Point Pattern Analysis
Part II

Outline

- Assessing point patterns statistically → Spatial Statistical Analysis

Review (PPT Statistics)

- Density-based :
  - Quadrat Counts
  - Kernel density estimation
- Distance-based:
  - Mean NND
  - G & F functions
  - K function
- Proximity polygon

The “Null” Spatial Process

- The standard process using in statistical test is CSR:
  - No 1st order effect: equal probability
  - No 2nd order effect: independence
- Statistically speaking, SSA is to:
  - Compare the observed patterns (measured in certain statistics) to the expected patterns under CSR
  - Then determine if we should reject or not reject this “null” hypothesis

A Reminder on Null Hypothesis in Statistical Test

- In framing a question statistically, we ask the question:
  - “*How probable are the observed patterns on the assumption that the null hypothesis is true?*”
- In other words:
  - “*How likely are the observed patterns on the assumption that CSR is the underlying process?*”

Spatial Analysis Framework

- Assume observed patterns are a realization of a process!!!
- Describe the pattern using certain statistics
- **Calculate the expected values (same statistics) for patterns produced by a hypothesized process**
- **Compare these two**

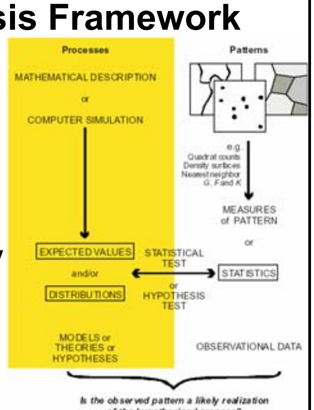

Illustration (Quadrat Count)

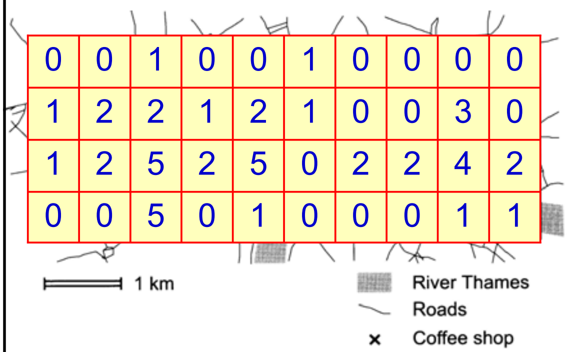

Illustration (Quadrat Count)

# events	#quadrats
0	18
1	9
2	8
3	1
4	1
5	3

- Compare to the expected counts by **CSR** using Chi-Square Test

Calculation of Chi-Square Test

- For each quadrat count determine the difference of the observed number of events from the expected number of events per quadrat (*n/a*)
- Square the differences and sum them
- Divide the total by the expected count per quadrat to get a chi-square test statistic
  - Large value → clustered; small value → even spacing
- Use this result to determine a *p* value from a table of Chi-square values
  - The degrees of freedom is given by the number of quadrats -1

Illustration (Quadrat Count)

- Total 47 events in 40 quadrats → expected # of events per quadrat = 47/40 = 1.175

# events	# quadrats	Difference from expected	Difference squared	Total difference squared
0	18	-1.175	1.380625	24.85125
1	9	-0.175	0.030625	0.275625
2	8	0.825	0.680625	5.445
3	1	1.825	3.330625	3.330625
4	1	2.825	7.980625	7.980625
5	3	3.825	14.630625	43.891875
				85.775

Illustration (Quadrat Count)

- Divide 85.775 by 1.175 → a of 73.0
- Larger than a value for Chi-square test statistic at *p* = 0.05 (54.57), 0.01(62.43), and even 0.001(72.06)
- Conclude: reject the null hypothesis – the pattern is significantly different from randomness with 99.9% confidence

Extending the Same Idea

- In the same way, if we know expected values of other pattern measures → statistic test
- Mean NND can be predicted based on mean event density
- G(d), F(d), K(d) can be predicted also

CSR and Other Measures

- Mean NND $E(d) = \frac{1}{2\sqrt{\lambda}}$
- G(d) & F(d): $E(G(d)/F(d)) = 1 - e^{1\lambda\pi d^2}$
- K(d): $E(K(d)) = \pi d^2$

Transformation of K(d)

- A problem with K(d):
  - Notice due to the dependence of d^2, K increases rapidly and comparison of two curves can be difficult
- So K(d) → L(d) with expected value (CSR) = 0:

$$K - E(K(d)) = 0 \qquad K = \pi d^2 \qquad L(d) = \sqrt{\frac{K}{\pi}} - d$$

- For exploratory purpose, we can plot L(d) and look at its relationship to zero:
  - Clustered L(d) > 0: more points than expected
  - Even spacing L(d) < 0: fewer points than expected

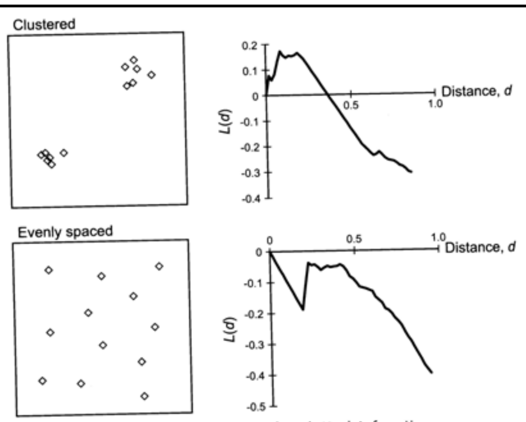

Figure 4.19 Example of a plotted *L* function.

A Problem with Distance Statistics

- Rely on NNDs → Edge Effect and shape of the study area
  - Points near the edge of the study area tend to have higher NNDs, especially when the total # of points is small
    - e.g. K(*d*) does not consider points outside when *d* is large → solution: guard zone
  - Thus the expected value given by the equations need to adjust → too complex depends on the shape of the study area

Monte Carlo Simulation

- We can generate a simulated prediction of L: **Monte Carlo Simulation !!!**
  - Randomly generate a large number (e.g. 1000) patterns of *n* events in the study region → sampling under CSR
  - Calculate L(d) (mean NND or G or F or K) for each simulation
  - These form an envelop or range of certain confidence level (*p*) and can be used to assess the observed L(d)

Pros & Cons of Monte Carlo Simulation

- Pros:
  - Any required significance level can be produced
  - Remove Edge Effect
  - Not as sensitive to estimation of intensity
  - Processes other than CSR can be conveniently tested
- Cons:
  - Computation intensive, especially for higher significance levels

GeoComputation!!!!

Points for Discussion

- Is CSR a useful null hypothesis?
- What other processes might be used?
- How would we test these processes?
- If CSR is the null hypothesis, what is the "alternative hypothesis"?
- Something wrong here?

Review

- Statistical test: compare observed patterns with expected patterns under CSR
  - For quadrat count, use a Chi-square test
  - For NND, G(d) & F(d), K(D) there are analytical results that we can test against
  - Analytical result too complex → use Monte Carlo simulation