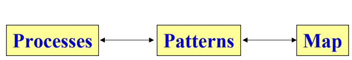
Map as Outcomes of Processes

Outline

- Definitions: processes & patterns
- A Starting point: Complete Spatial Randomness
- More definition: Stationary

Map as Outcomes of Processes

- The basic assumption of Spatial Analysis:
  - Maps have the ability to show patterns in the phenomena they represent
  - Patterns provide clues to a possible causal process
  - Thus, maps can be understood as outcomes of processes



What Is Process?

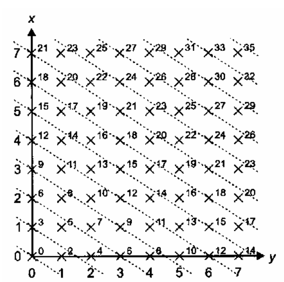
- Definition: A spatial process is a description of how a spatial pattern might be generated
  - Human activities (residence, employment, leisure) determine urban structures
  - Geological forces form different landscapes
- Types of Processes: the chance of a specific outcome
  - Deterministic: 100%
  - Stochastic: < 100%

Deterministic Processes

- Always the same outcome at a location

$$z = 2x + 3y$$

Where x and y are two spatial coordinates while z is the numerical value for a variable



Stochastic Processes

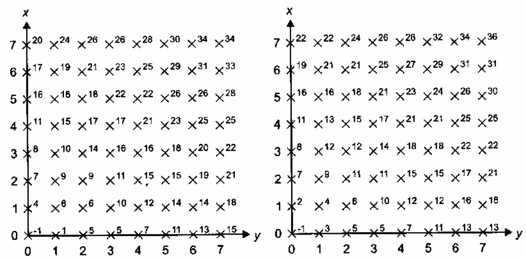
- More often, geographic phenomena appear to be the result of a chance process
  - Outcome is subject to variation that cannot be given precisely by a mathematical function
    - e.g. the individual or collective results of human decisions
  - Some spatial patterns are the results of deterministic physical laws, but they appear as if they are the results of chance process.

$$z = 2x + 3y + d$$

Where d is a randomly chosen value at each location, -1 or +1.

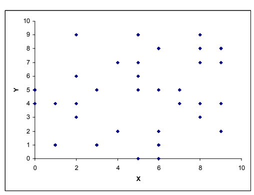
Stochastic Processes (Cont.)

- **Stochastic:** two realizations of $z = 2x + 3y \pm 1$
  - 2^{64} = 18,446,744,073,709,551,616 possible realizations



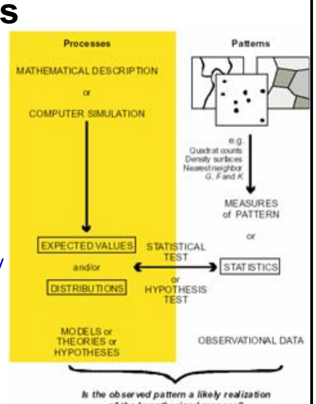
Exercise - 4

- Dot map with randomly distributed points
  - Created 40 (20 each column) Random numbers from Excel
    - Int(10 * Rand())
  - Use them as x and y coordinates for a plot
  - Repeat this process



SA & Processes

- Basic Question of SA: "*How likely is this (observed) pattern a realization of that (hypothesized) process?*"
  - Observed pattern is only one potential realization of a hypothesized process!!!

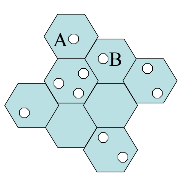


Complete Spatial Randomness (CSR)

- Suppose we assume that:
  - *Anything can happen at anywhere, with equal probability*
- We have Complete Spatial Randomness (**CSR**)
  - Means that no Geographic effects
  - The most commonly used "*Standard*" process or null hypothesis
  - Also called Independent Random Process (IRP)

Illustration: IRP/CSR

- Quadrat Analysis
  - Count the numbers of events in each quadrat
  - Event: a point in the map, representing an incident.
  - Quadrats: a set of equal-sized & nonoverlapping areas



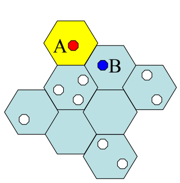
Pattern

↕

CSR Process

Illustration: IRP/CSR (Cont.)

- Assumptions:
  - Equal probability
  - Independence



P (event A in Yellow quadrat) = 1/8

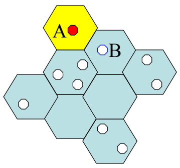
P (event A not in Yellow quadrat) = 7/8

P (only event A in the Yellow quadrat)
= P (event A in Yellow quadrat and other events not in the Yellow quadrat)

$$= \frac{1}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8}$$

A B C D E F G H I J

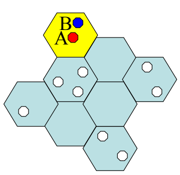
Illustration: IRP/CSR (Cont.)



P (one event only)
= P (event A only) + P (event B only) + … +
P (event J only)
= 10 × P (event A only)

$$= 10 \times \frac{1}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8}$$

Illustration: IRP/CSR (Cont.)



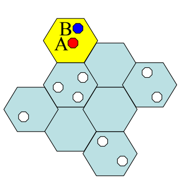
P (event A & B in Yellow quadrat) = 1/8 ×1/8

P (event A & B in Yellow quadrat only) =
P ((event A & B in Yellow quadrat) and (other events not in Yellow quadrat))

$$= \frac{1}{8} \times \frac{1}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} \times \frac{7}{8}$$

A B C D E F G H I J

Illustration: IRP/CSR (Cont.)

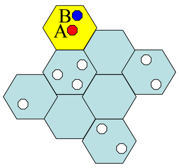


P (two events in Yellow quadrat)
= P(A&B only) + P(A&C only) + … + P(I&J only)

=(no. possible combinations of two events) × $\left(\frac{1}{8}\right)^2 \times \left(\frac{7}{8}\right)^8$

How many possible combinations?

Illustration: IRP/CSR (Cont.)



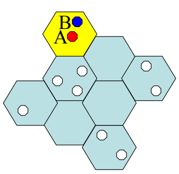
The formula for number of possible combinations of k events from a set of n events is given by

$$C_k^n = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 1$$

In our case, n = 10, and $k = 2$

Illustration: IRP/CSR (Cont.)



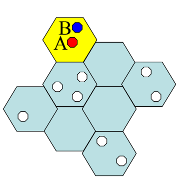
$$\text{P } (k \text{ events}) = C_k^{10} \times \left(\frac{1}{8}\right)^k \times \left(\frac{7}{8}\right)^{10-k}$$

$$= \frac{10!}{k!(10-k)!} \times \left(\frac{1}{8}\right)^k \times \left(\frac{7}{8}\right)^{10-k}$$

$$P(n,k) = \binom{n}{k} p^k (1-p)^{n-k}$$

p = quadrat area / area of study region

Binomial Distribution



Binomial distribution

$$P(k,n,x) = \binom{n}{k}\left(\frac{1}{x}\right)^k\left(\frac{x-1}{x}\right)^{n-k}$$

x is the number of quadrats used
n is the number of events
k is the number of events in a quadrat

Poisson Distribution

- Binomial distribution is computation intensive
  - 50! = 30,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
- Poisson distribution is a good approximation of binomial distribution
  - Reduced computation burden

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$\lambda = \frac{n}{x}$ Intensity of events

e is a constant, = 2.7182818

Poisson vs. Binomial

k	*Binomial*	*Poisson*
0	0.26307558	0.28650480
1	0.37582225	0.35813100
2	0.24160002	0.22383187
3	0.09203810	0.09326328
4	0.02300953	0.02914478
5	0.00394449	0.00728619
6	0.00046958	0.00151796
7	0.00003833	0.00027106
8	0.00000205	0.00004235
9	0.00000007	0.00000588
10	0.00000000	0.00000074

Real Processes Differ From IRP/CSR

- The CSR is mathematically elegant and forms a useful starting point for spatial analysis, but its use is often exceedingly naive and unrealistic.
- If real-world spatial patterns were indeed generated by unconstrained randomness, geography would have little meaning or interest, and most GIS operations would be pointless

Real Processes Differ From IRP/CSR

- First-Order effect: variations in the density of a process across space due to variations in environment properties: **No** equal probability
  - e.g. plants are always clustered in the areas with favored soils
  - e.g. the locations of disease cases tends to cluster in more densely populated areas
- Second-Order effect: interaction between locations: **NO** independence
  - e.g. physicians tend to cluster around a major medical facility
  - e.g. stores of McDonald tend to be far away from each other

Real Processes Differ From IRP/CSR

- First-order and second-order effects shift a process from being stationary to changing over space
  - 1st Order stationary: no variation in the intensity of point events over space
  - 2nd Order stationary: no interaction in events
  - Non-stationary: violate either 1st or 2nd → a model can not universally apply

Real Processes Differ From IRP/CSR

- BUT, in practice it is close to impossible to distinguish from variation in the environment or interaction by the analysis of spatial data
  - What cause the clusters of crime in a city?
    - Environment factors: low income, low employment rate …
    - Or vulnerability of population?

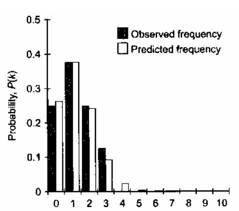
Summary

- Map can be regarded as outcome of a process
  - A pattern is the result of a process thus able to offer clues for processes
- Deterministic vs. stochastic processes
  - Spatial processes are more likely to be stochastic
- A starting point: Complete Spatial Randomness (CSR)
  - Equal probability: 1st order
  - Independence: 2nd order
- Standard SA: Compare observed pattern to CSR

Illustration: IRP/CSR (Cont.)

- Observed vs. Predicted

k	*Number of quadrats*	*Fraction observed*	*Fraction predicted*
0	2	0.250	0.2630755
1	3	0.375	0.3758222
2	2	0.250	0.2416000
3	1	0.125	0.0920391
4	0	0.000	0.0230095
5	0	0.000	0.0039445
6	0	0.000	0.0004696
7	0	0.000	0.0000383
8	0	0.000	0.0000021
9	0	0.000	0.0000001
10	0	0.000	0.0000000



More Differentiations

- Anisotropic: directional effects in spatial variation of data
  - e.g. Down-stream areas are polluted by up-stream sources
- Isotropic: ***NO*** directional effects in spatial variation of data
  - e.g. the infestation rate of disease simply spread outward