(r) - r$

It, in fact, makes some sense to standardise this. The $L-$function linearises $K(r)$ and is defined as

$$L(r) = \left\{ \frac{K(r)}{\pi} \right\}^{1/2}.$$

In the case of a homogeneous Poisson process, $L(r) = r$. We can further define the $L(r) - r$ function, which in the case of homogeneous Poisson process equals 0.
Therefore, we can broadly characterise a point process as follows

- Clustered: $L(r) - r > 0$
- CSR: $L(r) - r = 0$
- Regular: $L(r) - r < 0$

Figure: $L(r) - r$ function for CSR, Clustered and Regular data

Relating Ripley's K function with the second-order intensity and pair correlation

Let us consider the relationship between the K function and the second-order intensity $\gamma(r)$ of a stationary, isotropic point process N on $\mathbb{R}^2$.
It can be shown that [Diggle,Cressie]

$$\begin{aligned}\lambda K(r) &= \frac{2\pi}{\lambda}\int_0^r s\gamma(s)\mathrm{d}s; \qquad r > 0\\ &= 2\pi\lambda\int_0^r sg(s)\mathrm{d}s; \qquad r > 0.\end{aligned}$$

Question

Let N be a stationary and isotropic process with pair correlation function

$$g(r) = \begin{cases} 0 & r < \delta \\ 1 & r \geq \delta. \end{cases}$$

Which of the following is FALSE?

(a) $K(\delta) = 0$.

(b) $K(2\delta) = 4\pi\delta^2$

(c) E{number of events within distance δ of an arbitrary event} $\geq$ 0

(d) number of events within distance δ of an arbitrary event $= 0$.

(e) $K(0) = 0$

Solution

Estimating Ripley's K-function

First, let $E(r) = E\{N_0(r)\}$. It seems sensible to construct an estimator for $E(r)$ as follows. Let $d_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$, and define a crude estimator for $E(r)$ as

$$\tilde{E}(r) = n^{-1} \sum_{i=1}^{n} \sum_{j \neq i} I(d_{ij} \leq r).$$

Here, I is the indicator function which takes a value of 1 if the argument is true, and a value 0 if it is false.

Estimating Ripley's K-function

$\tilde{E}$ will be negatively biased. **Why?**

Several methods have been proposed to correct for this source of bias. We will focus on Ripley's method.

Edge Correction

- Let $w(\boldsymbol{s}, r)$ be the proportion of the circumference of the disc with centre $\boldsymbol{s}$ and radius r which lies within A.
- We will use the notation w_{ij} for $w(\boldsymbol{s}_i, \|\boldsymbol{s}_i - \boldsymbol{s}_j\|)$.
- Then, for any stationary isotropic processes, w_{ij} is the conditional probability that an event is observed, given there is an event a distance $r_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ away from the ith event $\boldsymbol{s}_i$.

Figure: Edge correction

Note: in general $w_{ij} \neq w_{ji}$.

Edge correction

Therefore, an unbiased estimator for $E(r)$ is

$$\hat{E}(r) = n^{-1} \sum_{i=1}^{n} \sum_{j \neq i} w_{ij}^{-1} I(d_{ij} \leq r).$$

Finally, to get an estimator for $K(r)$, replace the unknown intensity by an estimator for the intensity. The obvious choice is $n/|A|$.

The Ripley's estimator for $K(r)$ is

$$\hat{K}(r) = n^{-2}|A| \sum_{i=1}^{n} \sum_{j=1} w_{ij}^{-1} I(d_{ij} \leq r)$$

This estimator is approximately unbiased for small r

Weights - an aside

- The formula for the weights is a geometry problem and can be written down for a small number of cases.
- Consider the rectangle $A = (0, a) \times (0, b)$.
- Write $\boldsymbol{s} = (s_1, s_2)$ and let $\delta_1 = \min(s_1, a - s_1)$, $\delta_2 = \min(s_2, b - s_2)$.
- The values of δ_1 and δ_2 are the distances from the point $\boldsymbol{s}$ to the nearest vertical and horizontal edges of A, respectively. To calculate $w(\boldsymbol{s}, r)$, we need to distinguish two case:
  - if $r^2 \leq \delta_1^2 + \delta_2^2$, then

$$w(\boldsymbol{s}, r) = 1 - \pi^{-1}[\cos^{-1}(\min(\delta_1, r)/r) + \cos^{-1}(\min(\delta_2, r)/r)]$$

  - if $r^2 > \delta_1^2 + \delta_2^2$, then

$$w(\boldsymbol{s}, r) = 0.75 - (2\pi)^{-1}[\cos^{-1}(\delta_1/r) + \cos^{-1}(\delta_2/r)]$$

R implementation

Here is some example code:

```
# load point pattern from csv
N=read.table('test_data_dense.csv',sep=',')

# convert it into the ppp data format for SpatStat
N=as.ppp(N,c(0,1,0,1))

# estimate K(r) using Ripley's edge correction
K = Kest(N,correction = 'Ripley')

# plot K(r) and L(r)-r
plot(K)
plot(K,sqrt(./pi)-r ~ r)
```

Simple tests for complete spatial randomness

Hypothesis tests and Monte Carlo

Hypothesis tests

- A hypothesis test is a method of statistical inference.
- It lies at the very heart of scientific method.
- A hypothesis is proposed for either
  - the statistical relationship between two data sets, or
  - a data set obtained by sampling is compared to an idealized model for the data.
- We will restrict this discussion to the latter of these.

Null and alternative hypothesis

- The null hypothesis H_0 states that the sampled data comes from a specified model. It is typically associated with a contradiction to a theory one would like to prove.

- The alternative hypothesis H_A is typically associated with a theory one would like to prove.

- Examples:
  - H_0: The Higgs Boson does not exist
    vs
    H_A : The Higgs Boson does exist

  - H_0: This drug treatment does not cure the illness
    vs
    H_A : This drug treatment does cure the illness.

Test statistic

- For the data set I am testing the hypothesis with, I need to extract a summary statistic from it, this is known as the test statistic.
- I then wish to know how consistent this test statistic is with my null hypothesis,
- i.e. if, assuming the null hypothesis is true, this value of the test statistic would be very strange/extreme then I *reject* the null hypothesis.
- If the value of the test statistic is not strange/extreme under the null hypothesis then I *fail to reject* it.

p-value and significance level

- To do this, we determine the *p*-value. The *p*-value is defined as the probability under the null hypothesis that I see a test statistic at least as extreme as the value I observe.
- The threshold at which I deem the statistic to be strange or not is called the significance level. E.g., if my *p*-value < 0.05 I reject H_0 at the 5% level.
- The critical region for the test statistic is that in which I reject the null hypothesis.

Example 1

I think a coin I have is unfair. That is to say, I don't think the probability it lands heads (H) and the probability it lands tails (T) are both equal to 0.5. In fact, I think it favours heads.

Null hypothesis: H_0 : it is a fair coin.

Alternative hypothesis: H_A : the coin has a higher probability of falling H than T.

How would we write these hypotheses mathematically?

Null hypothesis: $H_0 : P(H) = P(T) = 0.5$.

Alternative hypothesis: $H_A : P(H) > 0.5$.

Example 1

I wish to run a hypothesis test at the 5% level. My test statistic is going to be the number of heads I observe, denoted $\#\text{Heads}$.

The distribution for the number of heads is Bionomial$(n, 0.5)$, and the corresponding probability mass function is

$$P(\#\text{Heads} = k) = {}^{n}C_k 0.5^k 0.5^{n-k} = {}^{n}C_k 0.5^n \qquad k = 0, ..., n$$

Therefore, the p-value for a particular observation t of the test statistic is

$$p = P(\#\text{Heads} \geq t) = \sum_{k=t}^{n} {}^{n}C_k 0.5^n.$$

Example 1

I now observe some data.

$$H\ H\ T\ H\ H\ T\ H\ T\ H\ H\ H\ H\ H\ T\ T\ H\ H\ H$$

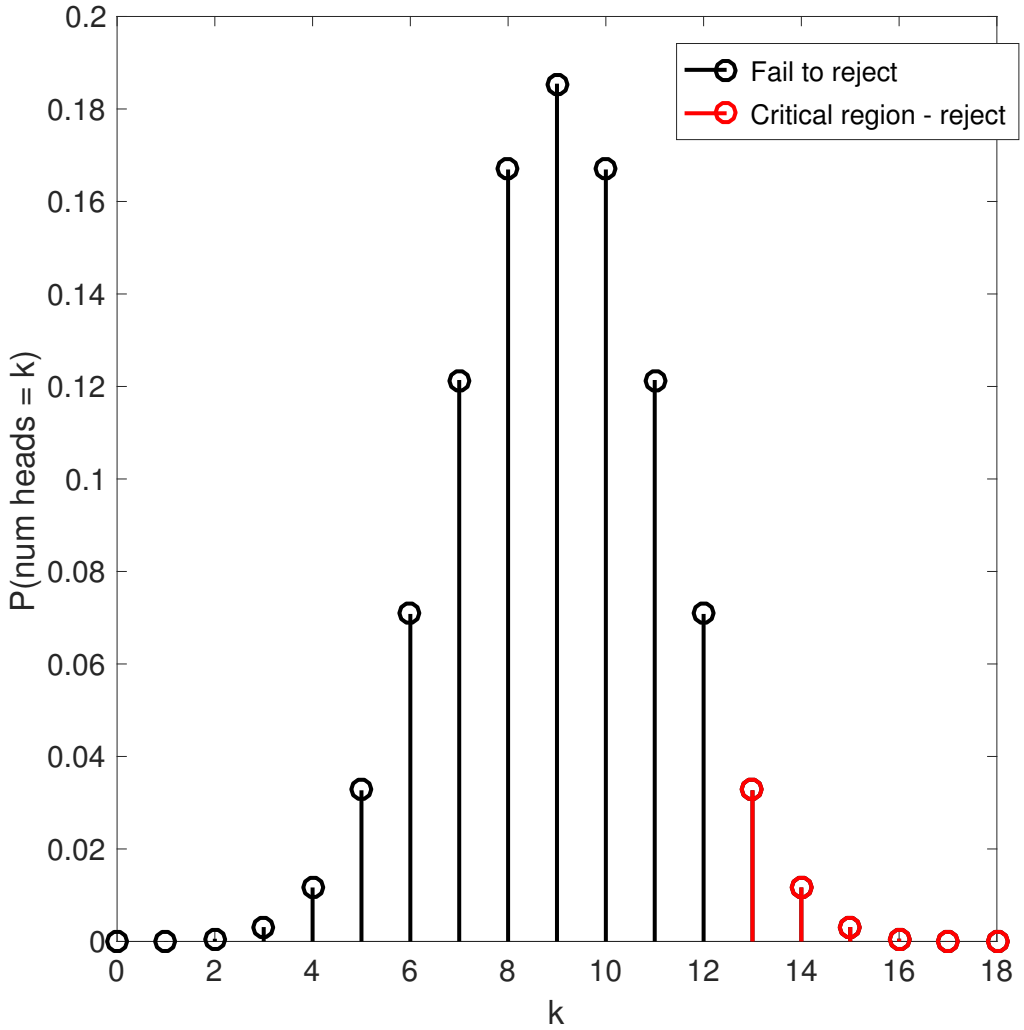
Therefore, here, my test statistic takes the value $t = 13$ and $n = 18$.

To determine whether this is significant, I determine the p-value for these data under the null hypothesis.

$$\begin{aligned} p = P(\#\text{Heads} \geq 13) &= \sum_{k=13}^{18} P(\#\text{Heads} = k) \\ &= \sum_{k=13}^{18} {}^{18}C_k 0.5^k 0.5^{18-k} = 0.0481 \end{aligned}$$

This is less than the 0.05, and I therefore reject my hypothesis at the 5% level.

Example 1



For this test, the critical region is $\mathcal{C} = \{13, 14, 15, 16, 17, 18\}$

Example 2

Consider a population with $X \sim N(\mu, \sigma^2)$, with both μ and σ^2 unknown. I wish to test the null hypothesis

$$H_0 : \mu = 0$$

against the alternative hypothesis

$$H_A : \mu \neq 0.$$

Suppose I observe n samples from the population $X_1, ..., X_n$. In this setting, my test statistics is

$$T = \frac{\bar{X}}{\hat{S}/\sqrt{n}}$$

where $\bar{X}$ is the sample mean and $\hat{S}$ is the sample standard deviation.

Example 2

Under the null hypothesis, $T \sim t_{n-1}$ ("Student" t-distribution with $n-1$ degrees of freedom).

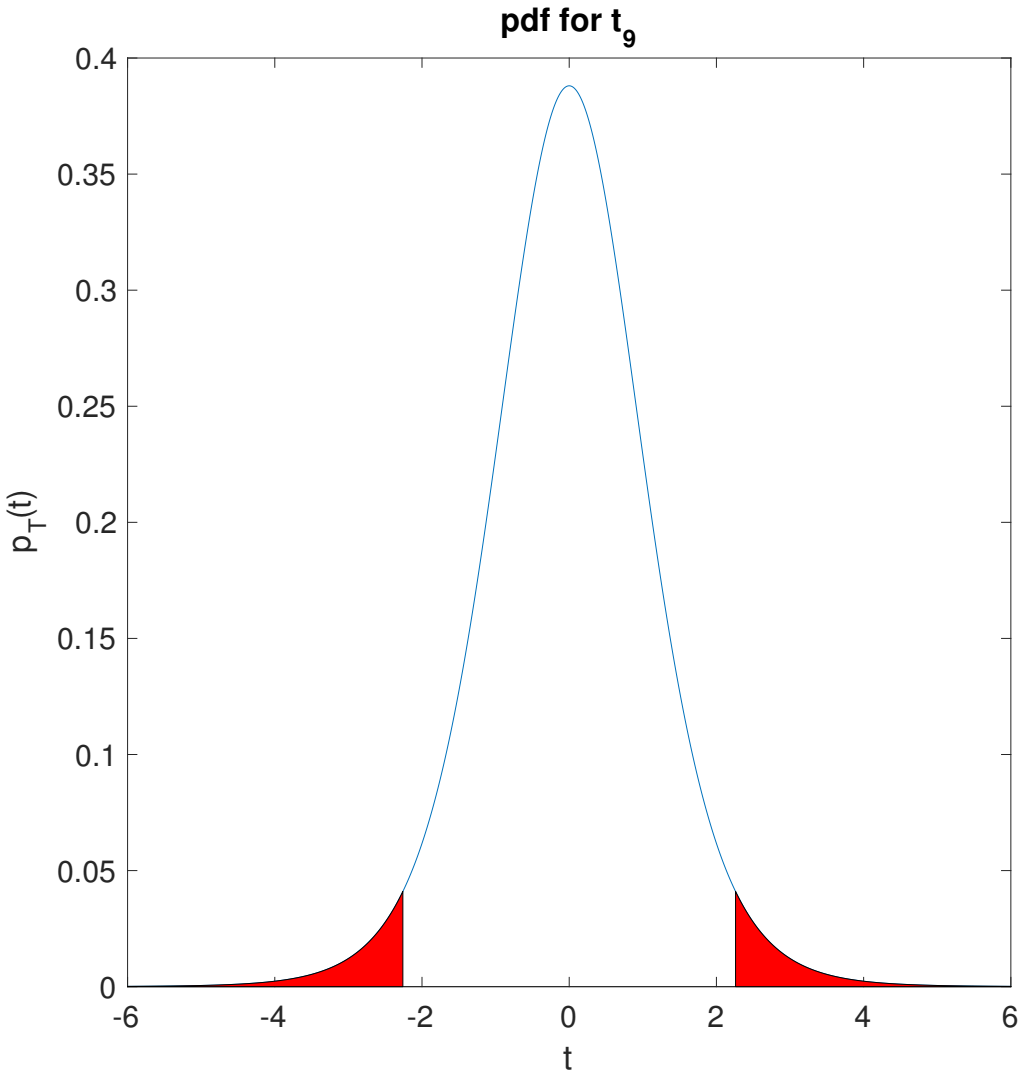
I observe the data

$$-0.0479, \quad 1.7013, \quad -0.5097, \quad -0.0029, \quad 0.9199,$$

$$1.4049, \quad 1.0341, \quad 0.2916, \quad -0.7777, \quad 0.1498, .$$

Here, $n = 10$, $\bar{x} = 0.4164$ and $\hat{s} = 0.8182$, thus $t = 1.6092$. It is a fact that $P(T < 1.6092) = 0.9290$. Therefore we fail to reject this at the 5% level.

Example 2



For this test, the critical region is $\mathcal{C} = (-\infty, -2.26] \cup [2.26, \infty)$

KEY POINT!!!!!!!!!!!

THE TEST STATISTIC IS RANDOM!!!

EACH TIME A NEW DATA SET IS SAMPLED FROM THE POPULATION, A DIFFERENT TEST STATISTIC IS OBTAINED!!!!

OUR GOAL IS TO DETERMINE WHAT THE DISTRIBUTION OF THIS TEST STATISTIC IS UNDER THE NULL.

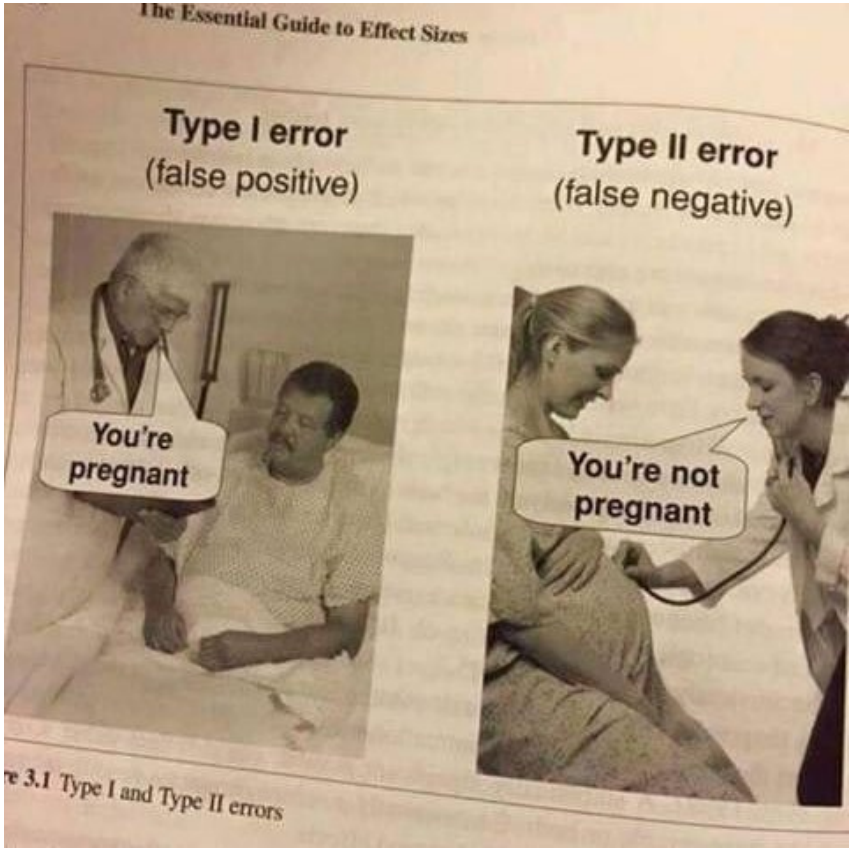
Statements of inference

- Correct things to say:
  - *"We reject the null hypothesis at the 5% level"* or *"The data is not consistent with the null hypothesis"*.
  - *"We fail to reject the null hypothesis"* or *"The data is consistent with the null hypothesis"*
- Incorrect thing to say:
  - *"We accept the null hypothesis"* or *"The alternative hypothesis is wrong"*.
  - *"The null hypothesis is wrong"*.
  - *"The p-value is the probability the null hypothesis is correct."*

Type I and Type II errors

- Type I error is the probability of incorrectly rejecting the null hypothesis when it is in fact true.
- This is set by the level of the test.
- Type II error is the probability of incorrectly failing to reject the null hypothesis when the alternative is in fact true.
- Power of test $=$ 1-type II error.

Type I and Type II errors



Question

I run a hypothesis test. The p-value of my test statistic is 0.02. Which of the following is a correct statement of inference?

(a) I reject the null hypothesis at the 1% level.

(b) I fail to reject the null hypothesis at the 10% level.

(c) The probability the null hypothesis is wrong is 0.02.

(d) I reject the null hypothesis at 5% level.

(e) The null hypothesis is wrong.

Monte Carlo testing

- In the previous examples, we have known what the distribution of the test statistic is under the null hypothesis.
- Suppose now this is not the case - we don't know the distribution of the test statistic under the null hypothesis.
- How can we proceed?

Example

- Return again to Example 1, but now suppose I've never been taught that the distribution for the number of heads from n tosses of a fair coin is Binomial(n,0.5). How can I test the null hypothesis?
- Easy! If I have a coin that I know is fair, and toss it n times (in our case 18 times) and record the number of heads, I have taken a sample from the null distribution.
- If I repeat this s times, I have taken s samples from the null distribution.
- I can then compare the original sample I observed with these s samples from the null to determine whether it is consistent with the null or not.

Example: Method

- Toss a fair coin 18 times, and record number of heads.
- Repeat this s times (s is typically 99, but the larger the better).
- Sort these s samples in ascending order.
- Find the $(1-\alpha)s$ sample of this list. This is my Monte Carlo estimate of the critical value.

Example: Implementation

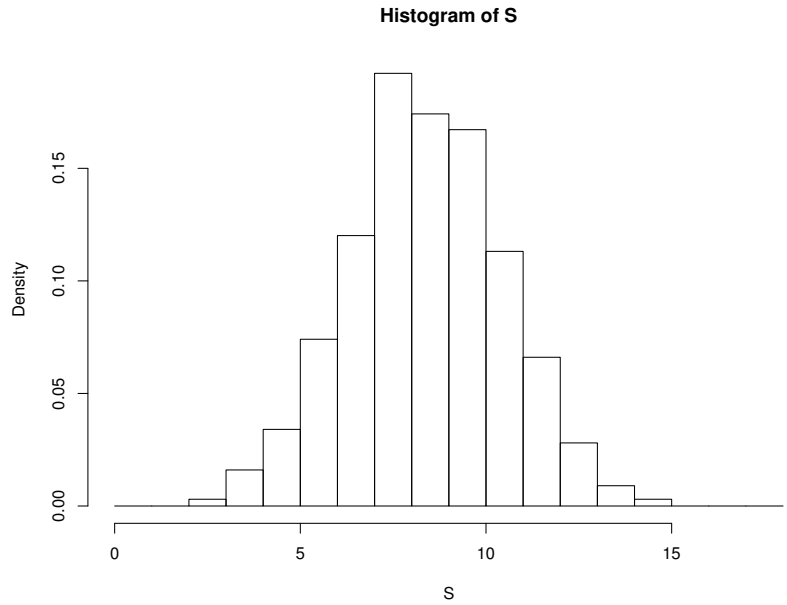
```
s = 99
S = sort(rbinom(s,18,0.5))
S[95]
```

```
 3  3  4  5  5  6  6  6  6
 6  6  6  6  6  6  7  7  7
 7  7  7  7  7  8  8  8  8
 8  8  8  8  8  8  8  8  8
 8  8  8  9  9  9  9  9  9
 9  9  9  9  9  9  9  9  9
 9  9  9  9  9 9 10 10 10 10
10 10 10 10 10 10 10 10 11 11
11 11 11 11 11 11 11 11 11 12
12 12 12 12 13 13 13 13 13 13

13 14 14 14 16
```

Example: Implementation

```
s = 999
S = sort(rbinom(s,18,0.5))
S[950]
```




```
S[950] = 13
```

Testing for complete spatial randomness

An immediate question a data analyst will ask when presented with event data is:

Are these events completely spatially random? If not, do they exhibit clustering or regularity?

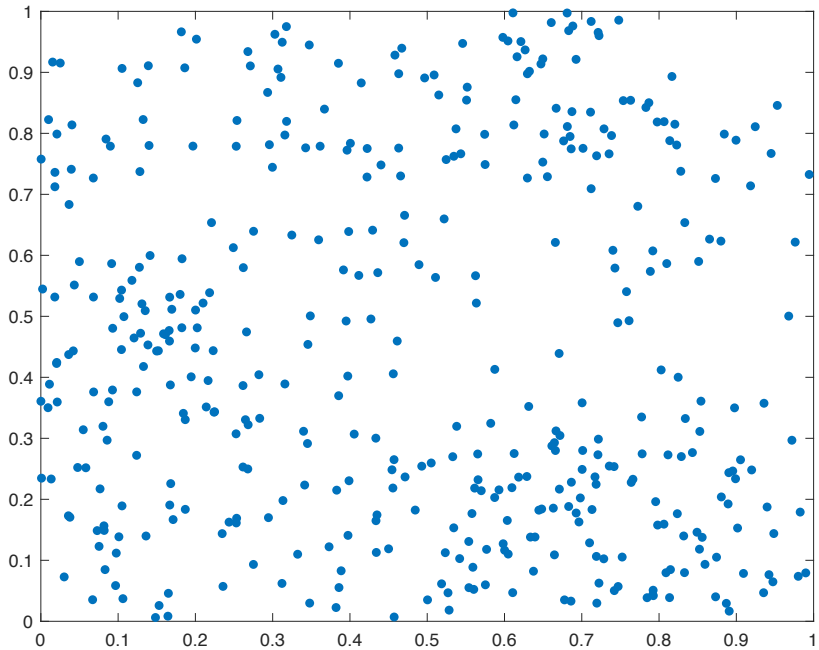


Figure: Dataset: test_data_dense.csv

Testing for complete spatial randomness

In this case, the null hypothesis is

H_0: The spatial point process N that generates these data is completely spatially random.

or, equivalently

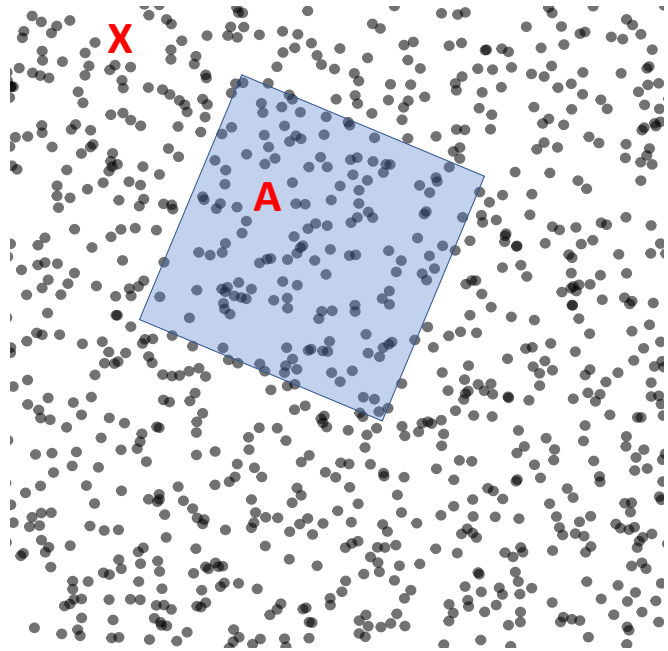
H_0: The spatial point process N that generates these data is homogeneous Poisson.

Versus the alternative Hypothesis

H_A: N is not completely spatially random/homogeneous Poisson process.

Sampling

- Typically, we will observe the point pattern on a limited region.
- When testing, we have to take into account that the point pattern continues outside of the region.



Testing for complete spatial randomness

- To test the null hypothesis we will need to summarise the spatial properties of the observed data.
- We will then wish to know what this summary would normally look like if the null hypothesis is true to determine if what we observe is consistent with it being completely spatially random, or if it is unusual/weird for CSR and hence evidence that the null hypothesis is not actually correct.
- Let us first consider some methods for summarising the spatial data.

Nearest Neighbour Distances

- For n events in a ROI A, let d_i denote the distance from the ith event to the nearest other event in A.
- The set $\{d_1, d_2, ..., d_n\}$ are called the *nearest neighbour distances*.
- Typically, the n nearest neighbour distances will contain duplicates if nearest neighbour pairs are reciprocal.

Figure: Nearest neighbour distances

Nearest neighbour distribution

Definition (Nearest neighbour distribution function)

For a stationary process, let D denote the distance from an arbitrary event to the nearest other event. The nearest neighbour distribution (NND) function is defined as

$$G(d) \equiv P(D \leq d).$$

Another way to write this is

$$G(d) = 1 - P(\text{no event within distance } d \text{ of some arbitrary event})$$

Question

The nearest neighbour distribution function $G(d)$ for the homogeneous Poisson process with intensity λ is

(a) $1 - \exp(-\lambda)$

(b) $\exp(-\lambda\pi d^2)$

(c) πd^2

(d) $\lambda\pi d^2$

(e) $1 - \exp(-\lambda\pi d^2)$.

Nearest Neighbour Distances

We can define the Empirical Distribution Function (EDF) for the nearest neighbour distribution function as

$$\hat{G}(d) = n^{-1}\#(d_i \leq d).$$

Figure: EDF for Nearest Neighbours Distances

Computing $G(d)$ in R

Here is some example code:

```
# load point pattern from csv
N=read.table('test_data_sparse.csv',sep=",")

# convert it into the ppp data format for SpatStat
N=as.ppp(N,c(0,1,0,1))

# estimate G(d)
G = Gest(N)

# plot G(d)
plot(G)
```

Testing

- We wish to compare the estimated EDF with what it should be under CSR. Large deviations away from it would indicate the null hypothesis should be rejected

- The theoretical distribution of nearest neighbour distance D under CSR depends on n and A, and is not expressible in closed form because of complicated edge effects.

Large n approximation‘

- If we, for the moment, ignore edge effects, then by noting $|A|$ denotes the area of A, then $\pi d^2 |A|^{-1}$ is the probability under CSR that an arbitrary event is within distance d of a specified event. Since the events are located independently, the distribution of D can be approximated as

$$G(d) = 1 - (1 - \pi d^2 |A|^{-1})^{n-1} : \ d \geq 0$$

  For large n, a further approximation is

$$G(d) = 1 - \exp(-\lambda \pi d^2) : \ d \geq 0.$$

- This has therefore reduced to the theoretical result for the nearest neighbour distribution of a Poisson process.

Testing G(r)

- Remember: Our EDF is random. It is dependent on the realization. A different realization *of the same process* would give a different EDF.
- Our goal is to determine if the observed EDF is consistent with CSR data (fail to reject H_0), or if it seems abnormal (reject H_0).
- To do so, we need to determine the sampling distribution of $\hat{G}(d)$ under complete spatial randomness.
- Analytically, this is very difficult!!

Simulation envelopes

- To tackle this problem, we consider simulation envelopes. Proceed with the following:
- Call our EDF $\hat{G}_0(d)$
- Create s (s typically 99) independent simulations of n uniformly distributed points in ROI A. For each one of these simulated set of points, compute $\hat{G}_i(d)$, $i = 1, .., s$.
- Define the upper and lower simulation envelopes,

$$U(d) = \max_{i=1,..,s} \{\hat{G}_i(d)\}; \qquad L(d) = \min_{i=1,..,s} \{\hat{G}_i(d)\}.$$

These two could be plotted against d, and have the property that under CSR, and for each d,

$$P\{\hat{G}_0(d) > U(d)\} = P\{\hat{G}_0(d) < L(d)\} = s^{-1}.$$

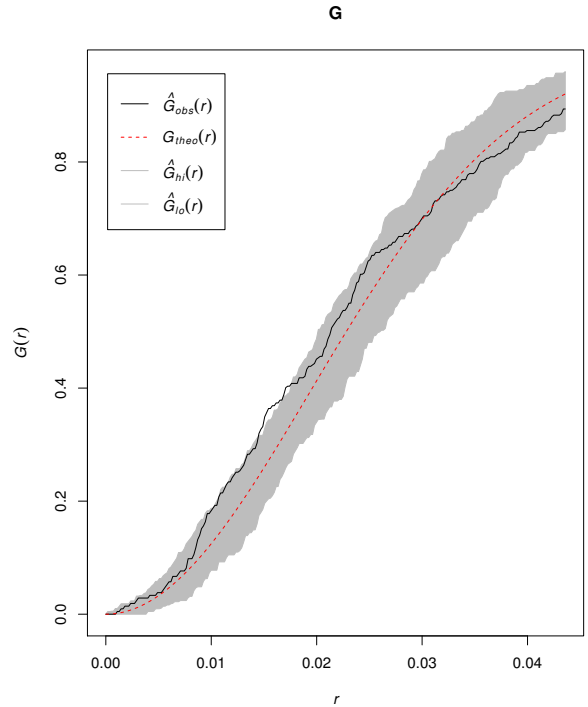
Formation of envelope plots

Figure: Formation of envelope plots

Producing envelope plots in R

The following code will produce envelope plots for $\hat{G}(r)$ in R

```
Genv = envelope(N, fun = Gest)
plot(Genv)
```



Question

Which of the following is a correct statement of inference?

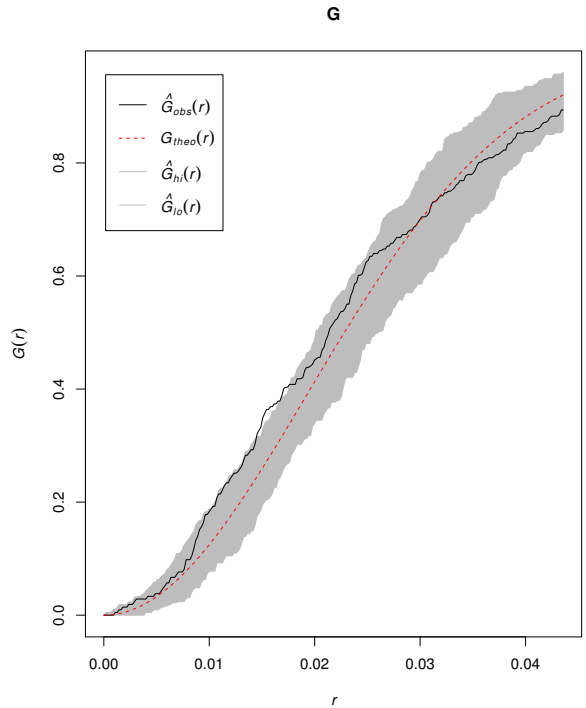
(a) The data is consistent with complete spatial randomness.

(b) The process is not completely spatially random.

(c) The data is not consistent with complete spatial randomness.

(d) The process is clustered.

(e) None of the above.



- Simulation envelopes are not a rigorous test, but instead intended to be a visual aid to assess how well the data matches the hypothesis of CSR.

- If we instead wanted to do a more formal Monte Carlo test for CSR we could choose one of the following methods.

Monte Carlo test 1

- For the observed point pattern compute the sample mean of the n nearest neighbour distances $\{d_1, ..., d_n\}$

$$\bar{d}_0 = \sum_{i=1}^{n} d_i.$$

- Simulate s realizations (s typically 99) of a CSR process on the same ROI. For $j = 1, ..., s$, compute the sample mean of the n nearest neighbour distances

$$\bar{d}^{(j)} = \sum_{i=1}^{n} d_i^{(j)}.$$

- Sort them and relabel such that $\bar{d}^{(1)} < \bar{d}^{(2)} < ... < \bar{d}^{(s)}$.
- To obtain the critical region for an α level test, take the $\alpha s/2$ and $(1-\alpha)s/2$ elements of this list.

Monte Carlo test 2

An alternative test statistic for large n is

$$u_0 = \int \{\hat{G}_0(y) - G(y)\}^2 \mathrm{d}y$$

where $G(\cdot)$ is the theoretical nearest neighbour distribution function given earlier.

We can then simulate s realizations of a CSR process on the same ROI. For $j = 1, ..., s$, compute

$$u_j = \int \{\hat{G}_j(y) - G(y)\}^2 \mathrm{d}y$$

and test in the same way.

Point to nearest event distribution

A closely related analysis looks at the distance from an arbitrary point in space to the nearest event.

Definition (Point to nearest event distribution)

For a stationary process, let D denote the distance from an arbitrary point to the nearest event. The point to nearest event distribution (PNED) function is defined as

$$F(d) \equiv P(D \leq d).$$

Another way to write this is

$$F(d) = 1 - P(\text{no event within distance } d \text{ of some arbitrary point})$$

Empirical PNED

- We estimate $F(d)$ by analysing the distances x_i from each of m sample points in A to the nearest of the n events.

- The EDF
$$\hat{F}(x) = m^{-1}\#(x_i \leq x)$$
measures the *empty* spaces in A.

- This is because $\{1 - \hat{F}(x)\}|A|$ is an estimate of the area $|B(x)|$ of the region $B(x)$ consisting of all points in A a distance at least x from every one of the n events in A.

Figure: point to nearest event distances and forming the ECDF

Point to nearest event distances

- Under the same reasoning as the nearest neighbour distribution function,

$$F(x) = 1 - \exp(-\pi\lambda x^2) \qquad x \geq 0,$$

  approximately, where $\lambda = n|A|^{-1}$.

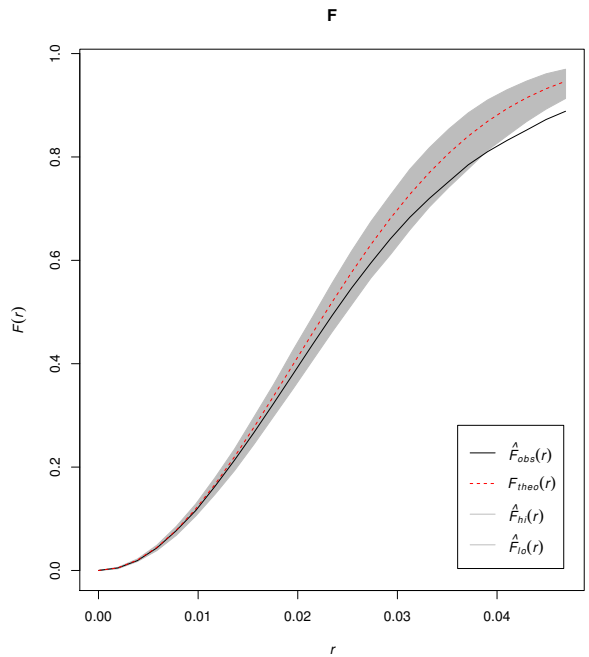
- We are faced with the challenge of how to choose m, the number of points at which we measure the nearest event distance. There has been some recommendation that a $k \times k$ grid should be chosen where $k \approx \sqrt{n}$.

- A Monte Carlo test of CSR can be performed in an analogous manner to that used the nearest neighbour distances.

Producing envelope plots in R

The following code will produce envelope plots for $\hat{F}(r)$ in R

```
Fenv = envelope(N, fun = Fest)
plot(Fenv)
```



Ripley's K-function

Using Ripley's K-function for testing CSR follows a very similar approach to that demonstrated previously. Although approximate distributions exist in the case of CSR in rectangular regions, they are messy, and instead we consider the Monte Carlo approach.

Simulation envelopes

The first exploratory data technique is to consider simulation envelopes. The procedure is as follows:

- Compute $\hat{K}_0(r)$, the estimate of Ripley's K-function for the point pattern under analysis.
- Estimate the intensity of the point process as $\hat{\lambda} = n/|A|$.
- Simulate s (s typically 99) Poisson processes with intensity $\lambda = \hat{\lambda}$.
- For each one, compute $\hat{K}_i(r)$, $i = 1, ..., s$ on a vector of values for r.
- Compute the upper and lower simulation envelopes

$$U(r) = \max_{i=1,..,s} \{\hat{K}_i(r)\}; \qquad D(r) = \min_{i=1,..,s} \{\hat{K}_i(r)\}.$$

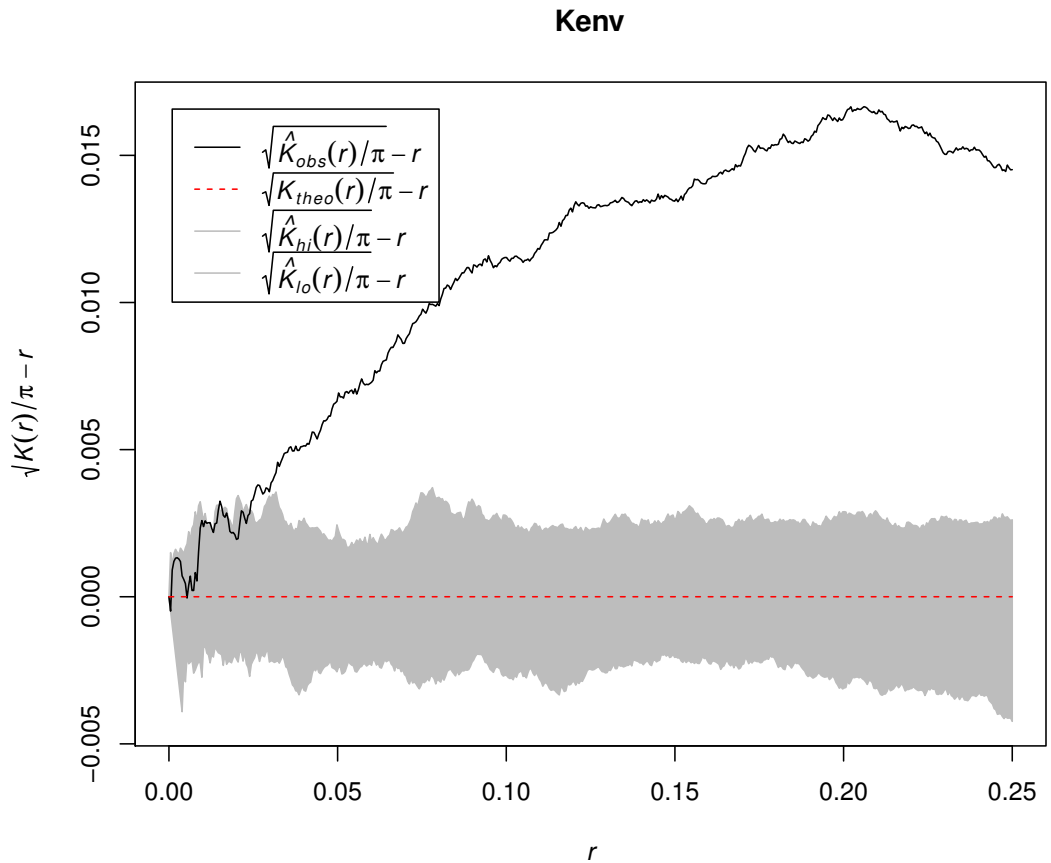
These could be plotted against r and have the property that under CSR, and for each t,

$$P\{\hat{K}_0(r) > U(r)\} = P\{\hat{K}_0(t) < D(r)\} = s^{-1}.$$

Envelope plots

It is typically more useful to look at $\hat{L}_0(r) - r = (\hat{K}_0(r)/\pi)^{1/2} - r$ and the simulation envelopes $(U(r)/\pi)^{1/2} - r$ and $(D(r)/\pi)^{1/2} - r$

```
Kenv = envelope(N,Kest)
plot(Kenv,sqrt(./pi)-r ~ r)
```



Hypothesis test

An alternative approach is to obtain a summary test statistic from $\hat{K}_0(r)$. There are two common test statistics that have been proposed.

- The first measures *globally* how different $\hat{K}_0(r)$ is from the theoretical value of πr^2,

$$T_I = \int_0^{r_{\max}} \left\{ \hat{L}_0(r) - r \right\}^2 \mathrm{d}r.$$

- The second measures the maximum deviation $\hat{K}_0(r)$ takes from the null hypothesis value of πr^2

$$T_S = \sup_r |\hat{L}_0(r) - r|$$

Figure: Test statistics

Quadrat counts

- The methods discussed thus far are what we call distance based approaches (as they analyse the distances to or between events).
- An alternative to a distance-based approach is to partition A into m sub-regions, or *quadrats*, or equal area and use the event counts in each of the m quadrats to test for CSR.
- How we go about choosing the m quadrats is completely arbitrary, but if we have a square ROI, it makes sense to chop this up into a $k \times k$ grid of square sub-regions, so that $m = k^2$.

Quadrat counts

- Denote $n_1, n_2, ..., n_m$ to be the event counts in each quadrat. Let $\bar{n} = n/m$, the sample mean of n_i. A sensible statistic with which to test for departures from the uniform distribution, implied by CSR, is

$$X^2 = \sum_{i=1}^{m} (n_i - \bar{n})^2 / \bar{n}.$$

- If the events are uniformly distributed on A, we expect this to be small. If they are not evenly distributed we expect this to be large.

- We note, this does not necessarily tell us about whether it is truly *uniform* (in the distributional sense). For example, a regular grid of events will also give a low value.

- However, if we also note that this is $m-1$ times the ratio of the sample variance to the sample mean, we can obtain more insight.

- Recall that the expected value and variance of a Poisson distributed random variable equal one another. Therefore, in the case of CSR, where each quadrat count will be a Poisson random variable, we expect this ratio to be close to $m-1$.

- If it is higher than this then it means the variability amongst quadrats is high, this could imply clustering. If this ratio is low it means the variance between quadrat counts is low could imply regularity.

- It has been shown that in the case of CSR data $X^2 \sim \chi^2_{m-1}$, and a hypothesis test can be constructed using this.

Figure: Hypothesis testing for CSR using Quadrat counts

Example

```
> quadrat.test(N)

        Chi-squared test of CSR using quadrat counts
        Pearson X2 statistic

data:  N
X2 = 94.868, df = 24, p-value = 4.444e-10
alternative hypothesis: two.sided

Quadrats: 5 by 5 grid of tiles
```

Alternative models of spatial point processes

Thus far, the only *model* of a spatial point process we have encountered is the Poisson process. Here we consider some simple models for alternative processes.

Clustered processes

The following model is called the *Thomas process*. It is used to model clustered event data seen in ecology and microscopy images, amongst others.

1. Let the parent process be homogeneous Poisson (with intensity $\lambda_p(\boldsymbol{s}) = \lambda_p$).
2. Each parent produces a Poisson(ξ) distributed number of offspring.
3. The positions of the offspring relative to their parents are independent and identically distributed according to a 2-dimensional Gaussian distribution $N_2(\mu, \sigma^2 I)$.
4. The final process is composed of the superposition of offspring only.

Figure: Thomas Cluster Process

First and second order properties

We will consider the first and second order properties of a Thomas process N with parent intensity λ_p and Poisson(ξ) distributed number of offspring.

Proposition

Let N be a Thomas process with parent intensity λ_p and offspring rate ξ. The intensity of N is $\lambda = \lambda_p \xi$.

Proposition

Let N be a Thomas process with parent intensity λ_p and offspring rate ξ, then N is both stationary and isotropic.

First and second order properties

Proposition

Let N be a Thomas process with parent intensity λ_p and offspring rate ξ, the second order intensity is given by

$$\gamma(r) = \lambda^2 + \lambda_p \xi^2 \frac{r}{2\sigma^2} \exp(-r^2/4\sigma^2); \qquad r > 0,$$

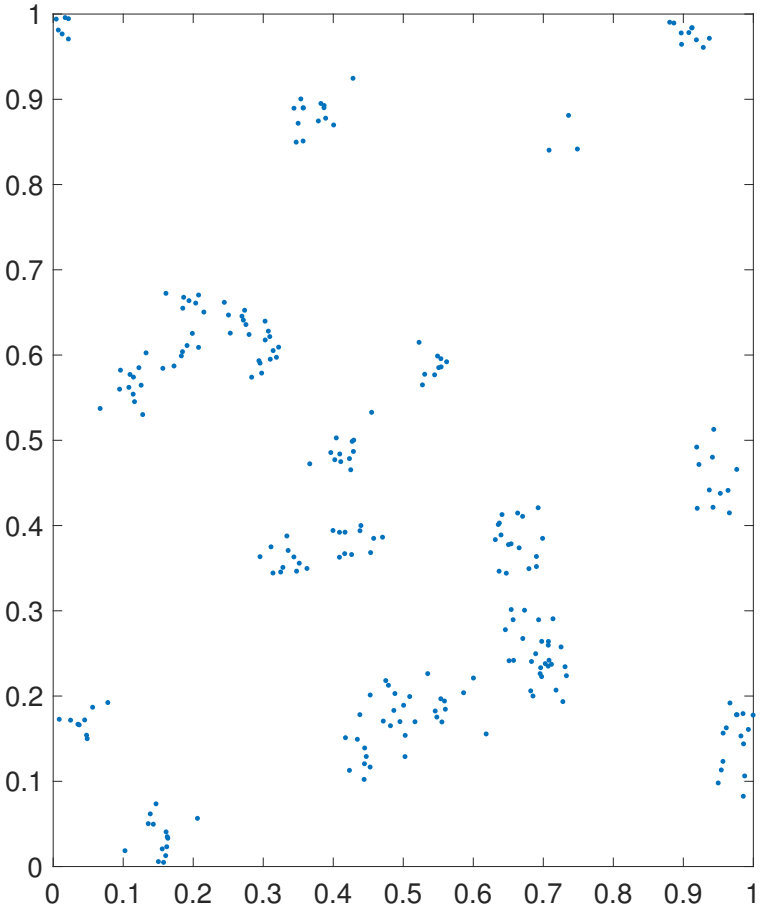
and consequently, the K-function is given as

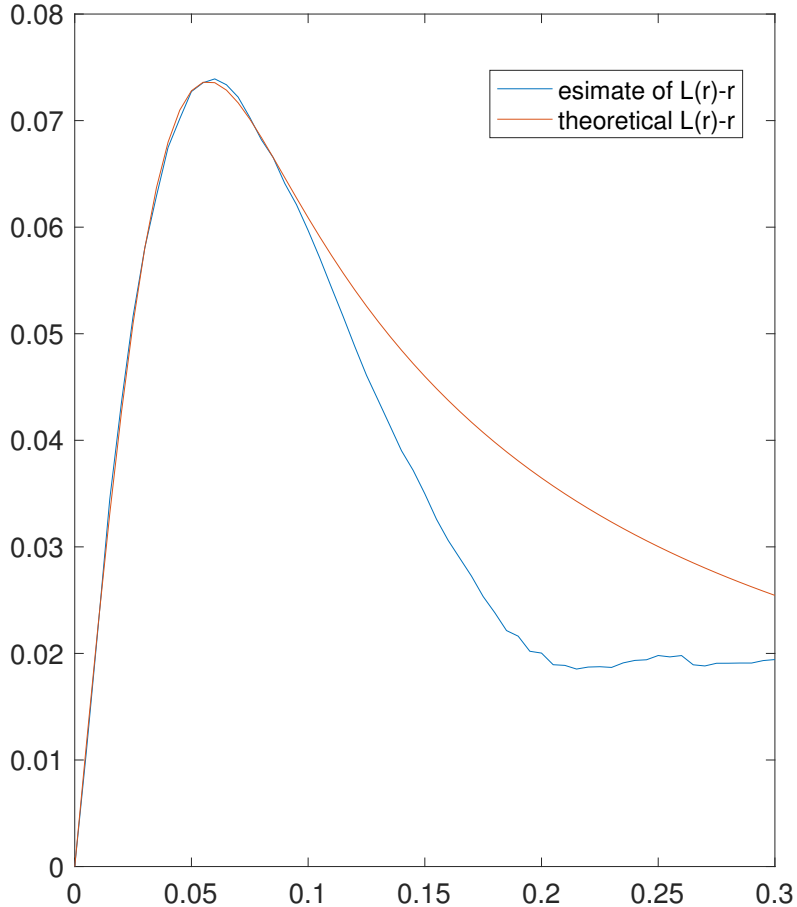
$$K(r; \sigma^2, \lambda_p) = \pi r^2 + \lambda_p^{-1}\{1 - \exp(-r^2/4\sigma^2)\}; \qquad r > 0.$$

The proof of this is reasonably involved, c.f. Cressie.

Thomas process

Thomas process with $\sigma = 0.02$, $\lambda_p = 20$, $\xi = 10$.





Model fitting

The K-function can be used for fitting a Thomas process model. The procedure is as follows:

- For the dataset, compute estimate $\hat{K}(r)$ of Ripley's K function.
- Find values of σ and λ_p that minimize

$$\int_{r_{\min}}^{r_{\max}} \left| \{K(r; \sigma^2, \lambda_p)\}^{0.25} - \{\hat{K}(r)\}^{0.25} \right|^2 \mathrm{d}r.$$

This can be performed in `R` with `thomas.estK`.

Model fitting

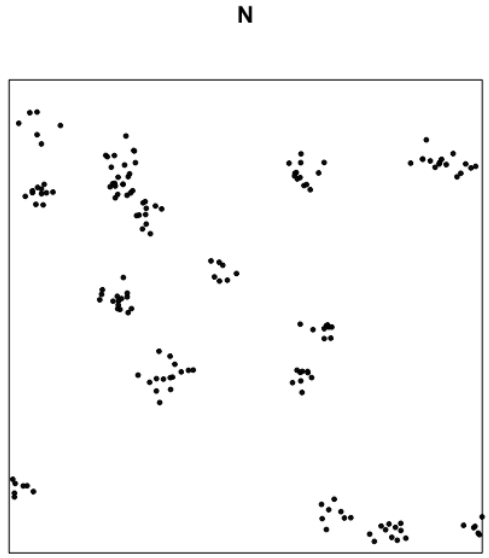
```
> thomas.estK(N)
Minimum contrast fit (object of class "minconfit")
Model: Thomas process
Fitted by matching theoretical K function to Kest(N)

Internal parameters fitted by minimum contrast ($par):
       kappa       sigma2
1.996230e+01 3.640801e-04

Fitted cluster parameters:
      kappa       scale
19.96229948  0.01908088

Converged successfully after 441 function evaluations

Starting values of parameters:
 kappa sigma2
     1      1
Domain of integration: [ 0 , 0.25 ]
Exponents: p= 2, q= 0.25
```



Regular processes

In this section we will introduce two related models for regular (inhibited) point processes. We will soon see why these are important in imaging. The first is the *Matérn I* process.

A point process N_I is called Matérn I if:

1. Events are first generated as a homogeneous Poisson process with intensity ρ.
2. There exists some fixed "hardcore" distance $\delta > 0$ such that all events within distance δ of another event are removed.
3. The remaining events form the point process N_I.

First order properties

Let us consider the first order properties of this process.

Proposition

Let N_I be a Matérn I process generated from a Poisson process N_0 with intensity ρ and hardcore distance δ. The intensity is given as

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

Proof.

The probability that an arbitrary event in N_0 is retained is the same as the probability that an arbitrary event as no other events within distance δ of it. This is equal to $\exp(-\rho\pi r^2)$. Therefore,

$$\begin{aligned}
\lambda(\boldsymbol{s})|\mathrm{d}\boldsymbol{s}| &= P\{N_I(\mathrm{d}\boldsymbol{s}) = 1\}|\mathrm{d}\boldsymbol{s}| \\
&= P\{N_I(\mathrm{d}\boldsymbol{s}) = 1 | N_0(\mathrm{d}\boldsymbol{s}) = 1\} P\{N_0(\mathrm{d}\boldsymbol{s}) = 1\} \\
&= \exp(-\rho\pi\delta^2)\rho|\mathrm{d}\boldsymbol{s}|.
\end{aligned}$$

Second-order properties

Proposition

Let N_I be a Matérn I process generated from a Poisson process N_0 with intensity ρ and hardcore distance δ. The second order intensity is given as

$$\gamma(r) = \begin{cases} \rho^2 \exp\{-\rho U_\delta(r)\} & r \geq \delta \\ 0 & r < \delta \end{cases}$$

where $U_\delta(r)$ is the area of the union of two discs of distance radius δ, separated by r.

Second-order properties

Proof.

$$\begin{aligned}
\gamma(\boldsymbol{s},\boldsymbol{u})|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}| &= P\{N_I(\mathrm{d}\boldsymbol{s})=1, N_I(\mathrm{d}\boldsymbol{u})=1\} \\
&= P\{N_I(\mathrm{d}\boldsymbol{s})=1, N_I(\mathrm{d}\boldsymbol{u})=1|N_0(\mathrm{d}\boldsymbol{s})=1, N_0(\mathrm{d}\boldsymbol{u})=1\}P\{N_0(\mathrm{d}\boldsymbol{s})=1, N_0(\mathrm{d}\boldsymbol{u})=1\} \\
&= P\{N_I(\mathrm{d}\boldsymbol{s})=1, N_I(\mathrm{d}\boldsymbol{u})=1|N_0(\mathrm{d}\boldsymbol{s})=1, N_0(\mathrm{d}\boldsymbol{u})=1\}P\{N_0(\mathrm{d}\boldsymbol{s})=1, N_0(\mathrm{d}\boldsymbol{u})=1\} \\
&= P\{N_I(\mathrm{d}\boldsymbol{s})=1, N_I(\mathrm{d}\boldsymbol{u})=1|N_0(\mathrm{d}\boldsymbol{s})=1, N_0(\mathrm{d}\boldsymbol{u})=1\} \quad \rho^2 \quad |\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|
\end{aligned}$$

Let us consider three cases for assessing $P\{N_I(\mathrm{d}\boldsymbol{s})=1, N_I(\mathrm{d}\boldsymbol{s})=1|N_0(\mathrm{d}\boldsymbol{s})=1, N_0(\mathrm{d}\boldsymbol{s})=1\}$:

1. If $r = \|\boldsymbol{s}-\boldsymbol{u}\| < \delta$ then they will definitely both be deleted, therefore

$$P\{N_I(\mathrm{d}\boldsymbol{s})=1, N_I(\mathrm{d}\boldsymbol{s})=1|N_0(\mathrm{d}\boldsymbol{s})=1, N_0(\mathrm{d}\boldsymbol{s})=1\} = 0.$$

2. If $r = \|\boldsymbol{s}-\boldsymbol{u}\| > 2\delta$ then we just need to know the probability there are no further events in either disc. This is given as

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

3. if $\delta \leq r = \|\boldsymbol{s}-\boldsymbol{u}\| < 2\delta$, then we need to know the probability that there are no further events in the area $U_\delta(r)$. This is given as $\exp\{-\rho U_\delta(r)\}$.

The result follows. □

Pair-correlation function and Ripley's K-function for Matern I

Following from this result we have the following identity for the pair correlation function takes the form

$$g(r) = \begin{cases} \exp[\rho\{2\pi r^2 - U_\delta(r)\}] & r \geq \delta \\ 0 & r < \delta \end{cases}$$

and the Ripley's K-function is given as

$$K(r) = \begin{cases} 2\pi \int_\delta^r r' \exp[\rho\{2\pi r'^2 - U_\delta(r')\}] \mathrm{d}r' & r \geq \delta \\ 0 & r < \delta \end{cases}$$

Figure: $g(r)$, $K(r)$ and $L(r) - r$ for Matérn I process

Matérn II

A second useful model that we will consider is the Matérn II process.

Definition (Matérn II)

A point process N_{II} is called Matérn I if:

1. Events are first generated as a homogeneous Poisson process with intensity ρ.
2. Independently mark the events $\{\boldsymbol{s}_1, \boldsymbol{s}_2, ...\}$ with numbers $\{Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2), ...\}$ from any absolutely continuous distribution F.
3. An event $\boldsymbol{s}$ of N_0 is deleted if there exists another event $\boldsymbol{u}$ with $\|\boldsymbol{s} - \boldsymbol{u}\| < \delta$ and $Z(\boldsymbol{u}) < Z(\boldsymbol{s})$.
4. The remaining events form the point process N_{II}.

First order properties of Matern II

Proposition

Let N_{II} be a Matérn II process generated from a Poisson process N_0 with intensity ρ and hardcore distance δ. The intensity is given as

$$\lambda = \frac{1 - \exp(-\rho\pi\delta^2)}{\pi r^2}.$$

Second-order properties of Matern II

Proposition

Let N_{II} be a Matérn II process generated from a Poisson process N_0 with intensity ρ and hardcore distance δ. The second order intensity is given as

$$\gamma(r) = \begin{cases} \frac{2U_\delta(r)1-\{\exp(-\rho\pi\delta^2)\}-2\pi\delta^2[1-\exp\{-\rho U_\delta(r)\}]}{\rho^2\pi\delta^2 U_\delta(r)\{U_\delta(r)-\pi\delta^2\}} & r \geq \delta \\ 0 & r < \delta \end{cases}$$

where $U_\delta(r)$ is the area of the union of two discs of distance radius δ, separated by r.

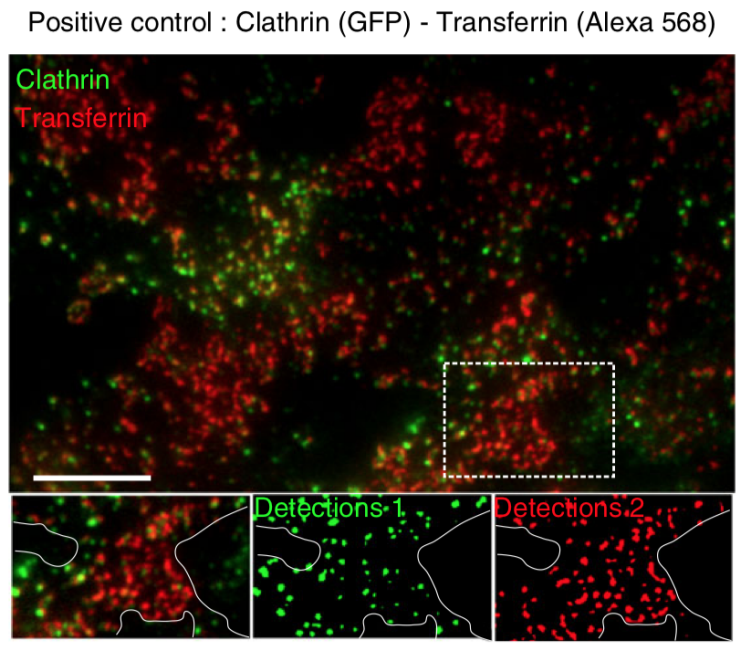
Multivariate Point Processes and Point Patterns

Multivariate Point Processes and Point Patterns

We now turn our attention to the case where we have two (or more) spatial point processes.

Some examples of this include

- two different types of protein molecule in an image. One is tagged with a green fluorophore and the other red.
- two different species of plant in a forest.
- the location of nuclear power plants and the location of leukaemia sufferers.



Lagache et al, Nature Communications (2018)9:698

Multivariate Point Processes and Point Patterns

- One immediate question we might ask ourselves is: "are these two processes independent of each other, or is there inter-dependency?"
- Another way to say this is, where I see an event of one process, am I more or less likely to see an event of another process than I would if they were completely independent.
- Before introducing a method for dealing with these types of data, we will introduce some theory on bivariate point processes.

Theoretical framework for bivariate point processes

- Let us consider two point processes. We will label these N_1 and N_2.
- Process N_1 has intensity $\lambda_1(\boldsymbol{s})$ and second order intensity $\gamma_1(\boldsymbol{s}, \boldsymbol{u})$ and process N_2 has intensity $\lambda_2(\boldsymbol{s})$ and second order intensity $\gamma_2(\boldsymbol{s}, \boldsymbol{u})$.

Cross-intensity

Definition (Cross-intensity)

The cross intensity between N_1 and N_2 at points $\boldsymbol{s}$ and $\boldsymbol{u}$ is defined as

$$\gamma_{12}(\boldsymbol{s}, \boldsymbol{u}) = \lim_{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}| \to 0} \frac{E\{N_1(\mathrm{d}\boldsymbol{s})N_2(\mathrm{d}\boldsymbol{u})\}}{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|}.$$

Cross-covariance intensity

From this we can further define the cross covariance intensity of N_1 and N_2.

Definition (Cross-covariance intensity)

The cross covariance intensity between N_1 and N_2 at points $\boldsymbol{s}$ and $\boldsymbol{u}$ is defined as

$$c_{12}(\boldsymbol{s}, \boldsymbol{u}) = \gamma_{12}(\boldsymbol{s}, \boldsymbol{u}) - \lambda_1(\boldsymbol{s})\lambda_2(\boldsymbol{u}).$$

- This function provides a measure of dependency between the two point processes.
- Let us consider the case where the two processes are independent. Recall, that for two independent random variables X and Y we have $E\{XY\} = E\{X\}E\{Y\}$.
- With this in mind, we have

$$\begin{aligned}
\gamma_{12}(\boldsymbol{s},\boldsymbol{u}) &= \lim_{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|\to 0} \frac{E\{N_1(\mathrm{d}\boldsymbol{s})N_2(\mathrm{d}\boldsymbol{u})\}}{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|} \\
&= \lim_{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|\to 0} \frac{E\{N_1(\mathrm{d}\boldsymbol{s})\}E\{N_2(\mathrm{d}\boldsymbol{u})\}}{|\mathrm{d}\boldsymbol{s}||\mathrm{d}\boldsymbol{u}|} \\
&= \lim_{|\mathrm{d}\boldsymbol{s}|\to 0} \frac{E\{N_1(\mathrm{d}\boldsymbol{s})\}}{|\mathrm{d}\boldsymbol{s}|} \lim_{|\mathrm{d}\boldsymbol{u}|\to 0} \frac{\{N_2(\mathrm{d}\boldsymbol{u})\}}{|\mathrm{d}\boldsymbol{u}|} \\
&= \lambda_1(\boldsymbol{s})\lambda_2(\boldsymbol{u})
\end{aligned}$$

and following from this the cross-covariance intensity is

$$c_{12}(\boldsymbol{s},\boldsymbol{u}) = 0$$

Joint stationarity and isotropy

The concepts of stationarity and isotropy also extend to the bivariate setting.

Definition (Joint stationary and isotropic)

Let N_1 and N_2 be a pair of spatial point processes. We say they are jointly stationary and isotropic if they are both individually stationary and isotropic and the cross-intensity is a function of $r = \|\boldsymbol{s} - \boldsymbol{u}\|$ only.

In this setting, the cross-intensity and cross-covariance intensity can be represented as $\gamma_{12}(r)$ and $c_{12}(r)$, respectively.

Cross-K-function

The cross-K-function provides a natural extension to Ripley's K-function.

Definition (cross-K-function)

Let N_1 and N_2 be a pair of jointly stationary and isotropic processes, and let $N_{ij}(r)$ represent the random number of type j events within a distance r of an arbitrarily chosen type i event ($i, j = 1, 2$, not including that event). The cross-K-function is defined as

$$\begin{aligned} K_{ij}(r) &= \lambda_j^{-1} E\{\text{number of type } j \text{ events} \\ &\qquad \text{within a distance } r \text{ of an arbitrary type } i \text{ event}\} \\ &= \lambda_j^{-1} E\{N_{ij}(r)\}. \end{aligned}$$

Cross-K-function for independence

- Let us consider the case where N_1 and N_2 are *independent* jointly stationary and isotropic processes. It will be the case that

$$E\{\text{number of type 2 events within a distance } r \text{ of an arbitrary type 1 event}\} = \lambda_2 \pi r^2$$

and therefore

$$K_{12}(r) = \pi r^2.$$

- Therefore, irrespective of the type of process N_1 and N_2 are, if they are independent then $K_{12}(r) = \pi r^2$. This therefore means that we can use $K_{12}(r)$ to test for independence.

Question

Let N_1 and N_2 be independent processes with intensities $\lambda_1 = 5$ and $\lambda_2 = 2$, respectively. What is

$E\{\text{number of type 2 events within a distance 3 of an arbitrary type 1 event}\}$?

(a) 18π.

(b) 45π.

(c) 9π.

(d) π.

(e) 4.5π

- Note also that analogously to the standard Ripley's K-function, we have the relationship

$$K_{12}(r) = \frac{2\pi}{\lambda_1 \lambda_2} \int_0^r r' \gamma_{12}(r') \mathrm{d}r'.$$

- We also make the observation that $\gamma_{12}(r) = \gamma_{21}(r)$ (this is because for random variables X and Y it is true that $E(XY) = E(YX)$). Therefore it follows that $K_{12}(r) = K_{21}(r)$.

Estimating the cross-K-function

- Estimating the cross K-function is very similar to how estimated the Ripley's K-function, however, this time, we measure distances between pairs of events of different types.

- Let u_{ij} be the distance between the ith event of type 1 and jth event of type 2, w_{ij} are the weights as used before, and the number of events of type 1 and type 2 are n_1 and n_2, respectively.

We can construct an estimator of $K_{12}(r)$ as

$$\hat{\lambda}_2 \tilde{K}_{12}(r) = n_1^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij} I(u_{ij} \leq r)$$

but we can also construct one as

$$\hat{\lambda}_1 \tilde{K}_{21}(r) = n_2^{-1} \sum_{i=1}^{n_2} \sum_{j=1}^{n_1} w_{ji} I(u_{ij} \leq r).$$

We can therefore average the two estimates to give

$$\begin{aligned} \hat{K}_{12}(r) &= (n_1 n_2)^{-1} |A| \left\{ n_1 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij} I(u_{ij} \leq r) \right. \\ &\qquad \left. + n_2 \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} w_{ij} I(u_{ij} \leq r) \right\} / (n_1 + n_2) \\ &= (n_1 n_2)^{-1} |A| \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij}^* I(u_{ij} < r) \end{aligned}$$

Question

where w_{ij}^* equals...?

(a) $n_1 w_{ij} + n_2 w_{ji}$.

(b) $(w_{ij} + w_{ji})/(n_1 + n_2)$.

(c) $w_{ij} + w_{ji}$.

(d) $(n_1 w_{ij} + n_2 w_{ji})/(n_1 + n_2)$.

(e) w_{ij}.

Testing for independence

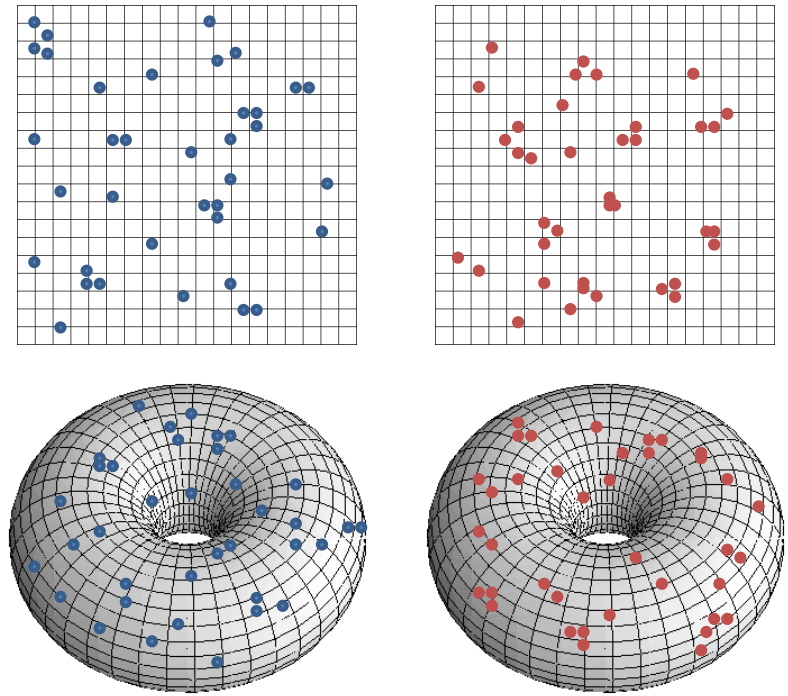
- Using the K-function to test

$$H_0 : N_1 \text{ and } N_2 \text{ are independent.}$$

  is somewhat troublesome.
- This is because the distribution of $\hat{K}_{12}(r)$ is dependent on the type of processes that are being tested. E.g. the distribution of $\hat{K}_{12}(r)$ when N_1 and N_2 are Poisson is different to when they are Thomas.
- Monte Carlo methods exist.

Toroidal Shift

- Toroidal shift method (Lotwick et al '82) is used to find critical value for a particular significance level.



- By performing random rotations and shifts, one is able to induce independence and and Monte Carlo sample from the null.