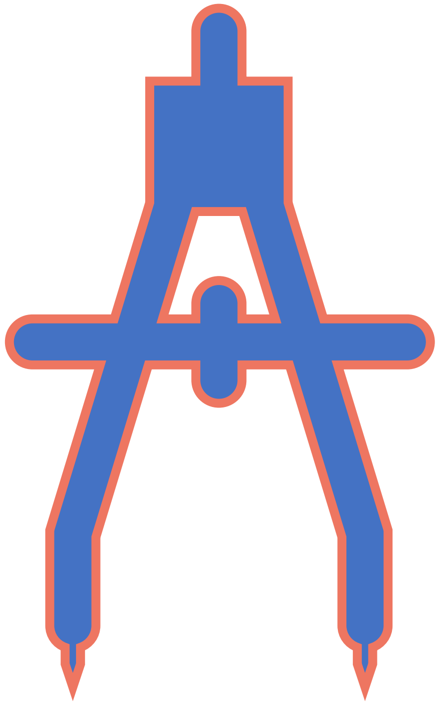
ECON6027 5D

Nearest Neighbour Distance Analysis (Clark and Evan’s Aggregation Index)

The approach

- The nearest neighbour approach compares the distances between nearest points, and distances that would be expected based on “chance”.
- Or simply measures the distance between an individual point and its nearest neighbour.
- Assumptions
  - Observation points represent a sample in a two- or more-dimensional Euclidean space.
  - Relationships between neighbouring points follow a Poisson distribution (null).
- Different ordered neighbour statistics (first-ordered, second-ordered, etc.) can be derived.



Compare...

Observed Distance:

Average distance between nearest neighbours, $\bar{r}$

Expected Distance: E(r)

HTTPS://EN.WIKIPEDIA.ORG/WIKI/POISSON_DISTRIBUTION

Theoretical derivation

- Let N be the sample size and A be the study area.
- The density of points is defined as:

$$\lambda = \frac{N}{A}$$

- For an arbitrary disk with radius r, the average density (aka intensity) per disk is,

$$\lambda \times \pi r^2$$

- Let the variable X be the number of neighbours within a radial distance of r.
- If we assume that variable X follows a homogeneous Poisson process (HPP), we can say,

$$X \sim Po(\lambda \pi r^2)$$

The *G*-function

- For a Poisson variable, the probability mass function is given by,

$$P(X = x) = \frac{e^{-\lambda\pi r^2}(\lambda\pi r^2)^x}{x!}$$

- Thus, the probability of an observation having no neighbours within a radial distance of r is,

$$P(X = 0) = e^{-\lambda\pi r^2}$$

- ***G*-function:** a cumulative distribution function (cdf) that generates the probability of having “no neighbours” within a radial distance of r,

$$G(r) = 1 - e^{-\lambda\pi r^2}$$

Note: $G(r) \to 0$ as $r \to 0$ and $G(r) \to 1$ as $r \to \infty$ as a valid cdf should.

The pdf of r

- The random variable r is the radial distance one can go before encountering the first neighbour.
- Thus, the analysis of the random variable r is focused on the analysis of "**second order effects**" in point patterns.
- The probability density function of r is given by taking the derivative of the cdf w.r.t. r,

$$f(r) = \frac{d}{dr}\left(1 - e^{-\lambda\pi r^2}\right) = 2\lambda\pi r e^{-\lambda\pi r^2}$$

Properties of *r*

- Expected value (mean) of *r*

$$E(r) = \mu_r = \frac{1}{2\sqrt{\lambda}}$$

- Variance of *r*

$$Var(r) = \sigma_r^2 = \frac{4 - \pi}{4\lambda\pi}$$

Inferences for *r*

- Standard error of $\bar{r}_N$:

$$s.e.(\bar{r}_N) \approx \frac{0.26136}{\sqrt{N\lambda}}$$

- Test statistic:

$$\frac{\bar{r}_N - E(r)}{s.e.(\bar{r}_N)} \sim N(0,1)$$

Activity A

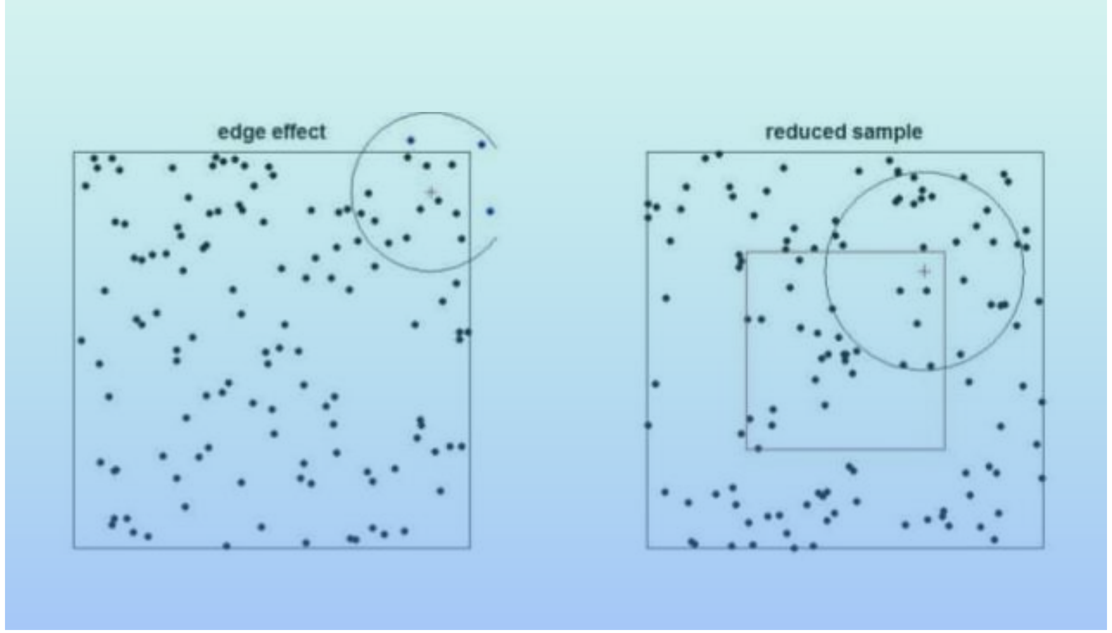
1. Prove the expressions for E(r) and Var(r) in slide 7.

(Hint: use integration by parts and Gaussian Integral)

2. Prove the expression for the standard error in slide 8.

(Hint: $s.e.(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$)

Clark & Evans Aggregation Index (1954)



- Clark & Evans Aggregation Index computes the ratio of observed distance to the expected distance
- The index is given by:

$$R = \frac{\bar{r}_N}{E(r)} = 2\bar{r}\sqrt{\lambda}$$

where $\lambda = \frac{N}{A}$ is the density of points as defined in slide 4.

- R is not "edge corrected" (recall MAUP). Thus, the points closer to the boundary ***may*** have (unnaturally) larger NN distance (why?) resulting in a biased R index.
- Without correction for edge effects, the value of R will be **positively biased (overestimated).**

HTTPS://WWW.JSTOR.ORG/STABLE/1931034?SEQ=1

Clark & Evans Aggregation Index: Interpretation

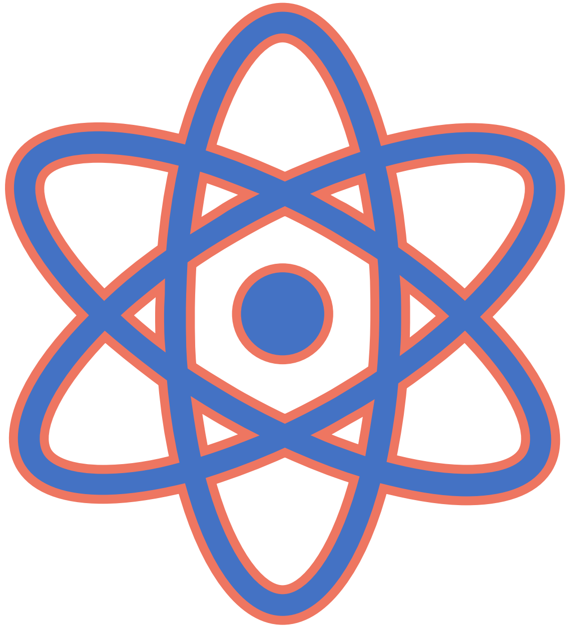
- CSR (HPP): If the observed distance is equal to the expected distance, $\bar{r}_N = E(r)$ and thus, $R = 1$.
- Clustering: If the observed distance is smaller than the expected distance, $\bar{r}_N < E(r)$ and thus, $R < 1$.
  - Maximum aggregation: all the observations occupy the same locus and thus, $R = 0$.
- Dispersion: If the observed distance is larger than the expected distance, $\bar{r}_N > E(r)$ and thus, $R > 1$.
  - Clark & Evans (1954) shows that the maximum value R can take is 2.1491.
- Linear interpretation of R:
  - For example, an R value of 0.5 would indicate that nearest neighbours are, on the average, half as far apart as expected under conditions of CSR.

Clark & Evans Test

- Hypotheses (two sided):

$$H_0: R = 1 \text{ vs. } H_0: R \neq 1$$

- The Clark & Evans Aggregation Index itself is not asymptotically pivotal (what does this mean?), however, when conducting the test using **clarkevans.test()**, the *p*-value for the test is computed by standardising *R* as proposed by Clark and Evans (slide 8) and referring the statistic to the standard Normal distribution. (https://rdrr.io/cran/spatstat/man/clarkevans.test.html)



HTTPS://WWW.SCIENCEDIRECT.COM/SCIENCE/ARTICLE/ABS/PII/0378375895000631

Donelly edge correction

- Donelly (1996) suggested an “edge corrected” version of the test statistic as follows:

$$E(r_c) = E(r) + \left(0.0514 + \frac{0.041}{\sqrt{N}}\right) \times \frac{A}{N}$$

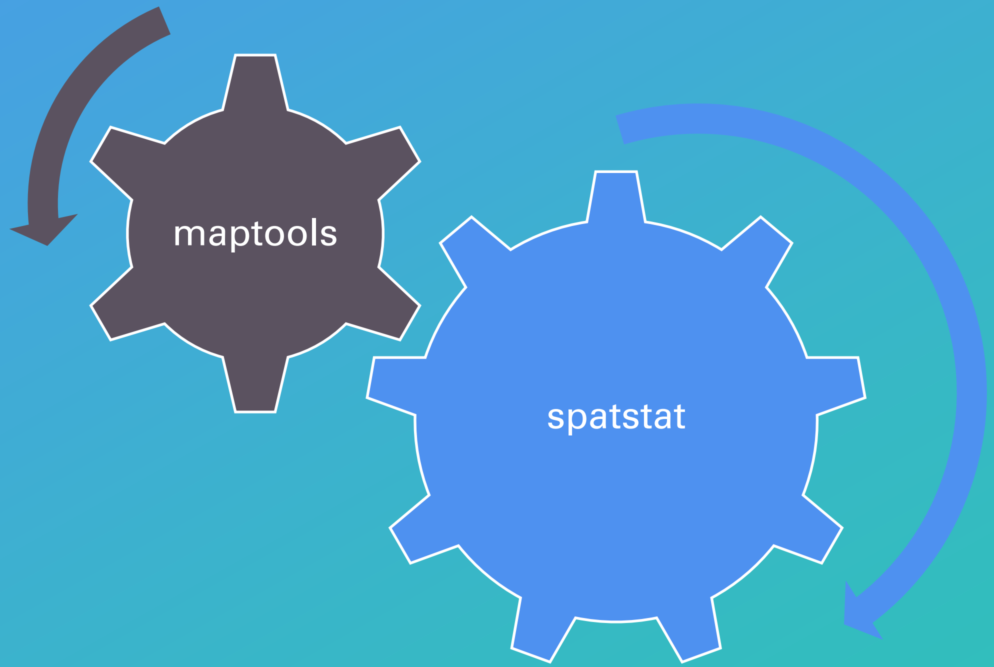
$$s.e.(\bar{r}_{Nc}) = \frac{\bar{r}_N - E(r_c)}{s.e.(\bar{r}_{Nc})} \sim N(0,1)$$

- The drawback however is that this is **only applicable for rectangular regions** which is generally unrealistic. Alternatives include,
  - Applying a guard (buffer) region (Activity B)
  - Using a cdf of second order neighbourhood effect such as *K*-function or *G*-function analysis (refer to 7e).

Steps

1. Calculate the density of points in an area, $\lambda = \frac{N}{A}$.
2. Derive observed average distances, $\bar{r}_N$.
3. Determine the hypothetical random pattern (the null hypothesis), Po$(\lambda \pi r^2)$.
4. Compute the R statistic and perform a statistical test.
5. Conduct a hypothesis test.

PACKAGES YOU NEED



Main results

```
> clarkevans(breach.ppp, correction="none")
```

[1] 0.53845

See: https://www.rdocumentation.org/packages/spatstat/versions/1.64-1/topics/clarkevans

A one tailed test: Is the R index significantly smaller than 1 (or do we have reasons to believe clustering in “breach”)?

```
> clarkevans.test(breach.ppp, correction="none",
alternative="less")

         Clark-Evans test

         No edge correction

         Z-test

data:  breach.ppp

R = 0.53845, p-value < 2.2e-16

alternative hypothesis: clustered (R < 1)
```

Conclusion: there seem to be significant clustering of incidences related to breaches of peace!

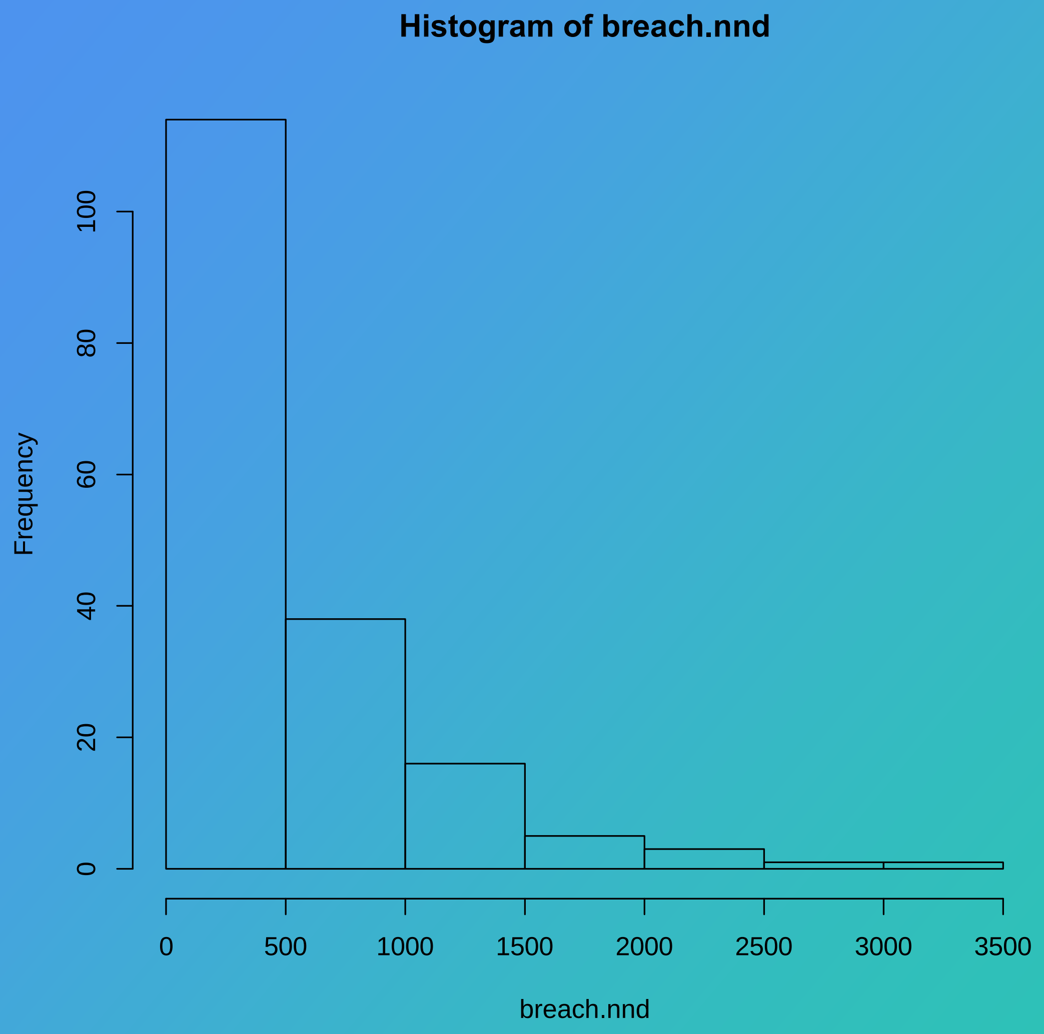
Activity B

1. Would you classify nearest neighbour distance analysis under first order effect or second order effect?
2. Define a clip region that is 1 mile outside of the Newhaven outline and re-run the test with an edge correction based on “guard”. Be sure to state your conclusion clearly

Hint:

```
> st_crs(breach)$proj4string
[1] " +proj=lcc +datum=NAD27
+lon_0=-72d45 +lat_1=41d52
+lat_2=41d12 +lat_0=40d50
+x_0=182880.3657607315 +y_0=0
+units=us-ft +no_defs
+ellps=clrk66
+nadgrids=@conus,@alaska,@ntv2_0.g
sb,@ntv1_can.dat"
```

> 1 mile = 5280ft



Calculations (optional)

Calculate the nearest neighbour distances.

```
> breach.nnd = nndist(breach.ppp, k=1)
```

A quick histogram tells us pretty much everything we need to know. The points are quite clustered with many points lying within a close distance to one another.

```
> summary(breach.nnd)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	14.4	387.9	480.7	612.1	3128.1

```
> hist(breach.nnd)
```

Confirm the calculations (DIY)

Start by converting the "breach.nnd" into a matrix.

```
> class(breach.nnd)
```

[1] "numeric"

```
> breach.nnd = as.matrix(breach.nnd)
```

Observed average r ($\bar{r}_N$),

```
> (r.bar = sum(breach.nnd)/nrow(breach.nnd))
```

[1] 480.7013

$E(r) = \mu_r = \frac{1}{2\sqrt{\lambda}}$ where $\lambda = \frac{N}{A}$

```
> (Er = 0.5*sqrt(area.owin(nh.owin)/nrow(breach.nnd)))
```

[1] 936.6724

Clark & Evans Aggregation Index

```
> (CE.index = r.bar/Er)
```

[1] 0.5132011 (as before)

```
> (Zstat = ((r.bar-Er)/
0.26136)*sqrt(nrow(breach.nnd)^2/area.owin(nh.owin)))
```

[1] -12.42483

(The p-value is computed based on this Zstat!)

Activity C

Conduct a Clark & Evans test for "bramblecanes" dataset that comes with "spatstat" using a Donelly edge correction. Make sure to state your conclusion clearly.

Take home points…

- What is the nearest neighbour distance all about?
- Observed distance vs. expected distance
- The theoretical derivation of the
  - G-function
  - Test statistic
  - Clark and Evan’s index
  - Donelly edge correction
- Conduct a nearest neighbour distance analysis

Important R functions

- clarkevans()
- clarkevans.test()
- nndist()

References

- ***Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations*** by Philip J. Clark and Francis C. Evans, Ecology, Vol. 35, No. 4 (1954), pp. 445-453
- ***Spatial Analysis*** by Tonny Oyana 2nd edition, Chapter 6.
- https://cran.r-project.org/web/packages/spatstat/spatstat.pdf
- https://spatstat.org