DSC 383: Advanced Predictive Models for Complex Data

Section: Spatial Statistics > Subsection: Spatial Point Patterns and Processes

INSTRUCTOR:
Catherine (Kate) Calder
Department of Statistics & Data Sciences
University of Texas at Austin

LECTURE: Spatial Point Patterns

SOME DEFINITIONS

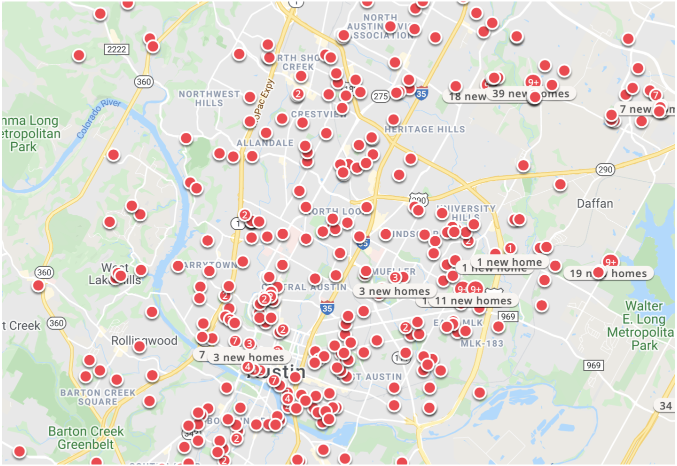
- **Definition:** A spatial point pattern (SPP) is a set of events (locations), irregularly distributed within a designated region, that can be viewed as being generated by some form of stochastic mechanism. (Diggle, 2003)
- Marks are attributes associated with the events. They can be categorical or continuous.
- Covariates are information that is explanatory of the spatial point pattern and/or marks, but not treated as part of the stochastic mechanism that generated the data.

EXAMPLE QUESTIONS

- How does the *intensity* of events vary across the study region?
- Do the events appear to *cluster*? Or, more generally, do they exhibit *stochastic dependence/interaction*?
- Does the intensity of events appear to depend on the value of a *covariate*? After controlling for the covariate, does there appear to be interaction between the events?
- For marked point patterns, does the distribution of marks vary across the study region (i.e., are the marks *segregated*)? How do the events with mark i interact with points of mark j?

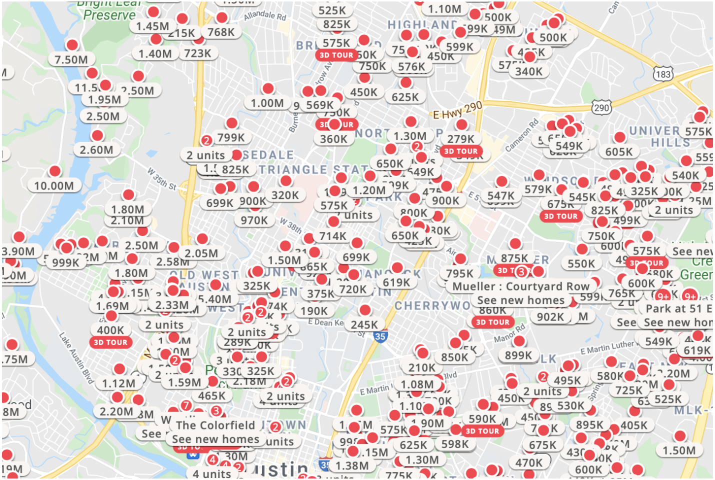
SOME EXAMPLES

- Houses for sale in Austin



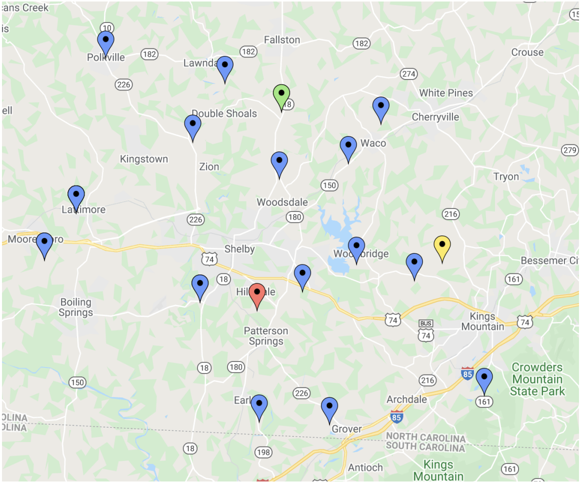
(Source: www.zillow.com)

- Houses for sale in Austin with prices



(Source: www.zillow.com)

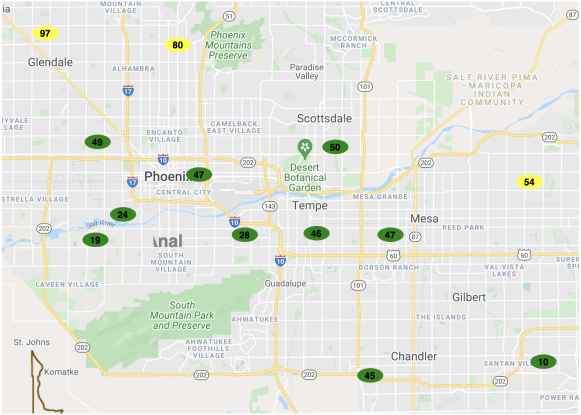
- Crimes in Cleveland County, NC



Crime Types: Property (blue); Assault (red); Missing person (yellow); Narcotics (green)

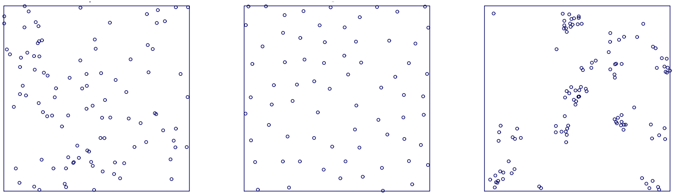
(Source: http://www.sheriffclevelandcounty.com)

- ~~Air quality readings from Maricopa County, AZ~~



(Souce: https://www.maricopa.gov/1643/Air-Quality-Status-and-Monitoring)

CHARACTERIZING SPPs



(Baddeley, 2010)

- **Complete Spatial Randomness**

LECTURE: Spatial Point Processes I

A SPATIAL POINT PROCESS

- **Definition:** A spatial point process is a stochastic process (probabilistic model) for random variables representing the locations of events in space
- A spatial point pattern is a realization of a spatial point process
- A spatial point process for CSR...

THE HOMOGENEOUS SPATIAL POISSON PROCESS

- **Definition**

a) The number of events N_A occurring within a finite region $A \subseteq D$, where D is the study area, is a random variable following a Poisson distribution with mean $\lambda|A|$ for some positive constant λ and $|A|$ denoting the area of region A

b) Given the total number of events N_A occurring within an area A, the location of the N_A events X_A represents an independent random sample of N_A locations, where each point location where an event could occur is equally likely to be chosen as an event

TESTING FOR CSR

- Quadrant Testing

  H_0 :

  H_1 :

  - Divide the study region D into m subregions (or, 'quadrants') $B_1, B_2, \ldots, B_m$ of equal area.
  - Let $N_j = n(x \in B_j)$ for $j = 1, \ldots, m$ be the observed number of events in quadrant j.
  - Under the null hypothesis, the expected number of events in quadrant j is $e_j = \hat{\lambda}|B_j|$ where $\hat{\lambda} = N/|D|$

- Then,

THE HETEROGENEOUS POISSON PROCESS

a) The number of events occurring within a finite region A is Poisson with mean $\int_A \lambda(s)ds$, where $\lambda(s)$ is a spatially varying function defined over $s \in D$

b) Given the total number of events N_A occurring with A, the N_A events represent an independent random sample of N_A locations, with probability of sampling a particular point s proportional to $\lambda(s)$

ESTIMATING $\lambda(\boldsymbol{s})$

1) **Kernel Smoothing**

The kernel density estimate of the intensity function λ at location $s_0 = (u_0, v_0) \in D$ $(D = D_1 \times D_2)$ is

$$\tilde{\lambda}(s_0) = E(s_0) \sum_{i=1}^{N} k\left(\frac{u_0 - u_i}{\sigma_u}\right) k\left(\frac{v_0 - v_i}{\sigma_v}\right)$$

Here, we are using a product kernel where $k(\cdot)$ is a one-dimensional function satisfying $\int_{\mathbb{R}} k(u)du = 1$. $E(s)$ is an edge correction:

$$E(s)^{-1} = \left[\int_{D_1} k(u)du\right]\left[\int_{D_2} k(v)dv\right]$$

- Common choices of $k(\cdot)$ include the Gaussian, uniform, triangle, quartic (biweight), and triweight kernels.
- Choosing the bandwidth (σ):

2) **Likelihood Estimation**

- The log likelihood of a *homogeneous* spatial Poisson process with intensity λ is

$$\log(L(\lambda; \mathrm{x})) = n(\mathrm{x}) \log(\lambda) - \lambda |D|$$

$\rightarrow$

2) **Likelihood Estimation cont.**

- The log likelihood of a *inhomogeneous* spatial Poisson process with parametric intensity function $\lambda_\theta(s)$ is

$$\log(L(\theta; \mathrm{x})) = \sum_{i=1}^{n} \log \lambda_\theta(x_i) - \int_D \lambda_\theta(u) du$$

$\rightarrow$

LECTURE: Spatial Point Processes II

CHARACTERIZING SPATIAL POINT PATTERNS

- So far, we have focused on first-order (mean) properties of spatial point processes.
- Now, we will talk about second-order (interaction) properties.

- **Pairwise distances**

  The distances between events

$$\{d_{ij} = |x_i - x_j| : i = 1, \ldots, n(A); j = 1, \ldots, n(A)\}$$

  is an *biased* sample of pairwise distances in the point process.

THE K FUNCTION

- **Definition:** The K function or reduced second moment measure (Ripley, 1977; Diggle, 1983) for a stationary point process, $\boldsymbol{X}$, is defined as

$$K(h) = \frac{1}{\lambda}\mathrm{E}\left[n(\boldsymbol{X} \cap b(u,h)\backslash\{u\}) \mid u \in \boldsymbol{X}\right]$$

  where $b(u,h)$ is a ball of radius h centered at arbitrary event u

- Intuitively, the K function is

- Ripley (1977) showed that specifying $K(h)$ for all h is equivalent to specifying $\mathrm{var}(N_A)$, the variance of the number of events occurring in subregion A

ESTIMATING THE K FUNCTION

1) Naive estimator:

$$\hat{K}(h) = \left(\frac{1}{\hat{\lambda}}\right) \frac{1}{N} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \delta(|x_i - x_j| < h)$$

2) Edge-corrected estimator:

$$\hat{K}_{ec}(h) = \left(\frac{1}{\hat{\lambda}}\right) \frac{1}{N} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij}^{-1} \delta(|x_i - x_j| < h)$$

where w_{ij} is a weight defined as the proportion of the circumference of the circle centered at event i with radius $|x_i - x_j|$ which lies in the study area

DIAGNOSTICS BASED ON THE K FUNCTION

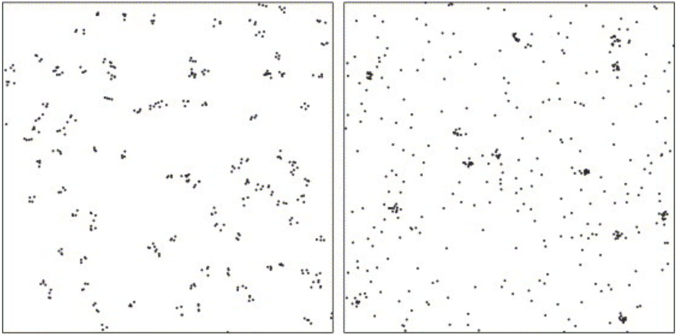
- If the spatial point pattern is a realization of a spatial Poisson process (CSR), then $K(h) = \pi h^2$.

- Strategy: compare the estimated K function, $\hat{K}_{ec}(h)$, to the theoretical K function under CSR, $K_{pois}(h)$

- Things to keep in mind:
  1. The K function is defined and estimated under the assumption of stationarity/homogeneity
  2. The K function *does not* completely characterize a spatial point process
  3. If the point process is not stationary/homogeneous, then differences between the theoretical and estimated K functions *do not* necessarily indicate interaction between events.

OTHER SPATIAL POINT PROCESSES

- **Neyman-Scott Process** (a type of **Poisson Cluster Process**)
  - The number/locations of the parents follow a homogeneous Poisson process (not observed)
  - The number of children per parent are independent and identically distributed according to the same distribution for each parent
  - The locations of the children around their respective parent are independent and identically distributed according to some bivariate probability distribution



(from Tscheschel and Stoyan, 2006)

- **Contagion/Inhibition Processes**

  These processes are modeled by directly specify the inter-event interactions in such as way that the occurrence of an event raises or lowers the probability of observing another event nearby

- **Cox Processes**

  Cox processes are similar to heterogeneous Poisson processes, but the intensity function $\lambda(s)$ is assumed to be random