Point Patterns

Terminology

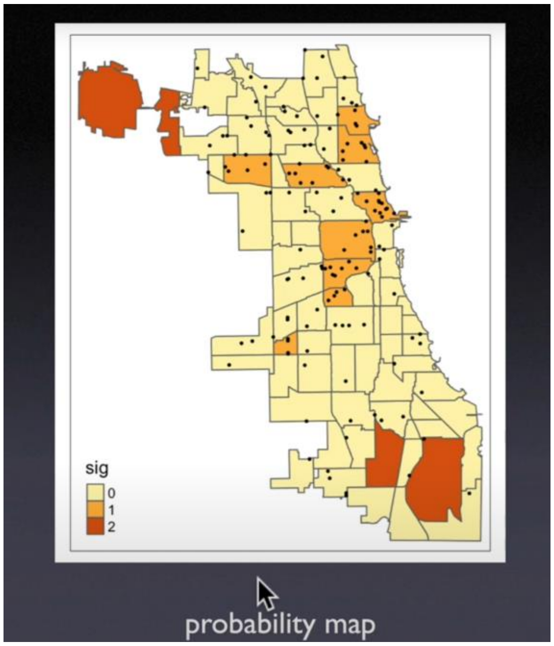
- Attraction / association
- Isotropic plane
- Network distance
- Population at risk / case control (proxy) design / background heterogeneity
  - Classic analysis: Lancashire cancers
- Concentration of events
- Patten and process (causation)
- Ripley-Rasson transformation to bounding box and convex hull
  - Push out the bounding by specific ratio
- Intensity
- CSR (Complete spatial randomness) - Use as reference

Homogeneous planar Poisson process

It is one way to simulate the CSR pattern, which used as a reference

- Each location has equal probability for an event
- Location of events are independent
- $N(A) \sim Poisson(\lambda|A|)$, λ is Intensity, equal to $\frac{N}{|A|}$, A is area
- $P(N(A) = y) = \frac{e^{-\lambda|A|} * (\lambda|A|)^y}{y!}$

We could use it to generate the probability map as follows (Dark area means the number of events is extremely lower than the expected value):



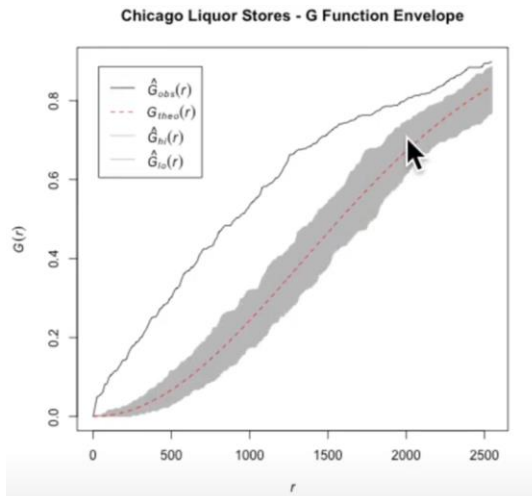
Kernel density

Kernel density is one famous None-parametric-approach for heterogeneity intensity investigating

- Intensity function (expected number) vs density function (probability of an event)
- $f(u) = \frac{1}{N_b}\sum_i K[\frac{u-u_i}{b}]$, u is any location, b is bandwidth, K is kernel function.
- Usually, we would use gaussian kernel function
  - $K(z) = (\frac{1}{\sqrt{2\pi}b}e^{(-\frac{1}{2b^2})z^2})$

Simulation envelope

Simulation envelope-mimic the point pattern under spatial randomness (dark area)



Nearest Neighbor Statistics

G-event to event, F- event to point, J-combination of both

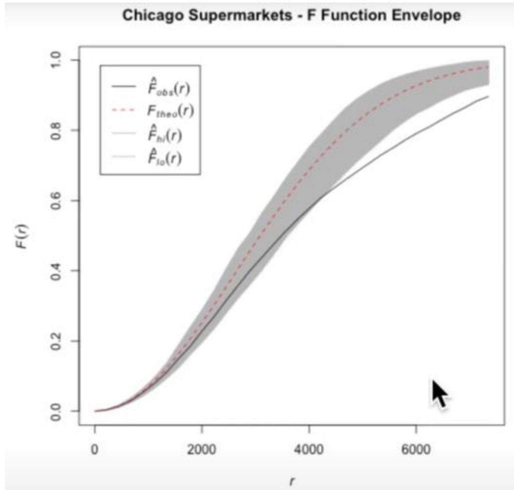
They all investigate the global property about one point pattern process

$$G(r) = \frac{\sum_i I(r_i \leq r)}{n}$$

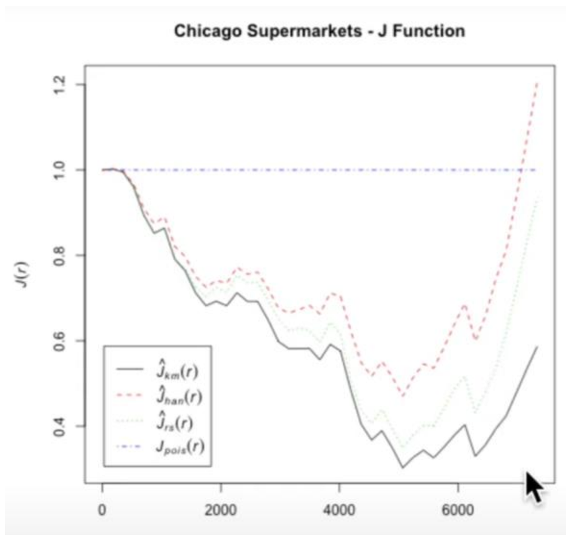
F function is very similar to G function, but it calculates the distance between events (interested points) to points (CSR Points). G function measures the events to events distances.

$$F(r) = \frac{\sum_i I(r_i \leq r)}{m}$$

From the Homogeneous planar Poisson process, when $A = \pi r^2$, we could get $P(y = 0) = e^{-\lambda \pi r^2}$. So, the probability of "observe at least one event within the buffer r" is $P(r_i < r) = 1 - e^{-\lambda \pi r^2}$. So, we compare the observed probability under r with theoretical one, and plot $P(r_i < r)$ against r



J function is the ratio between G and F function, it normalized the line under the CRS setting



Drawbacks

1. do not capture the property of the whole process
2. cannot measure the interaction between events from two different approach

Global K function and local K

Global K

Global function does not suggest the location of clusters.

$$K(r) = \frac{1}{\lambda}\frac{1}{(n-1)}\sum_i\sum_j I(d_{ij} < r) = \frac{|A|}{n(n-1)}\sum_i\sum_j I(d_{ij} < r)$$