Module organization

- Basic concepts
- Complete Spatial Randomness (CSR)
- Descriptive measures
  - Density based
  - Distance based
- Inference and interpretation

Basic concepts

- Spatial process underlying observed spatial pattern
- Spatial variation can be decomposed into:
  - Large scale variation: mean of spatial process
    \[ z_s = f(X_sB + e_s) \]
  - Small scale variation: covariance of spatial process
    \[ \text{Cov}(e_i, e_j) \neq 0 \]
- Objectives:
  - Descriptive measures and visual assessment
  - Inference to spatial process
    - What is the likelihood that the observed pattern resulted from a given stochastic spatial process?

Demonstration data 2: The Beach

Spatial continuous (fields) → Geostatistics

Points (objects) → Point Pattern Analysis

Irregular / Regular lattice (objects) → Spatial Econometrics

Volume of interaction among areas → Spatial Interaction Modeling
Complete Spatial Randomness (CSR)

- Baseline distribution:
  - clustered, random (CSR), or regular.

- Definition of CSR:
  - Equal probability:
    - Everywhere in domain $S$ has an equal chance of event.
  - Independent:
    - Event locations are mutually independent; location of one event has no impact on location of other events.

Point pattern description

- $n$ events
- $s = \{s_1, \ldots, s_n\}$ locations where $s_i = \{x_i, y_i\}$
- $A =$ study region
- $|A| = a =$ area of region

Density-based description

- Large scale variation; first-order intensity
  $$\lambda = \frac{n}{|A|} = \frac{\sum S \in A}{|A|}$$

- Density estimation
  - kernel-smoothing
  $$\lambda_p = \frac{\sum S \in C(p, r)}{\pi r^2}$$
  - $r$ is bandwidth, $p$ is set of nodes on grid.
Distance-based description

- **Distance from** $s_i, s_j$:
  - $d(s_i, s_j) = \text{distance from } s_i, s_j$.
- **Distance matrix**:

![Distance matrix diagram]

- Nearest neighbor
  - $G(d) = \frac{\#[d_{\min}(S_i) < d]}{n}$
    - Interpretation: CDF of nearest-neighbor distances
      - If $G$ increases faster than CSR, clustered
      - If $G$ increases slower than CSR, dispersed
  - $F(d) = \frac{\#[d_{\min}(p_i, S) < d]}{m}$
    - If $F$ increases slower than CSR, clustered
    - If $F$ increases faster than CSR, dispersed
  - $J(d) = \frac{1 - G(d)}{1 - F(d)}$ (1)
    - $J(d) > 1$ clustering; $J(d) < 1$ dispersion; $J(d) = 1$ random

Distance-based description

- **Interevent distances**
  - $K(d) = \text{sum}[S \text{ in } C(s_i, d)]$
    - $L(d)$ transforms to center on zero
    - $L(d) = 0$, random
    - $L(d) > 0$, clustered
    - $L(d) < 0$, dispersed
Inhomogeneous and marked point patterns

- What if $\lambda \neq \lambda(s)$?
- $K_{\text{inhom}}(S, \lambda(s))$ incorporates spatially varying mean.

- Marked patterns: May be interested in the dependence or inhibition among different types of events (suits vs. no suits)

Inference for distance measures

- Closed-form versus Monte Carlo simulation
- envelope(*) function in R
  - Allows for user-authored functions
Module III review

- Basic concept: process $\rightarrow$ pattern
- Measures incorporate scale-dependence (unlike lattice measures)
- MC simulation for inference
- Extensions:
  - linear model as function of environment instead of kernel smooth.
  - alternative process specifications for stochastic component.
Appendix

Expanded distance measures

• Exploring Spatial Point Patterns
  ▪ Spatial variation in mean (1st order)
    \[ \lambda(s) = \frac{1}{\delta(s)} \sum \frac{1}{r} \kappa\left(\frac{s - s_i}{r}\right) \]
  ▪ Covariation: Nearest Neighbor
    \[ G(u) = n^{-1} \sum I(u_i < u) \]
    \[ F(x) = m^{-1} \sum I(z_i < x) \]
    \[ J(r) = \frac{1 - G(r)}{1 - F(r)} \]

• Exploring Spatial Point Patterns
  ▪ Spatial covariance (2nd order)
    \[ K(t) = (n(n-1))^{-1}|A| \sum_{i \neq j} w_{ij}^{-1} I(u_{ij} < t), u_{ij} = |x_i - x_j| \]
    \[ K_f(t) = |A| \sum_{i \neq j} w_{ij}^{-1} I(u_{ij} < t) \lambda(x_i) \lambda(x_j) \]

▪ Interpretation of K-function