Topic 1: Distance methods

In this topic we will further investigate properties of spatial point processes. In particular, we consider the following data analysis methods:

- Introduction to **distance methods.**
- $\mathcal{G}(r)$ **function.**
- $\mathcal{F}(r)$ **function.**

Distance methods

Suppose that a point pattern appears to have constant intensity, and we wish to assess whether the pattern is Poisson.

The classical techniques for investigating interpoint interactions are distance methods, based on measuring the distances between points. Specifically we may consider

- pairwise distances $s_{ij} = ||x_i - x_j||$ between pairs of points x_i and x_j in the pattern;
- nearest neighbor distances $t_i = min_{j \neq i} s_{ij}$, the distance from each point x_i to its nearest neighbor;
- empty space distances $d(u, X) = \min_i ||u - x_i||$, the distance from a fixed reference location u to the nearest data point in X.

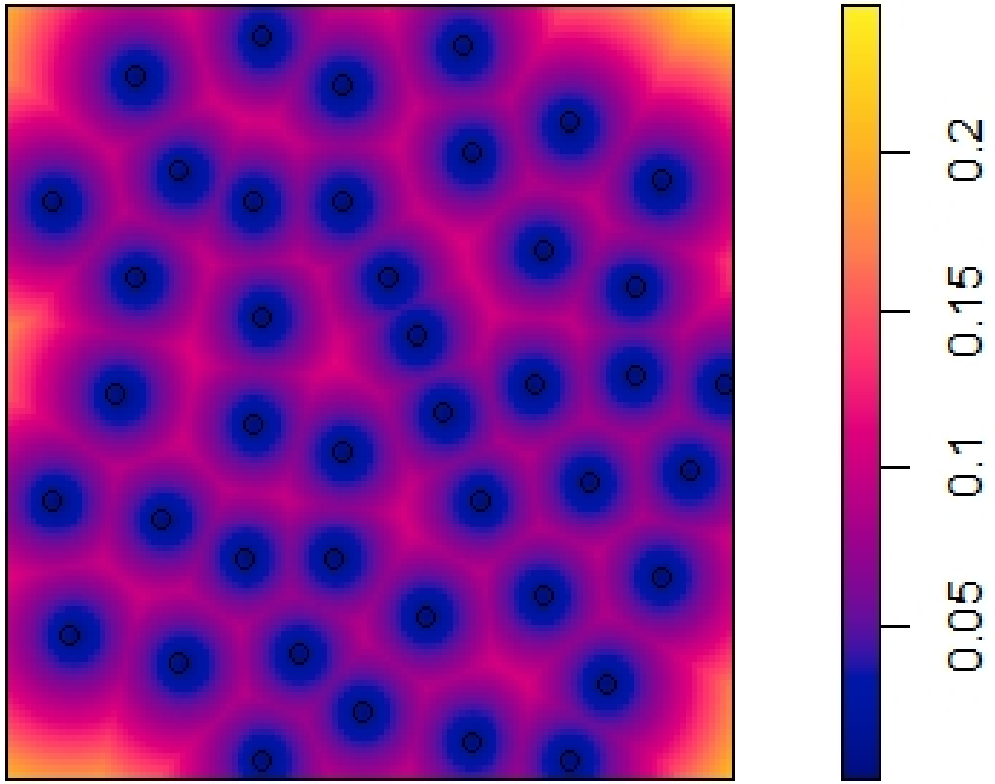
If you need to compute these distances use the SPATSTAT functions **pairdist, nndist** and **distmap** respectively.

If X is a point pattern object, then

- PAIRDIST(X) returns the matrix of pairwise distances.
- NNDIST(X) returns the vector of nearest neighbor distances.
- DISTMAP(X) returns a pixel image whose pixel values are the empty space distances to the pattern X measured from every pixel.

```
> data(cells)
> emp <- distmap(cells)
> plot(emp, main = "Empty space distances")
> plot(cells, add = TRUE)
```

Empty space distances



```
> pairdist(cells)
[,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.0000000 0.1503762 0.2880868 0.4250000 0.4854122
[2,] 0.1503762 0.0000000 0.1544960 0.2945980 0.3401294
[3,] 0.2880868 0.1544960 0.0000000 0.1402462 0.2024080
...
```

The vector of nearest neighbour distances suggests that there are no points that are too far away from the main cloud of points and each point has at least one closest neighbour at approximately same distance as other points:

```
> nndist(cells)
[1] 0.14583895 0.11486514 0.14024621 0.11180340 0.11180340
[6] 0.15449595 0.15449595 0.11486514 0.14577380 0.14243595
[11] 0.15206906 0.12356780 0.12356780 0.12500000 0.12356780
...
```

$\mathcal{F}$ function.

Assuming X is stationary, one can define the cumulative distribution function of the empty space distance

$$\mathcal{F}(r) = P(d(u, X) \leq r),$$

where u is an arbitrary location.

If the point process is stationary then $\mathcal{F}(r)$ does not depend on u. For a Poisson process

$$\mathcal{F}_{POI}(r) = 1 - exp(-\lambda \pi r^2).$$

The empirical distribution function of the observed empty space distances on a grid of locations $u_j,\ j = 1, ..., m$,

$$\hat{\mathcal{F}}(r) = \frac{1}{m} \sum_j \chi\{d(u_j, X) \leq r\}$$

is a biased estimator of $\mathcal{F}(r)$. Corrections for edge effect bias are required. Many edge corrections are available.

Plot the empirical function $\hat{\mathcal{F}}(r)$ against its theoretical expectation $\mathcal{F}(r)$.

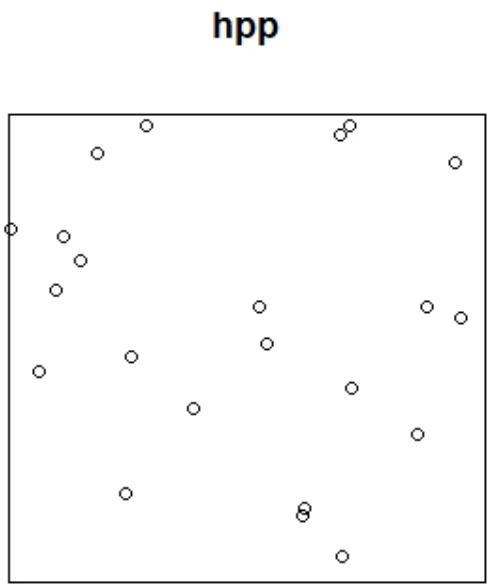
Values $\hat{\mathcal{F}}(r) > \mathcal{F}_{POI}(r)$ suggest that empty space distances in the point pattern are shorter than for a Poisson process, suggesting a regularly space pattern;
while values $\hat{\mathcal{F}}(r) < \mathcal{F}_{POI}(r)$ suggest a clustered pattern.

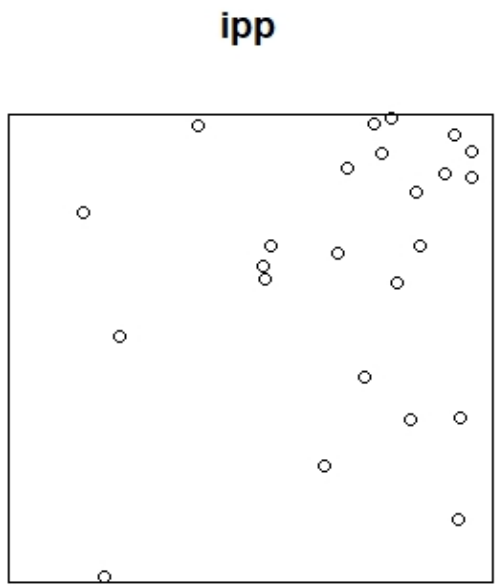
We use RPOISPP() to draw the first dataset from a Poisson point process.

```
> hpp <- rpoispp(30)
> class(hpp)
[1] "ppp"
> hpp$x
> hpp$y
> plot(hpp)
```

The second data set is a realization of inhomogeneous Poisson process which is an extension of the stationary Poisson process, where the intensity can vary over space as a function of position.

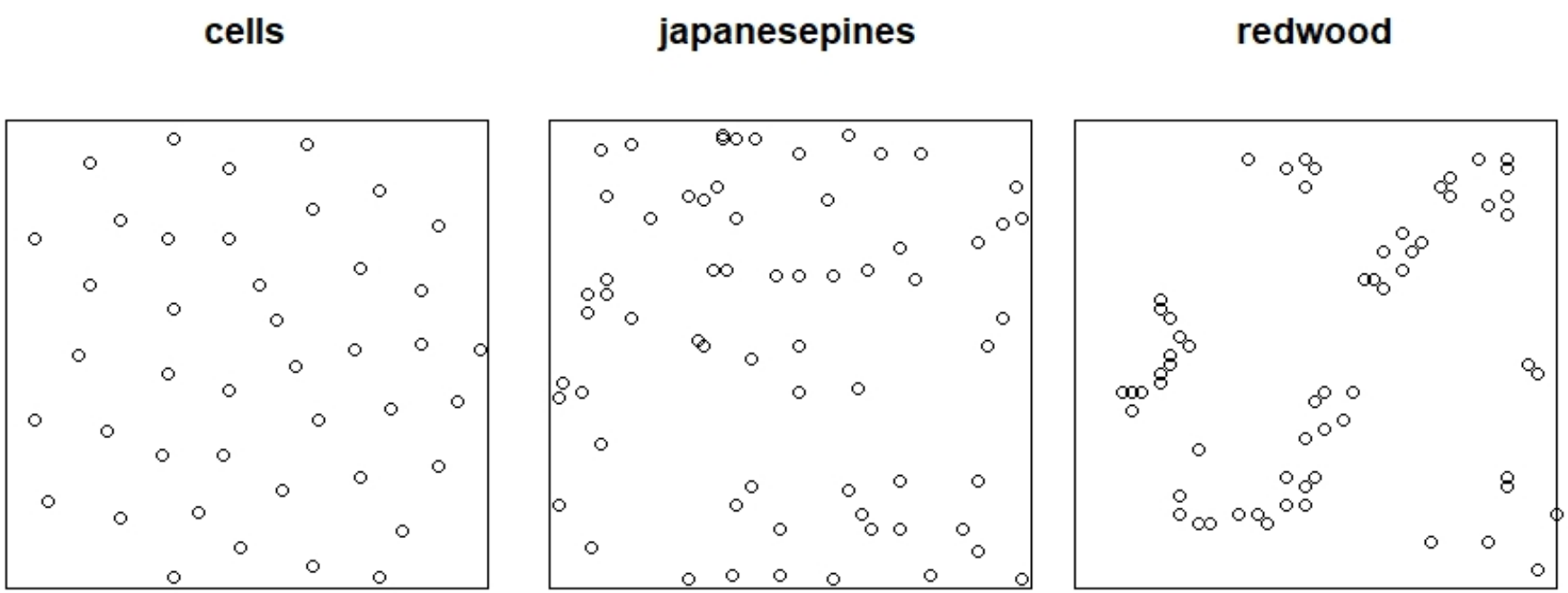
```
> ipp <- rpoispp(function(x,y) {30*(x^2 + y^2)})
> plot(ipp)
```





We also use the following real data:

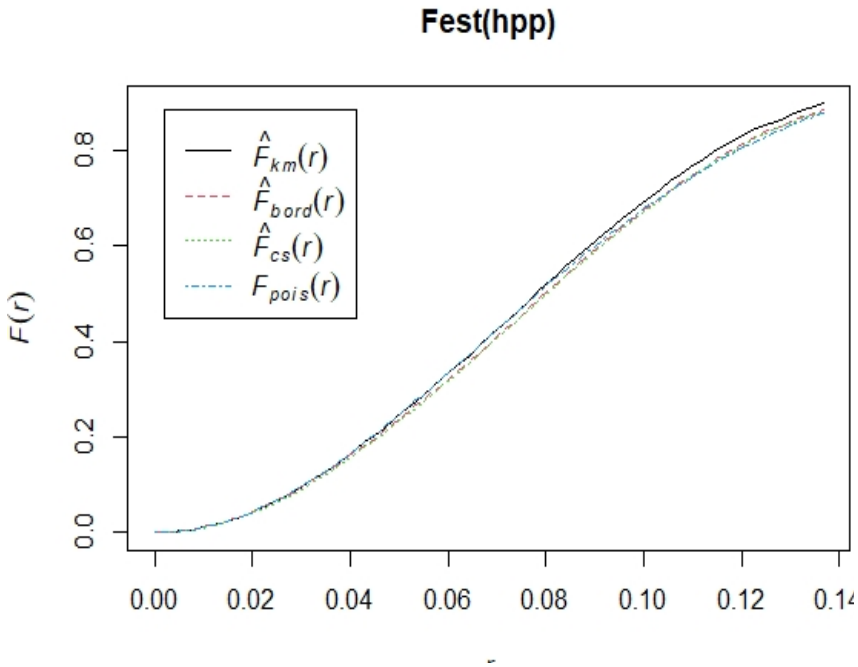
```
> data(cells)
> plot(cells)
> data(japanesepines)
> plot(japanesepines)
> data(redwood)
> plot(redwood)
```

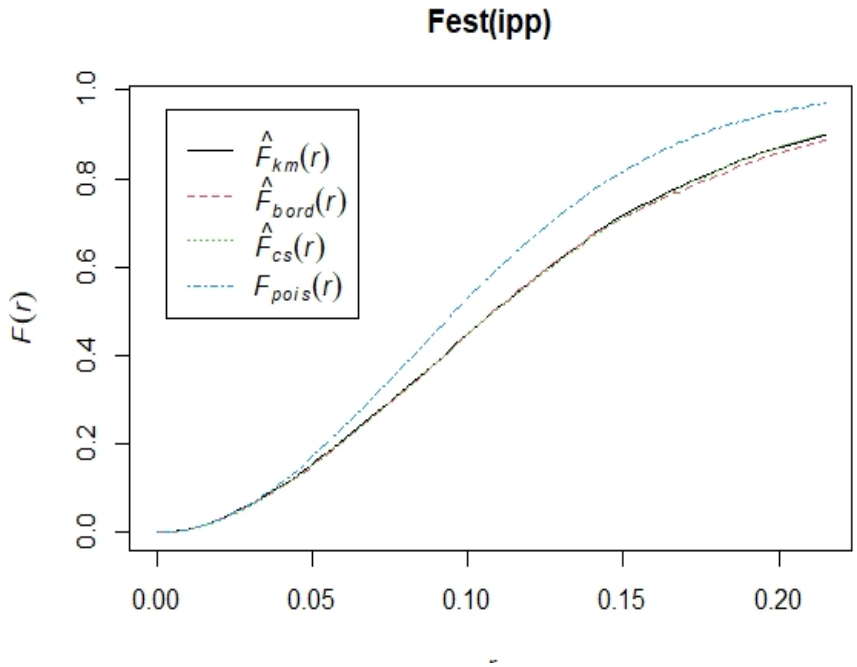


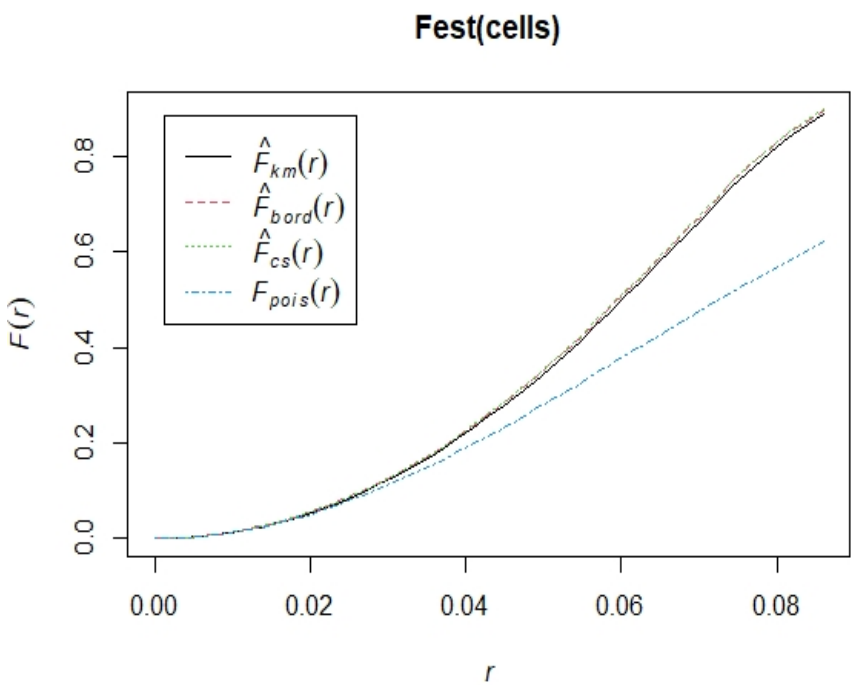
The function **Fest** computes estimates of $\hat{\mathcal{F}}(r)$ using several edge corrections, and the benchmark value for the Poisson process.

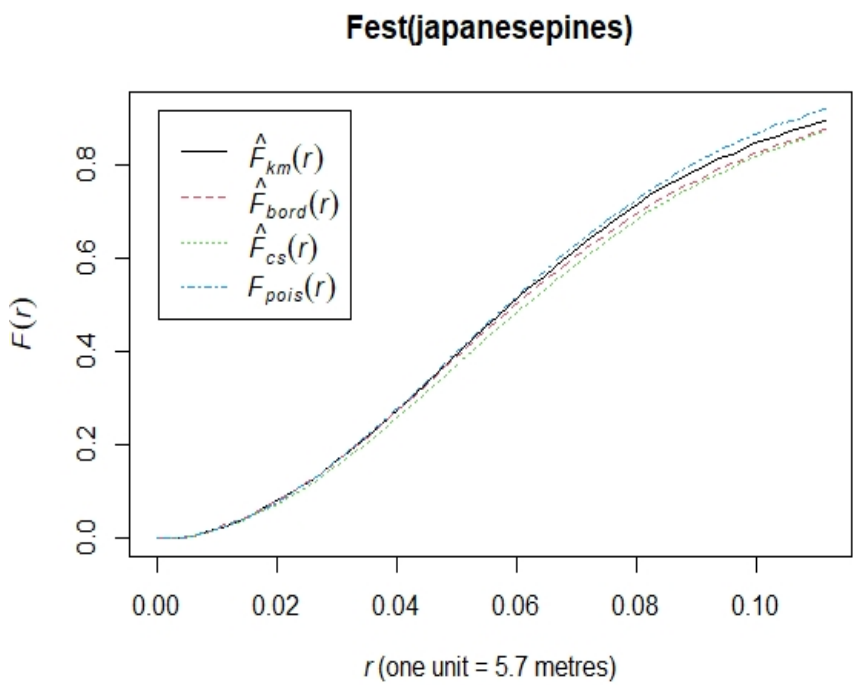
rs, km and **raw** contain different estimates of the empty space function $\mathcal{F}(r)$, namely the reduced sample estimator, the Kaplan-Meier estimator, and the uncorrected empirical distribution function, respectively.

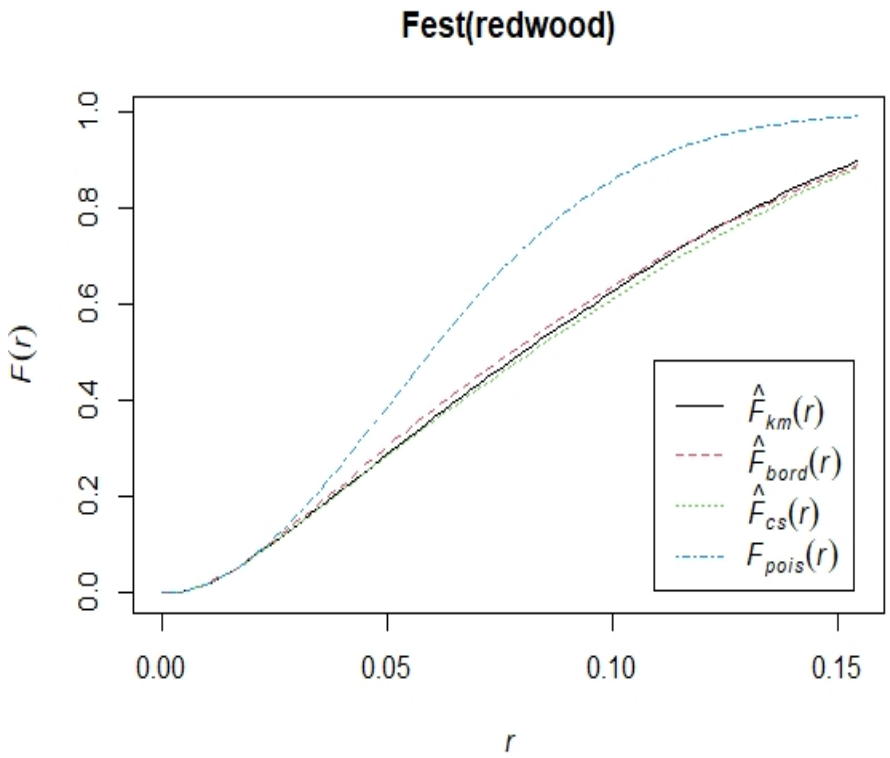
```
> plot(Fest(hpp))
> plot(Fest(ipp))
> plot(Fest(cells))
> plot(Fest(japanesepines))
> plot(Fest(redwood))
```


Fest(hpp)
$\hat{F}_{km}(r)$
$\hat{F}_{bord}(r)$
$\hat{F}_{cs}(r)$
$F_{pois}(r)$
F(r)
0.0
0.2
0.4
0.6
0.8
0.00
0.02
0.04
0.06
0.08
0.10
0.12


Fest(ipp)
$\hat{F}_{km}(r)$
$\hat{F}_{bord}(r)$
$\hat{F}_{cs}(r)$
$F_{pois}(r)$
F(r)
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20


Fest(cells)
$\hat{F}_{km}(r)$
$\hat{F}_{bord}(r)$
$\hat{F}_{cs}(r)$
$F_{pois}(r)$
F(r)
0.0
0.2
0.4
0.6
0.8
0.00
0.02
0.04
0.06
0.08
r


Fest(japanesepines)
$\hat{F}_{km}(r)$
$\hat{F}_{bord}(r)$
$\hat{F}_{cs}(r)$
$F_{pois}(r)$
F(r)
0.0
0.2
0.4
0.6
0.8
0.00
0.02
0.04
0.06
0.08
0.10
r (one unit = 5.7 metres)


Fest(redwood)
F(r)
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
r
$\hat{F}_{km}(r)$
$\hat{F}_{bord}(r)$
$\hat{F}_{cs}(r)$
$F_{pois}(r)$

$\mathcal{G}$ function.

For a stationary point process X one can define the cumulative distribution function of the nearest-neighbor distance for a typical point in the pattern

$$\mathcal{G}(r) = P\left(d(u, X\backslash\{u\}) \leq r | u \in X\right)$$

where u is an arbitrary location, and $d(u, X\backslash\{u\})$ is the shortest distance from u to the point pattern X excluding u itself. If the process is stationary then this definition does not depend on u.

For a Poisson process

$$\mathcal{G}_{POI}(r) = 1 - exp(-\lambda\pi r^2).$$

Let $n(X)$ be the number of points in X. The empirical distribution function of the observed nearest-neighbour distances

$$\hat{\mathcal{G}}(r) = \frac{\sum_i \chi\{t_i \leq r\}}{n(X)}$$

is a biased estimator of $\mathcal{G}(r)$. There are many edge corrections.

Interpretation of $\mathcal{G}(r)$ is the reverse of $\mathcal{F}(r)$.

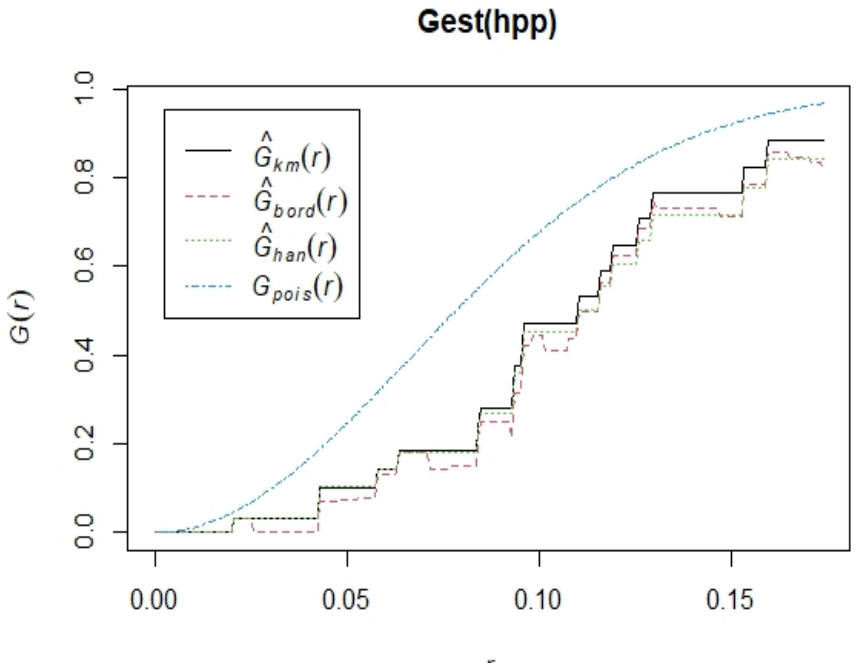
Values $\hat{\mathcal{G}}(r) > \mathcal{G}_{POI}(r)$ suggest that nearest neighbor distances in the point pattern are shorter than for a Poisson process, suggesting a clustered pattern;

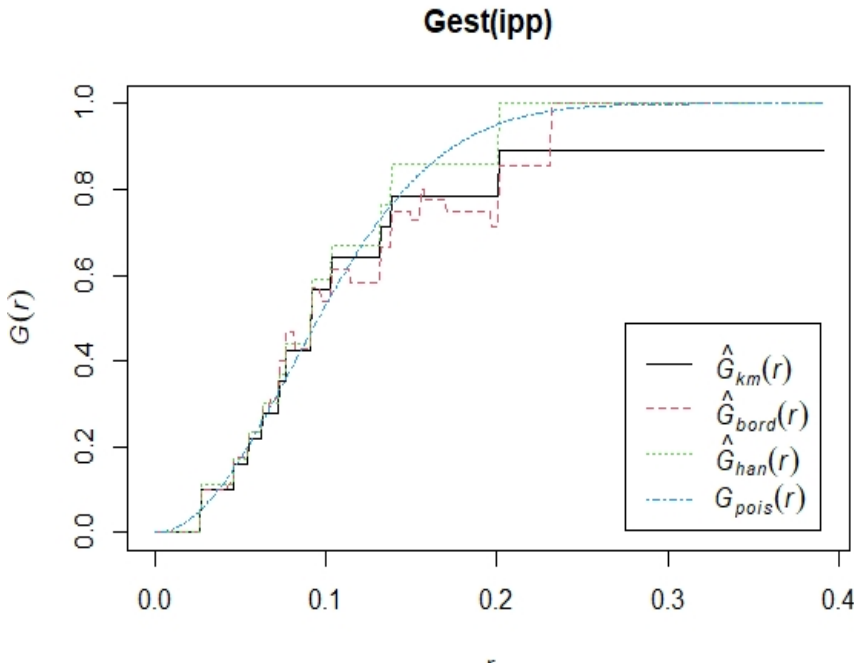
while values $\hat{\mathcal{G}}(r) < \mathcal{G}_{POI}(r)$ suggest a regular (inhibited) pattern.

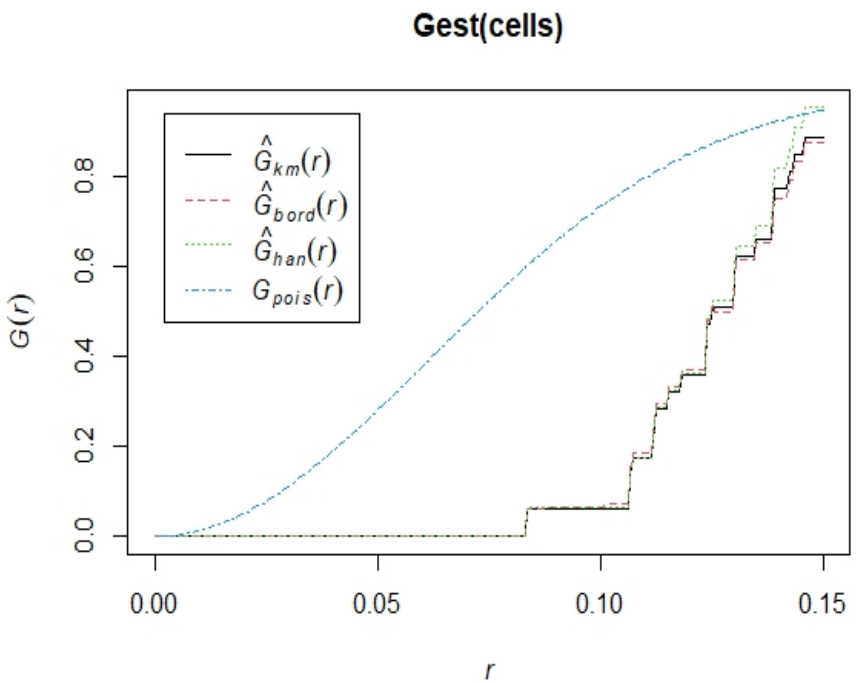
One can compare this theoretical to an empirical estimator and see if there are systematic deviations from the expected values using the function $\textsc{Gest}()$.

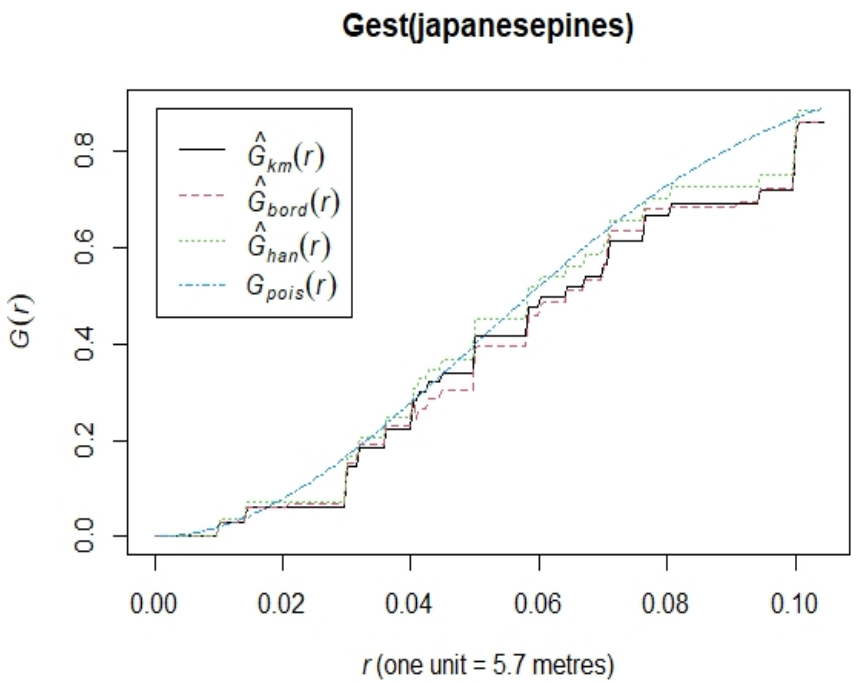
The function $\textsc{Gest}$ computes estimates of $\mathcal{G}(r)$ using several edge corrections, and the benchmark value for the Poisson process.

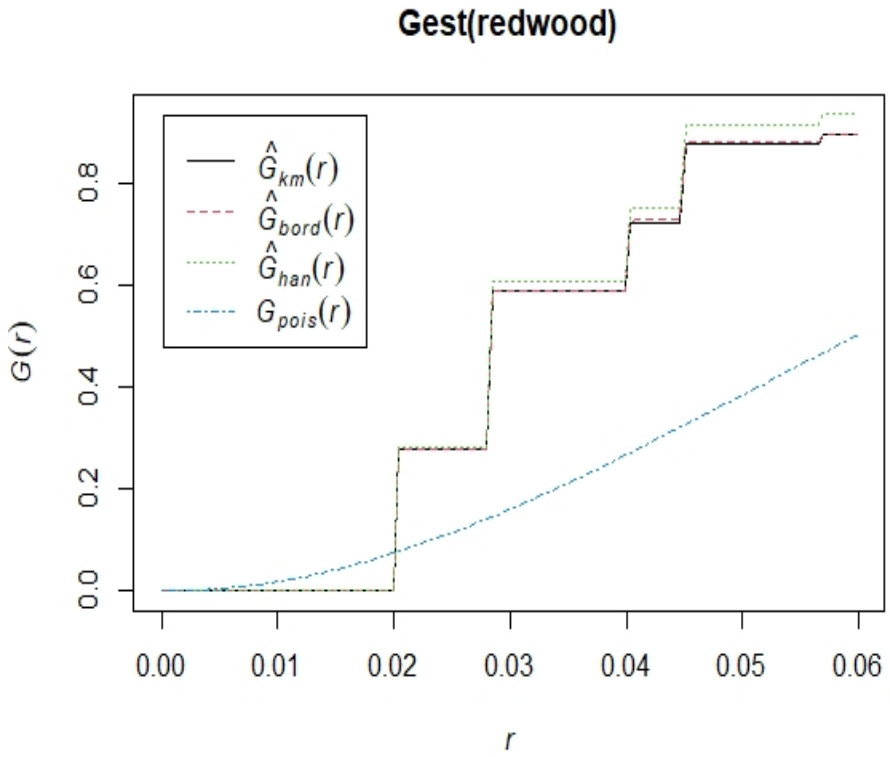
```
> plot(Gest(hpp))
> plot(Gest(ipp))
> plot(Gest(cells))
> plot(Gest(japanesepines))
> plot(Gest(redwood))
```


Gest(hpp)
G(r)
$\hat{G}_{km}(r)$
$\hat{G}_{bord}(r)$
$\hat{G}_{han}(r)$
$G_{pois}(r)$
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15


Gest(ipp)
G(r)
$\hat{G}_{km}(r)$
$\hat{G}_{bord}(r)$
$\hat{G}_{han}(r)$
$G_{pois}(r)$
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4


Gest(cells)
G(r)
$\hat{G}_{km}(r)$
$\hat{G}_{bord}(r)$
$\hat{G}_{han}(r)$
$G_{pois}(r)$
0.0
0.2
0.4
0.6
0.8
0.00
0.05
0.10
0.15
r


Gest(japanesepines)
G(r)
$\hat{G}_{km}(r)$
$\hat{G}_{bord}(r)$
$\hat{G}_{han}(r)$
$G_{pois}(r)$
0.0
0.2
0.4
0.6
0.8
0.00
0.02
0.04
0.06
0.08
0.10
r (one unit = 5.7 metres)


Gest(redwood)
$\hat{G}_{km}(r)$
$\hat{G}_{bord}(r)$
$\hat{G}_{han}(r)$
$G_{pois}(r)$
G(r)
0.0
0.2
0.4
0.6
0.8
0.00
0.01
0.02
0.03
0.04
0.05
0.06
r

Key R commands

pairdist(X)	*computes a matrix of pairwise distances*
nndist(X)	*returns the vector of nearest neighbour distances*
distmap(X)	*computes the distance map*
cells	*data with locations of the centres of 42 biological cells*
japanesepines	*data with locations of Japanese black pine*
redwood	*data with locations of 62 seedlings of California Giant Redwood*
Fest(X,...)	*estimates the empty space function F(r)*
Gest(X,...)	*estimates the nearest neighbour distance function G(r)*