Point Processes 3: Poisson Models

Michael Noonan

DATA 589: Spatial Statistics

Table of contents

Review

Review

Last lecture we saw how second moment descriptive statistics (Morisita's index, Ripley's K-function, the g-function) can help us further understand whether there are any correlations between points in a point process (clustering, avoidance, etc...).

These metrics are informative additions to first moment measures, but are correlative in nature and do not provide us with information on the underlying cause.

I mentioned that descriptive statistics are a great place to start, but that to fully understand a point process we need to be able to model it.

Today we will focus on how to formally model point processes using 'Poisson point process' models.

Applied Points Pattern Analysis

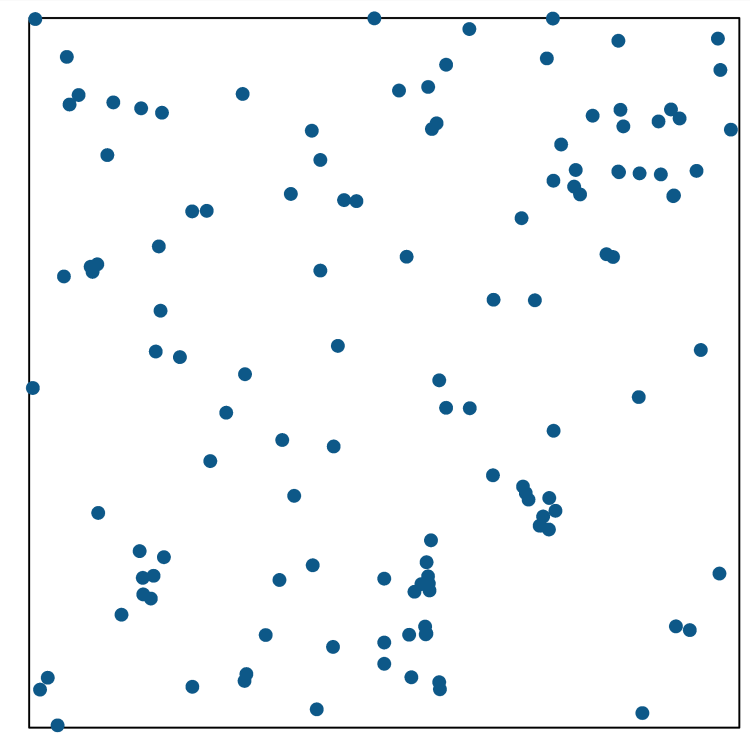
Data driven pathology

Jones-Todd *et al.* (2019) used machine learning and point pattern analyses to characterise the spatial distribution tumor cells in cancer patients.

Different spatial patterns could then be used to inform treatment options.

Motivation

Describing point processes

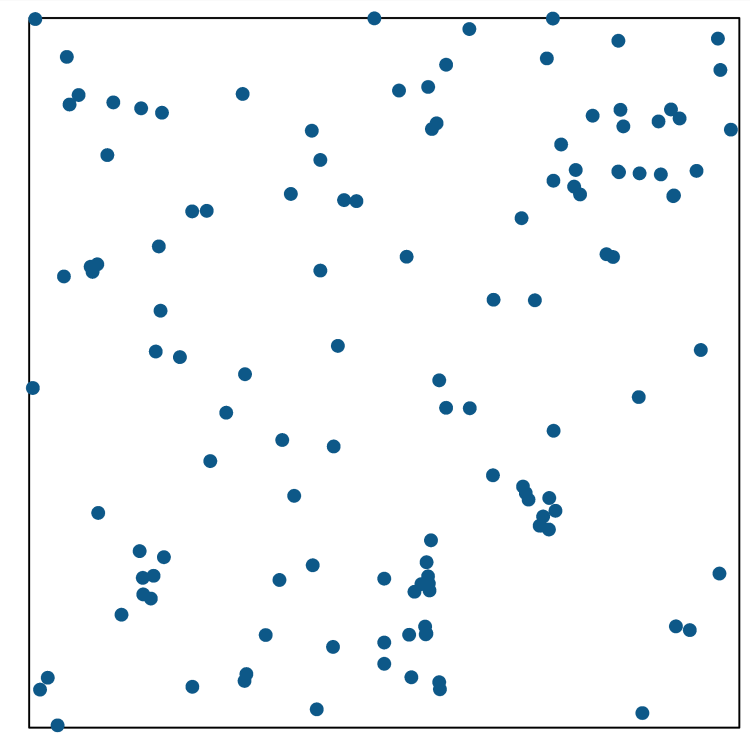


We have been covering methods for describing the spatial arrangement of points.

E.g., Are the points uniform? Does the intensity depend on a covariate? Are they clustered? and so on...

In asking these questions were are not really interested in the points *per se*, but in the **process** that generated the points.

Describing point processes cont.



Describing a collection of points within a window can provide valuable insight, but the findings are not generalisable (i.e., don't extend beyond the sampling window).

In order to be able to make general statements about how we expect points to be arranged, we need to model our system.

... and in order to model our system we need a formal framework for what these models should look like.

Defining the Point Process

Point processes definitions

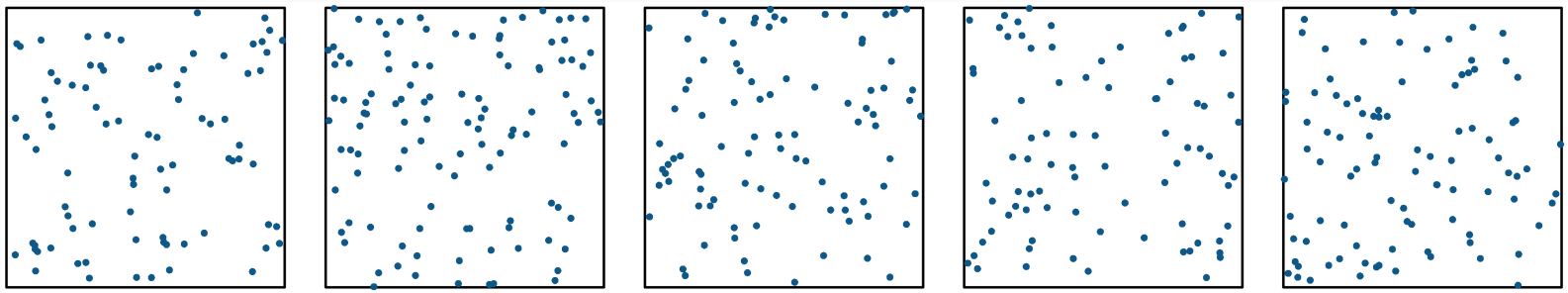
A point process, $\mathbf{X}$, is a random mechanism whose outcome is a point pattern.

A point pattern is a set $\mathbf{x} = \{x_1, x_2, ...\}$ of points in a two-dimensional space, which has a finite number of points in any bounded region B (i.e., $n(\mathbf{x} \cap B)$ is finite).

For any bounded region B, the number of points $n(\mathbf{x} \cap B)$ is a well-defined random variable.

Point processes derivation

We begin by considering a process that results in Complete Spatial Randomness (CSR).



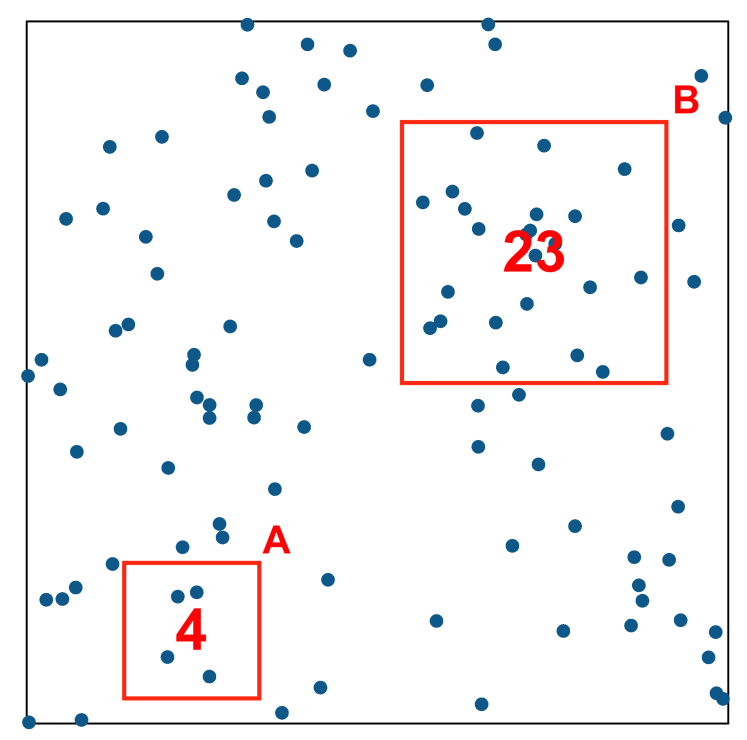
CSR implies:

1. **homogeneity** (points have no preference for any particular location), and
2. **independence** (knowing something about the number of points in one region of space provides no information on the number of points in other regions).

Homogeneity

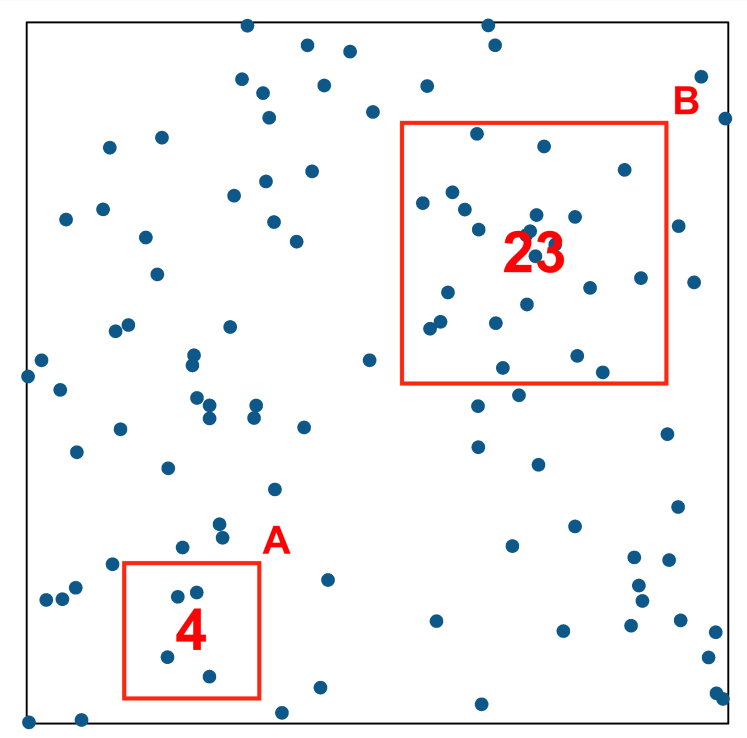
Under an assumption of **homogeneity**, the expected number of points falling in any region B is proportional to its area $|B|$:

$$\mathbb{E}n(\mathbf{x} \cap B) = \lambda|B|$$



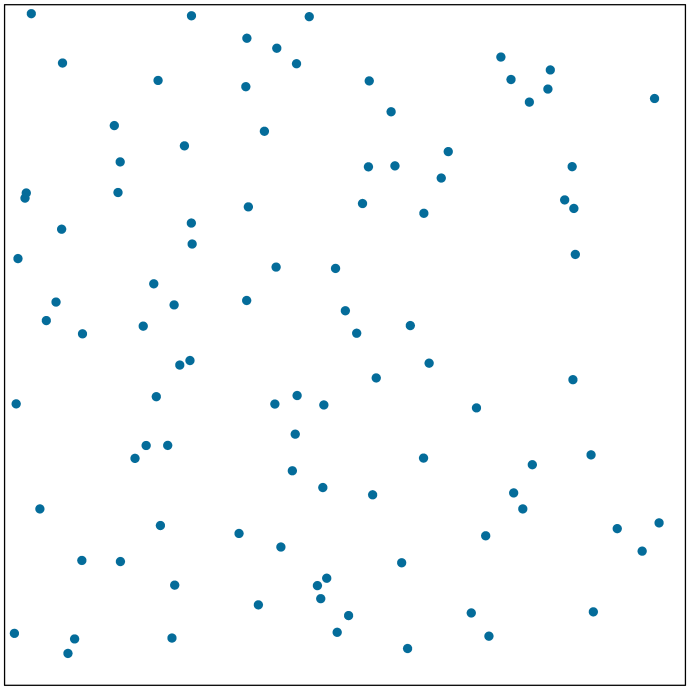
Independence

Under an assumption of **independence**, $n(\mathbf{x} \cap A)$ provides no information on $n(\mathbf{x} \cap B)$, which implies that the number of points falling in test regions are independent random variables.



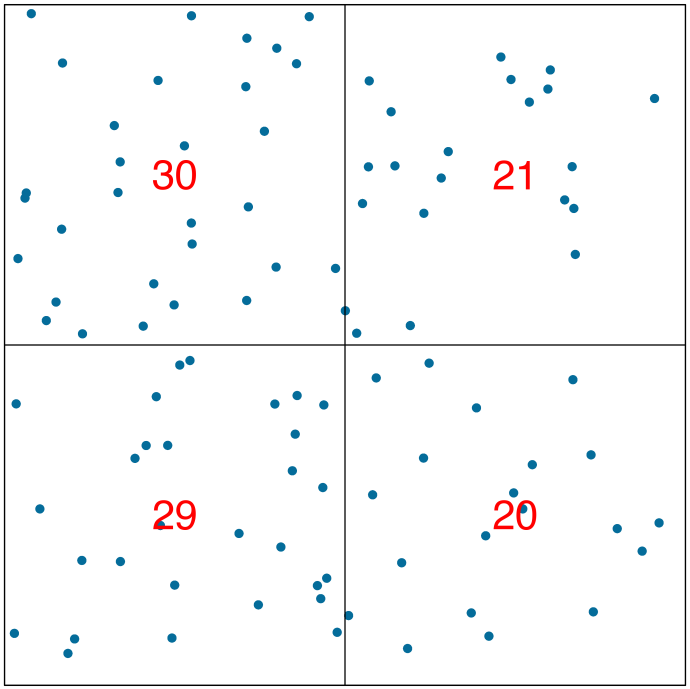
Independence cont.

The property of independence holds for regions of any shape and/or size... so taking finer and finer subdivisions results in more and more independent random variables.



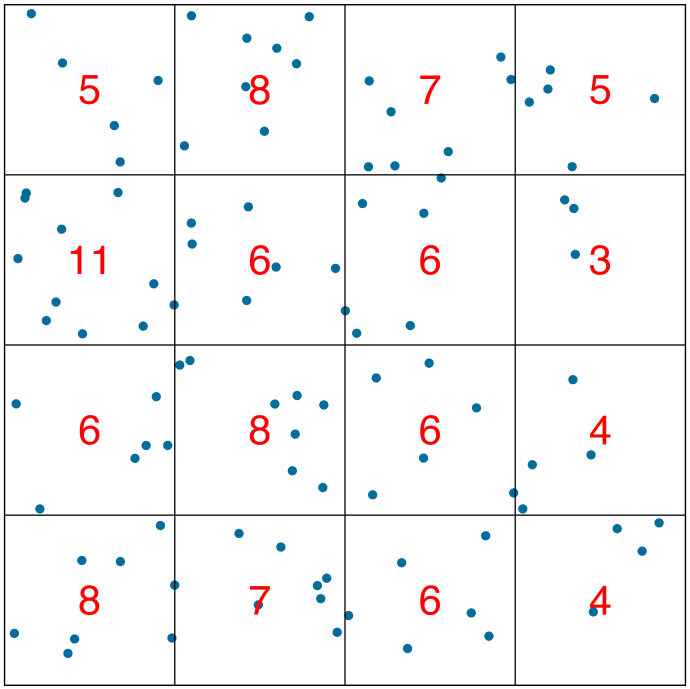
Independence cont.

The property of independence holds for regions of any shape and/or size... so taking finer and finer subdivisions results in more and more independent random variables.



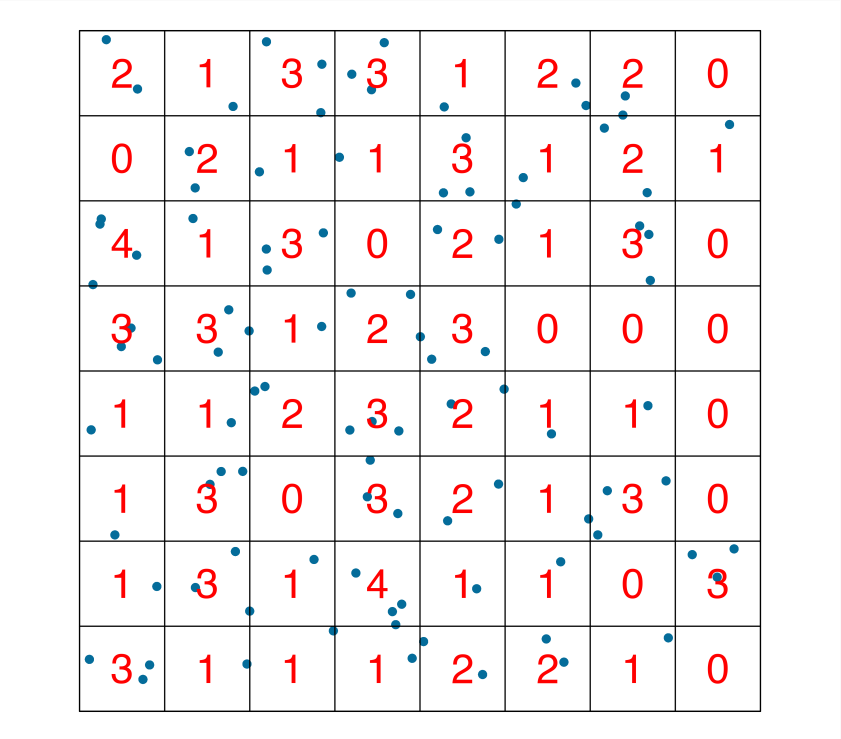
Independence cont.

The property of independence holds for regions of any shape and/or size... so taking finer and finer subdivisions results in more and more independent random variables.



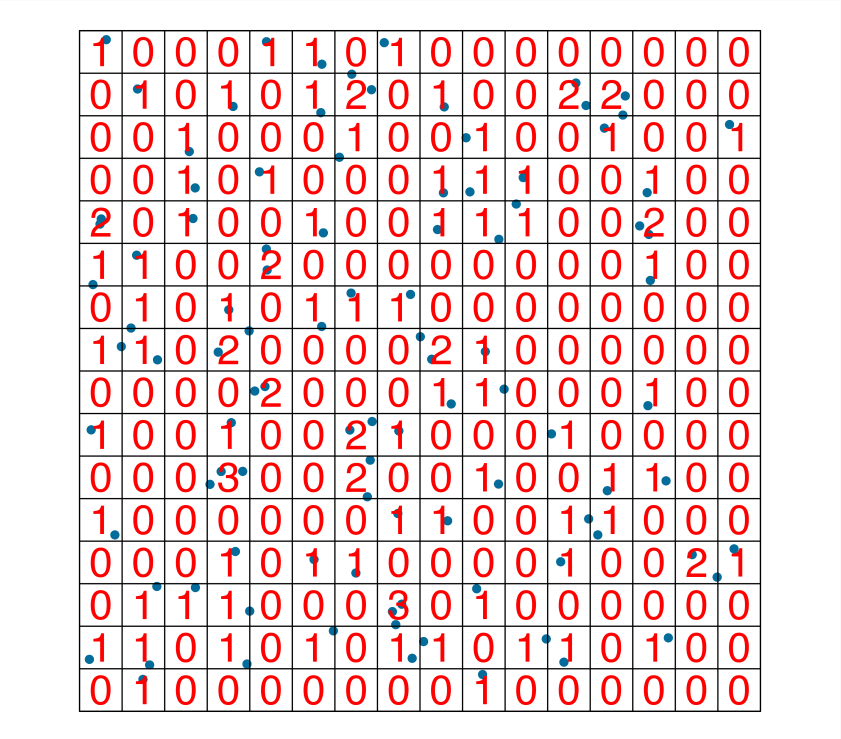
Independence cont.

The property of independence holds for regions of any shape and/or size... so taking finer and finer subdivisions results in more and more independent random variables.

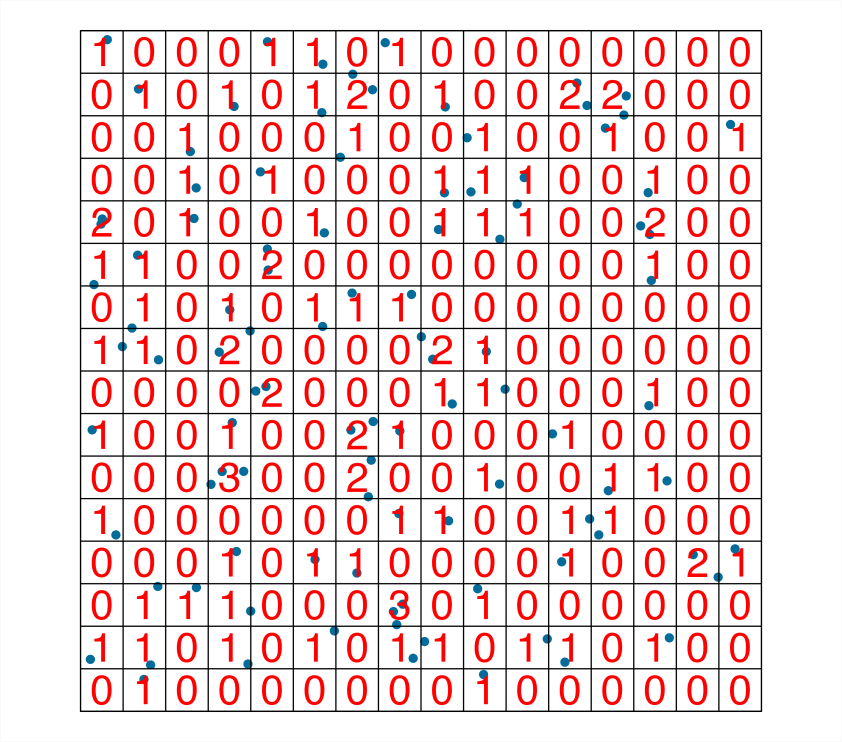


Independence cont.

The property of independence holds for regions of any shape and/or size... so taking finer and finer subdivisions results in more and more independent random variables.



Describing point processes cont.



When the size of the squares is extremely small, most will be empty and there is a negligible probability that a square will have >1 point.

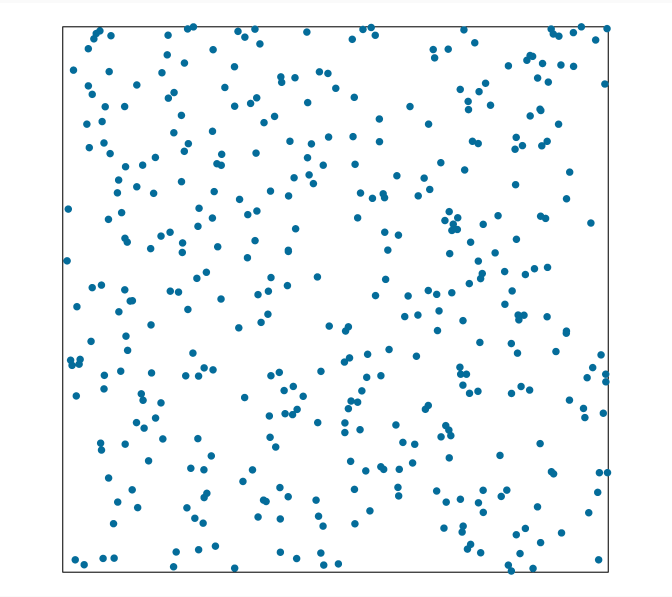
This implies that $n(\mathbf{X} \cap B)$ is the number of successes in a large number of independent trials, which implies that $n(\mathbf{X} \cap B)$ follows a Poisson distribution.

Because $\mathbb{E}n(\mathbf{x} \cap B) = \lambda|B|$, this means that $n(\mathbf{X} \cap B)$ is a Poisson distributed random variable with mean $\lambda|B|$.

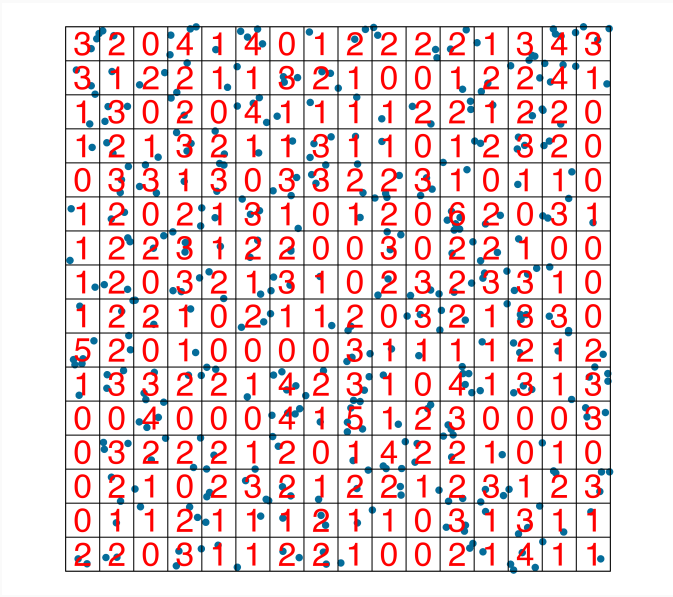
Poisson random variable

I just said that $n(\mathbf{X} \cap B)$ is a Poisson distributed random variable with mean $\lambda|B|$, but maybe you're not convinced?

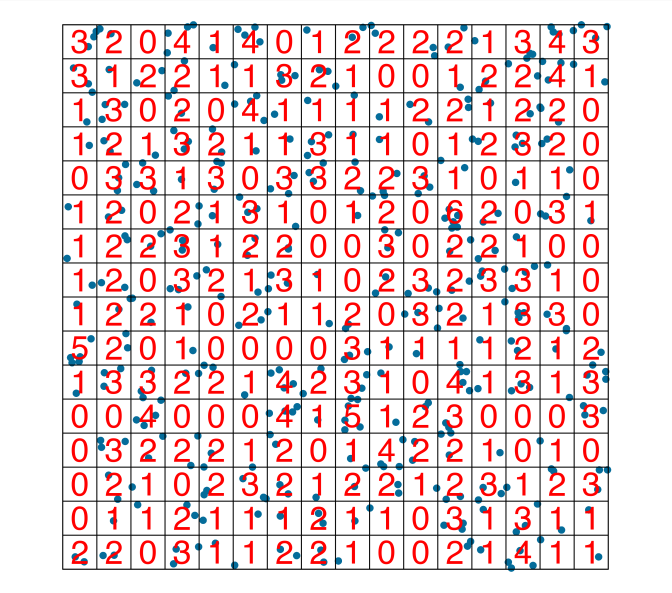
```
#Visualise a homogeneous point process
plot(ppp_example)
```



```
#Quadrat counting estimate of intensity
Q16by16 <- quadratcount(ppp_example,
                        nx = 16,
                        ny = 16)


plot(ppp_example)
plot(Q16by16, add = T)
```

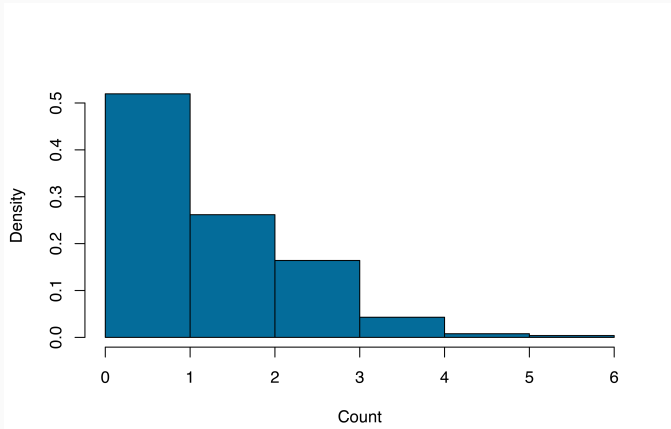


Poisson random variable cont.

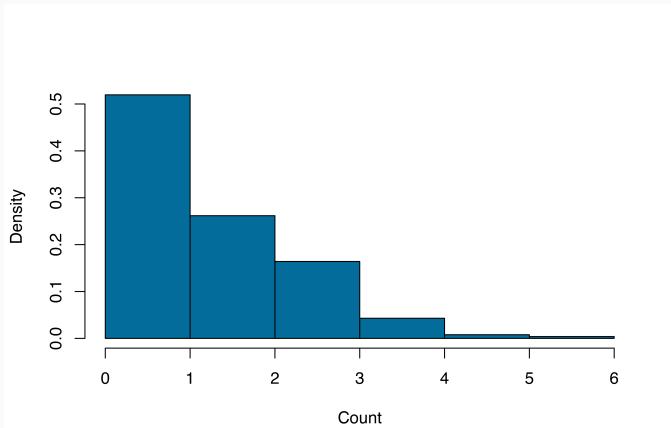



```
#Estimated intensity from quadrat counts
lambda_u <- mean(Q16by16)

lambda_u

[1] 1.5625

#Generate Poisson dist. vals. around lambda_u
R_Poisson <- rpois(n = 16^2,
                   lambda = mean(Q16by16))
```

```
#Visualise a histogram of the quadrat counts
hist(Q16by16)
```

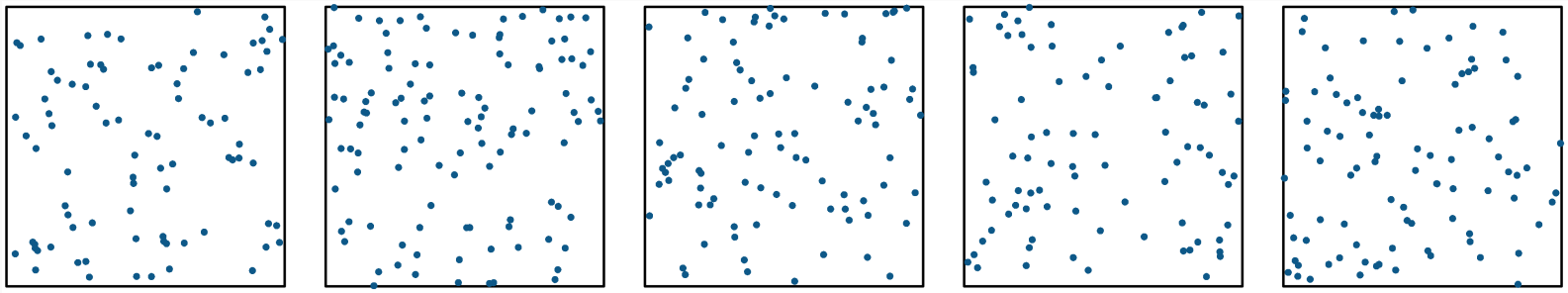



```
#Visualise a histogram of the Poisson values
hist(R_Poisson)
```



The Poisson Point Process

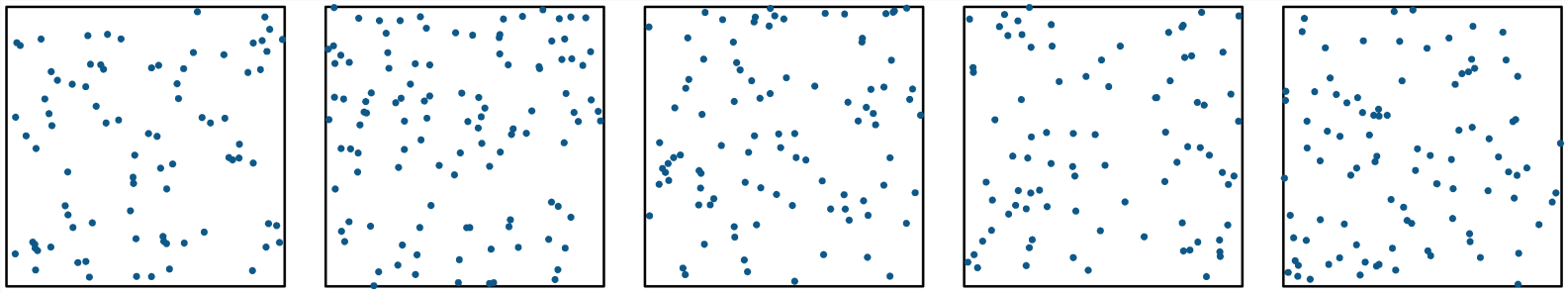
Poisson point processes



We now see that CSR implies:

1. **Homogeneity:** (points have no preference for any particular location).
2. **Independence:** (knowing something about the number of points in one region of space provides no information on the number of points in other regions).
3. **Poisson distribution:** $n(\mathbf{X} \cap B)$ follows a Poisson distribution.

Poisson point processes cont.



We are therefore dealing with a Poisson point processes that is described by an intensity function $\lambda(u)$

...which implies we need to build a statistical model that estimates $\lambda(u)$ (i.e., $\lambda(u) =$ some function of u).

Inhomogeneous Poisson point processes

If a point processes is homogeneous, $\lambda(u)$ is constant in space and defined by a 'baseline' intensity function.

$$\lambda(u) = \alpha$$

where α is an unknown baseline intensity that must be estimated.

If a point is processes is inhomogeneous, $\lambda(u)$ is not constant in space but rather a function of some covariate(s)

$$\lambda(u) = \alpha + \beta Z(u)$$

where α is the baseline intensity, $Z(u)$ is our spatial covariate, and β is our unknown covariate effect that must be estimated.

Inhomogeneous ppp cont.

$$\lambda(u) = \alpha + \beta Z(u)$$

defines an inhomogeneous point processes where $\lambda(u)$ is a function of some covariate(s)... but depending on the values of α and β, it is conceivably possible to obtain negative values for $\lambda(u)$, which is impossible.

As a fix, we add a log-link function and exponentiate our model

$$\lambda(u) = e^{\alpha+\beta_1 Z_1(u)+\beta_2 Z_2(u)...+\beta_i Z_i(u)}$$

Does this look familiar? (functionally similar to a Poisson GLM)

Inhomogeneous ppp cont.

Modelling an inhomogeneous Poisson point processes therefore means specifying the form of the model e.g.,

$$\lambda(u) = e^{\alpha+\beta_1 Z_1(u)+\beta_2 Z_2(u)...+\beta_i Z_i(u)}$$

...and estimating the unknown coefficients that best described the observed point pattern dataset

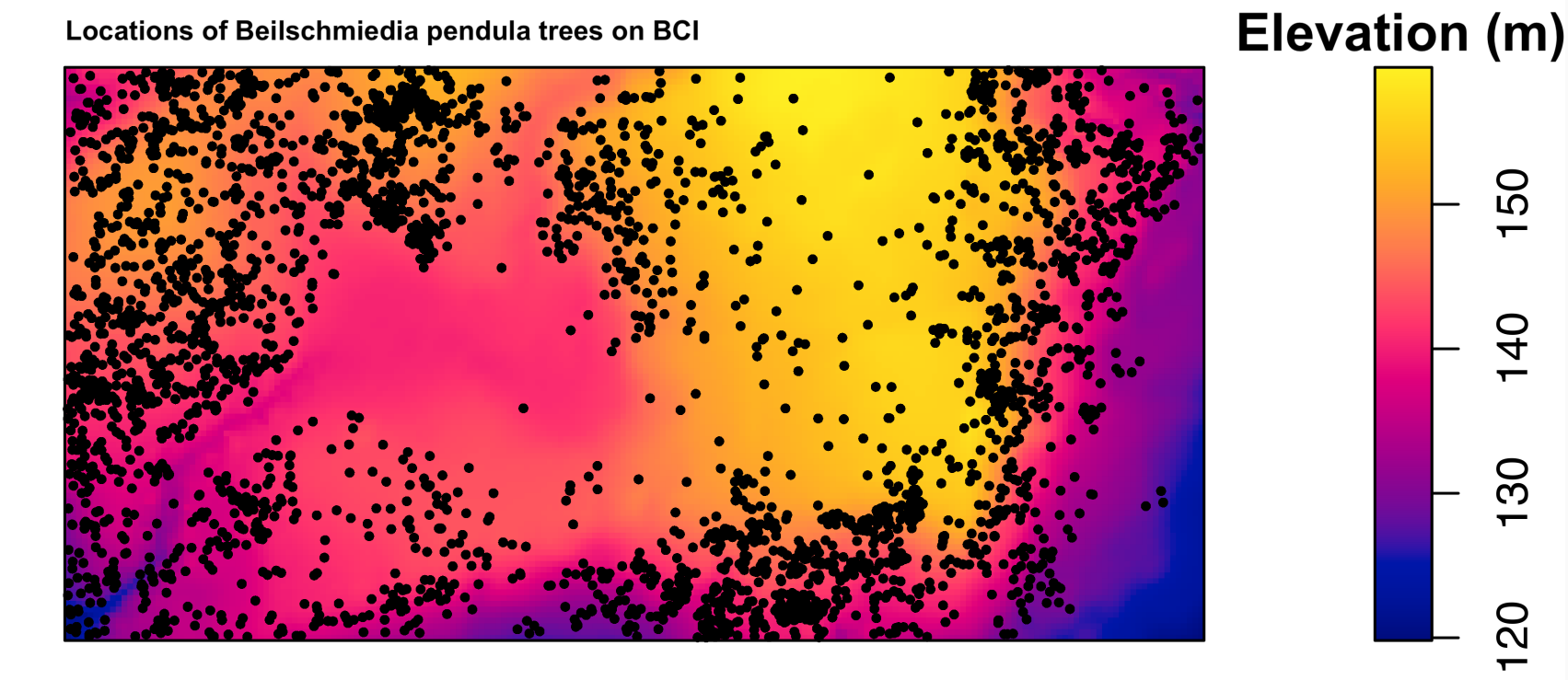
...and because we know what distribution $\lambda(u)$ should follow, we can approximate the likelihood function and estimate the parameters via some optimisation process (Ch. 9 in Baddeley *et al.* (2015) if you're interested).

Fitting Poisson models in R

Beilschmiedia pendula dataset

To demonstrate how to fit Poisson point processes to real data we will work with the *Beilschmiedia pendula* dataset.

Do we think the trees are homogeneous? Related to covariates?

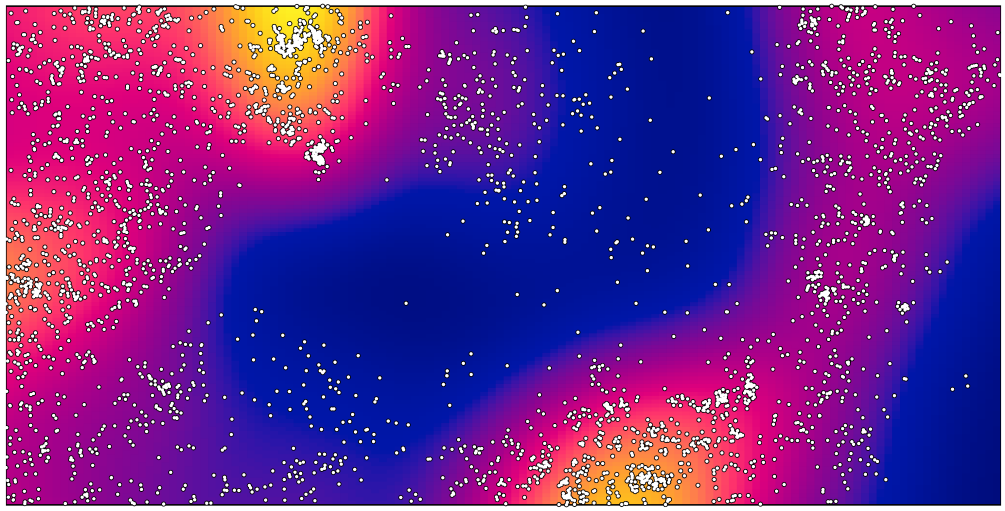


Source: `spatstat` package

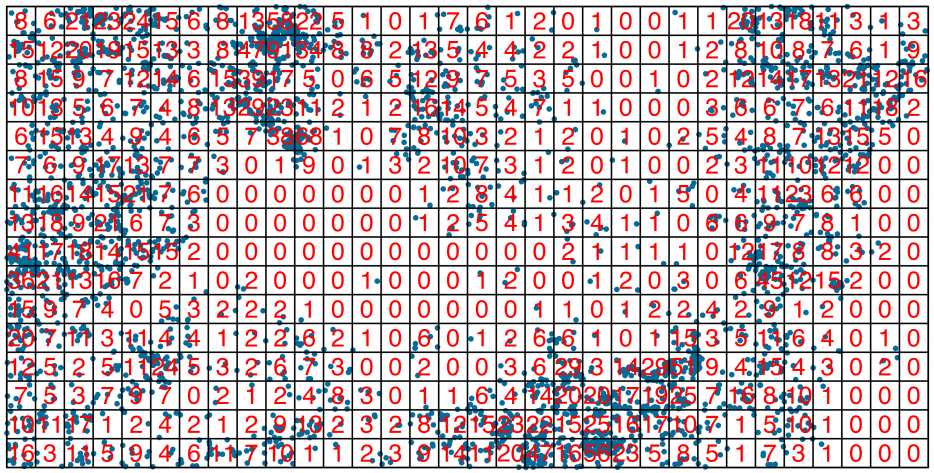
Start with the first moment

All analyses should start with describing the first moment. Do we think this point processes is homogeneous?

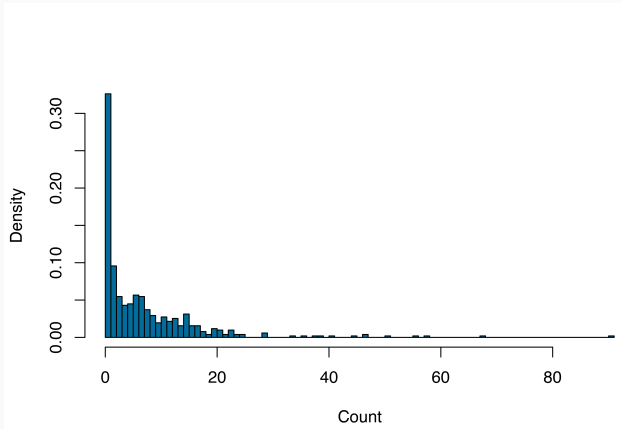
```
#Load in the data
data("bei")

#Estimate the intensity
lambda_hat <- density(bei)

#Visualise the first moment
plot(lambda_hat)
points(bei)
```



Homogeneous?

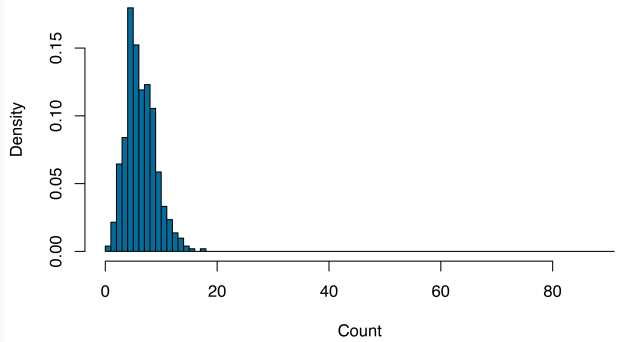


```
#Visualise a histogram of the quadrat counts
hist(Q16by16)
```




```
#Estimated intensity from quadrat counts
lambda_u <- mean(Qcount)

lambda_u

[1] 7.039062

#Generate Poisson dist. vals. around lambda_u
R_Poisson <- rpois(n = length(Qcount),
                   lambda = lambda_u)
```

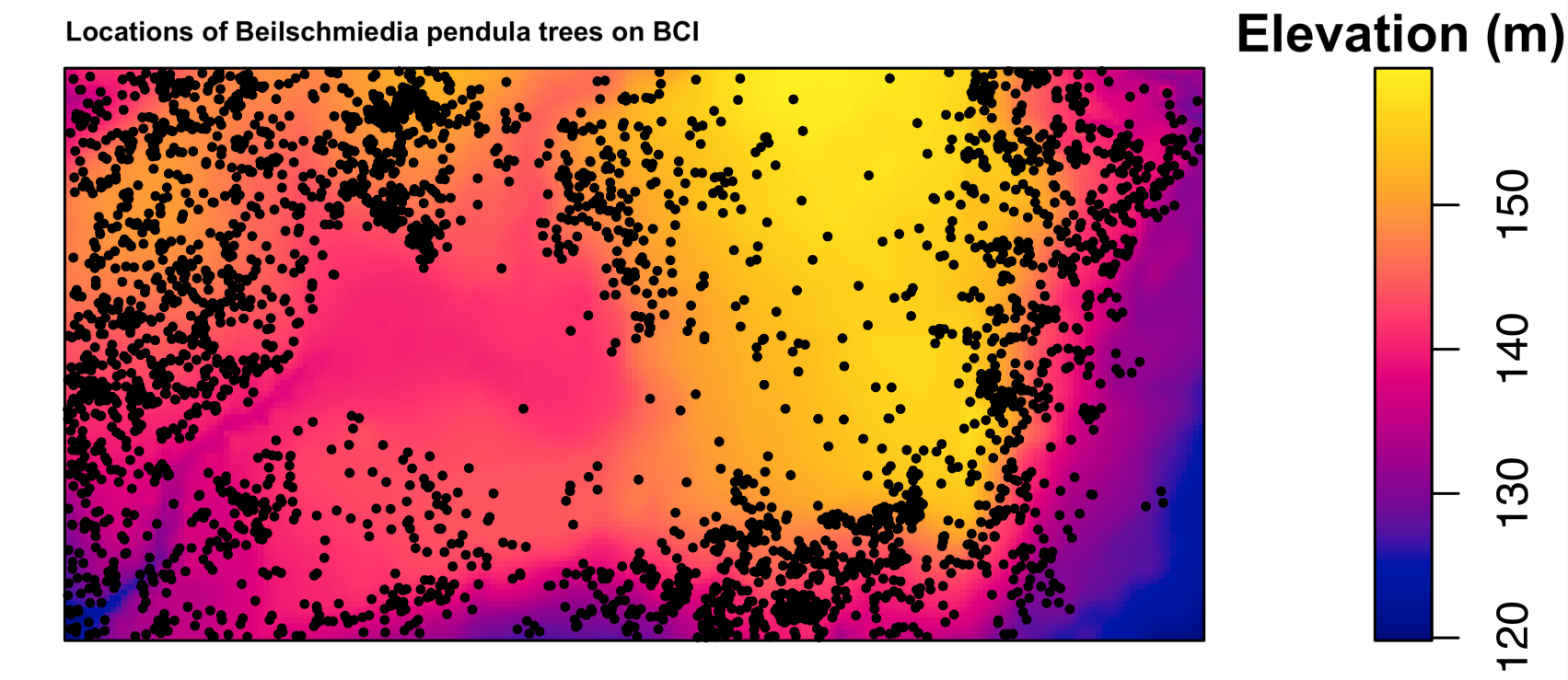
```
#Visualise a histogram of the Poisson values
hist(R_Poisson)
```



Inhomogeneous PPP

These data are clearly inhomogeneous, and a visual inspection suggests this may be related to an elevational preference.

This suggests we fit a model of the form $\lambda(u) = e^{\alpha + \beta_{\text{Elevation}} \times \text{Elevation}(u)}$



Source: `spatstat` package

Fitting the PPP

Fitting a PPP in R involves using the `spatstat::ppm()` function.

Model is of the form `ppm(X ~ trend, ...)`, where `X` is a point process and `trend` are our covariates (note: these are different objects in R env.).

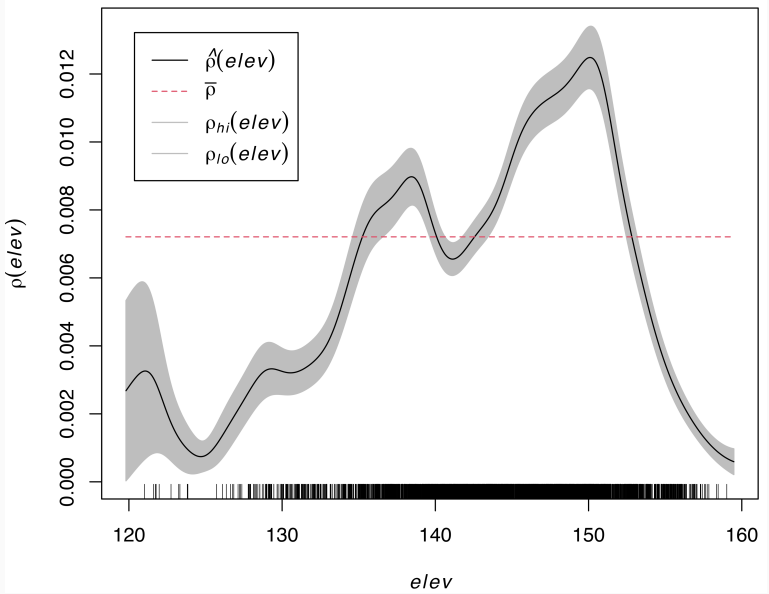
```
#Fit the PPP model
fit <- ppm(bei ~ elev, data = bei.extra)

---- Intensity: ----

Log intensity: ~elev
Model depends on external covariate    elev
Covariates provided:
  elev: im
  grad: im

Fitted trend coefficients:
(Intercept)        elev
-5.63919077  0.00488995

               Estimate        S.E.       CI95.lo      CI95.hi Ztest       Zval
(Intercept) -5.63919077 0.304565582 -6.2361283457 -5.042253203   *** -18.515522
elev         0.00488995 0.002102236  0.0007696438  0.009010256     *   2.326071
```

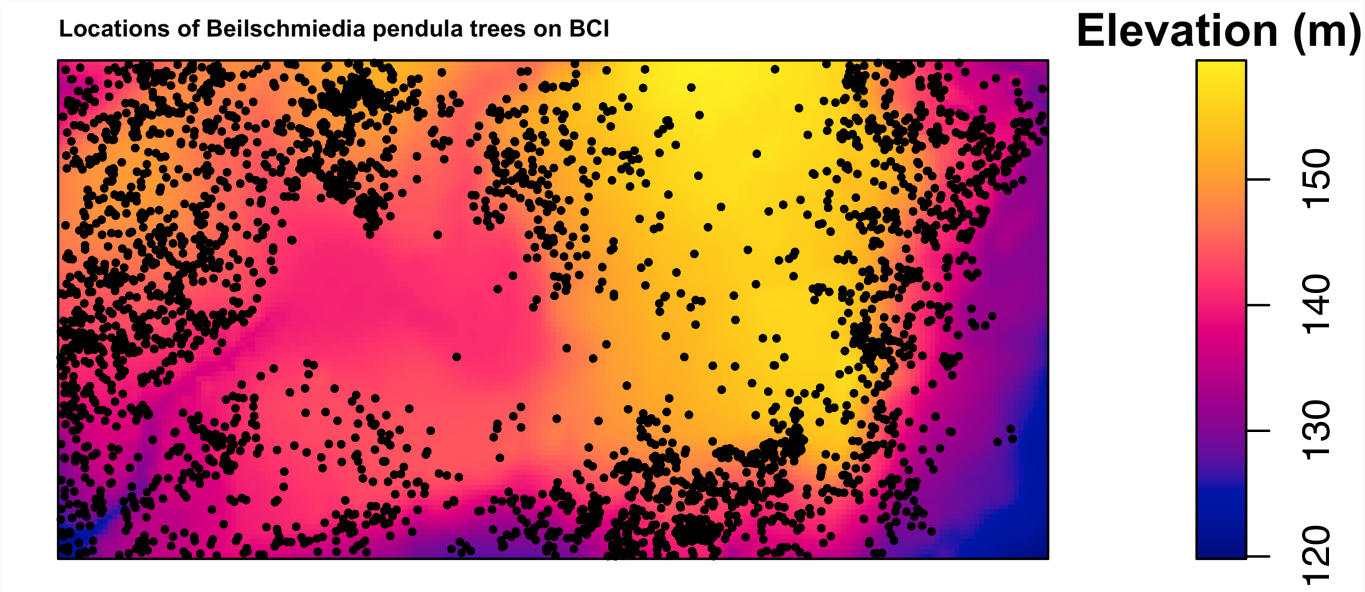
So our fitted model is $\lambda(u) = e^{-5.64+0.0049 \times \text{Elevation}(u)}$

Non-linearity?

...but if we remember from our lecture on intensity, it looked like the relationship with elevation was probably non-linear

...which means our linear model is probably inaccurate and will over-predict at high elevations.





PPP with multiple terms

The relationship looks quadratic, so we can just add a quadratic term to the model.

```
#Fit the PPP model
fit_quad <- ppm(bei ~ elev + I(elev^2), data = bei.extra)

---- Intensity: ----

Log intensity: ~elev + I(elev^2)
Model depends on external covariate    elev
Covariates provided:
  elev: im
  grad: im

Fitted trend coefficients:
  (Intercept)          elev     I(elev^2)
-1.379706e+02  1.847007e+00 -6.396003e-03

                 Estimate         S.E.       CI95.lo      CI95.hi Ztest      Zval
(Intercept) -1.379706e+02 6.7047209780 -1.511116e+02 -124.8295944   *** -20.57813
elev         1.847007e+00 0.0927883208  1.665145e+00    2.0288686   ***  19.90560
I(elev^2)   -6.396003e-03 0.0003207726 -7.024705e-03   -0.0057673   *** -19.93937
```

And our new fitted model is of the form:

$$\lambda(u) = e^{-138+1.85\times\text{Elevation}(u)-0.0064\times\text{Elevation}(u)^2}$$

PPP with multiple terms

On the surface, the more complex model makes sense given the observed patterns, but how do we know whether the additional complexity is supported by the data (i.e., are we overfitting?).

PPPs are amenable to standard model selection criteria (e.g., AIC or likelihood ratio tests).

We will only explore LRTs, but the selection process for PPP models is functionally identical to what you have seen for other models (e.g., GLMs).

Likelihood-Ratio Test

The likelihood-ratio test compares a pair of **nested** models based on the ratio of their likelihoods.

$$\lambda_{\mathsf{LR}} = -2 \ln \left[\frac{\mathcal{L}(\text{Reduced model})}{\mathcal{L}(\text{Full model})} \right]$$

The likelihood-ratio test statistic is often expressed as a difference between the log-likelihoods

$$\lambda_{\mathsf{LR}} = -2(\ln[\mathcal{L}(\text{Reduced})] - \ln[\mathcal{L}(\text{Full})])$$

The test statistic

So how does being able to quantify λ_{LR} help us identify the best model structure?

According to Wilks' theorem, as the sample size n approaches ∞, the test statistic λ_{LR} will be chi-squared distributed with degrees of freedom equal to difference in the number of parameters between the two models.

This implies that we can compare λ_{LR} to the χ^2 value corresponding to a desired statistical significance threshold (usually $\alpha = 0.05$) as an approximate statistical test.

LRT on PPP models

Nested PPP models are amenable to likelihood ratio tests.

```
#Conduct a likelihood ratio test on the quadratic term
anova(fit, fit_quad, test = "LR")

Analysis of Deviance Table

Model 1: ~elev   Poisson
Model 2: ~elev + I(elev^2)   Poisson
  Npar Df Deviance  Pr(>Chi)
1    2
2    3  1   536.05 < 2.2e-16 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
```

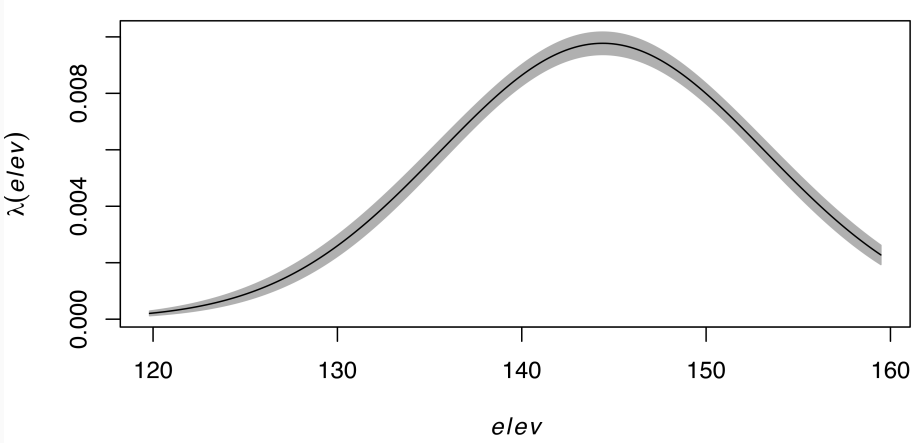
The p-value is tiny, so we reject the simple model in favour of the more complex quadratic form.

Model visualisation

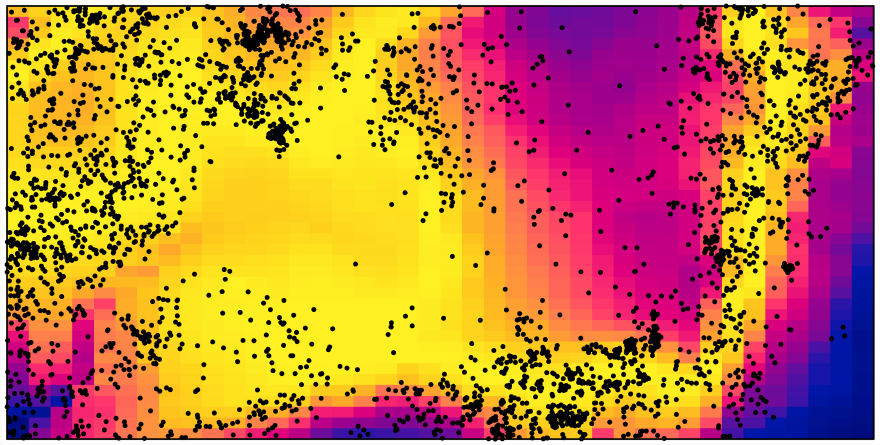
Seeing the summary output is useful, but perhaps not the easiest way to interpret the fitted model...

Visualisations help!

```
#Plot the elevation effect
plot(effectfun(fit_quad, "elev", se.fit = T))
```




```
#Plot the model predictions
plot(fit_quad)
```



Take home messages

We saw that for a homogeneous point process, the number of points falling in any region can be treated as Poisson distributed random variable.

This property allowed us to define a formal framework for modelling point processes, which in turn allows us to make general inference about our study system, and also make predictions from our fitted models.

We also saw that we can visualise fitted models and perform model selection to identify the best fit model for the data at hand.

...but (and importantly!) we didn't cover methods for validating our model, which we will focus on next lecture.

References

Baddeley, A., Rubak, E. & Turner, R. (2015). *Spatial point patterns: methodology and applications with R*. CRC press.

Jones-Todd, C.M., Caie, P., Illian, J.B., Stevenson, B.C., Savage, A., Harrison, D.J. & Bown, J.L. (2019). Identifying prognostic structural features in tissue sections of colon cancer patients using point pattern analysis. *Statistics in medicine*, 38, 1421–1441.