Intensity Analysis of Spatial Point Patterns
Geog 210C
Introduction to Spatial Data Analysis

Chris Funk
Spatial Point Patterns

**Definition**
- Set of point locations with recorded “events” within study region, e.g., locations of trees, disease or crime incidents
- Point locations could correspond to all possible events or to subsets of them (mapped versus sampled point pattern)
- Attribute values could have also been measured at event locations, e.g., tree diameter (marked point pattern)

**Objective:** Introduce statistical tools for quantifying spatial variations in event intensity
Some Objectives in Spatial Point Pattern Analysis

- **Intensity analysis objectives**
  - characterize patterns of spatial distribution of event locations
  - determine *where* events are more likely to occur in space
  - investigate relationships between spatial “clusters” and nearby sources or other factors

- **Note**
  - The above should be contrasted with the case of spatially continuous data, where analysis objectives often involve characterizing spatial variation in attribute values recorded at a fixed set of locations/units
Outline

✧ Concepts & Notation
✧ Quadrat Counts for Intensity Estimation
✧ 1D Kernel Density Estimation
✧ 2D Kernel Density Estimation
✧ Points To Remember
Some Notation

✦ Point events
✦ Set of N locations of events occurring in a study area:

\[ \{ \mathbf{u}_i, \ i = 1, \ldots, N \}, \quad \mathbf{u}_i \in D \subset \mathbb{R}^K \]

\( \mathbf{u}_i \) = coordinate vector of \( i \)-th event location, e.g., in 2D \( \mathbf{u}_i = (x_i, y_i) \), \( \in \) = belongs to, \( D \) = study domain, a subset \( \subset \) of the \( K \)-dimensional space \( \mathbb{R}^K \)

✦ Variable of interest

✦ \( y(s) = \text{number} \) of events (a count) within arbitrary domain or support \( s \) with measure (length, area, volume) \( |s| \); support \( s \) is centered at an arbitrary location \( \mathbf{u} \) and can also be denoted as \( s(\mathbf{u}) \); in statistics, \( y(s) \) is treated as a realization of a random variable (RV) \( Y(s) \)

✦ In other words, we are interested in studying the spatial distribution of the count realizations \( y(s) \) generated from the respective RVs \( Y(s) \)
Intensity of Events

- **Local intensity** \( \lambda(u) \)
  - Mean number of events per unit area at an arbitrary location or point \( u \)
    \[
    \lambda(u) = \lim_{|s| \to 0} \left\{ \frac{\mathbb{E}\{Y(s)\}}{|s|} \right\}, \quad u \in D
    \]
    - where \( \mathbb{E}\{Y(s)\} \) denotes the expectation (mean) of RV \( Y(s) \) defined over a region \( s(u) \) centered at \( u \).

- **Overall intensity** \( \lambda \)
  - Estimated as
  - Where where \( |D| = \) measure (area) of study region \( D \)

- **Region-specific intensity**
  - For a constant intensity \( \lambda(u) = \lambda \) the expected number of events within a region \( s \) is just a function of \( |s| \)
    - i.e., \( \mathbb{E}\{Y(s)\} = \lambda |s| \);
Local Intensity Estimation via Quadrat Counts I

Flowchart

1. partition \( D \) into \( L \) sub-regions or quadrats \( \{s_l ; l = 1 \ldots L\} \) of equal measure \( |s_l| \)
2. count number of events \( N(s_l) \) in each quadrat \( s_l \)
3. convert sample counts into estimated intensity values as:

\[
\hat{\lambda}(s_l) = \frac{N(s_l)}{|s_l|}
\]

Similar to a 2D or bivariate histogram, with quadrats playing the role of rectangular bins and coordinates playing the role of data from 2 attributes.
Characteristics

- estimated intensities $\lambda(s_i)$ over set of quadrats
- intended for revealing large-scale patterns in intensity variation over $D$
- larger quadrats yield smoother intensity maps; smaller quadrats yield "spiky" intensity maps with empty quadrats
- sliding and overlapping quadrats, as well as randomly placed quadrats, can also be used
- size, origin, and shape of quadrats matters a lot
Consider a hypothetical 1D point pattern comprised of $N = 10$ events (left) and estimate their local intensity, i.e., a 1D profile of average # of events per unit area:

**Statistical analogy**

- The objective is to describe the density of x-coordinates, and this problem has been treated extensively in the non-parametric density estimation literature; a first-cut at such a density profile is provided by the density histogram plot (right).

- **In other words**: the set of $N$ x-coordinates of events in a 1D point pattern can be viewed as $N$ values of an attribute, here the x-coordinate . . .
Key concepts

- Density estimation via a histogram calls for deciding on: (i) the number of attribute classes (bins), and (ii) their centers in the abscissa.
- Instead of choosing a limited number of bins, choose as many bins as the number of events in the data set.
- Bars in a histogram are rectangular, but nothing prevents us from using other shapes to build a density profile...

Another key concept: Each bar (left) or triangle (right) can be regarded as the influence of an observed event to the likelihood of seeing other events around that observed one...
Kernel function

Analytical expression for likelihood of a particular x-coordinate, or in other words for probability of observing an event at the particular x-coordinate, given presence of an event at coordinate $x_i$.

Kernel characteristics

Function of distance $h_i = |x - x_i|$ between arbitrary point at location $x$ and event at location $x_i$:

$$k(x, x_i) = k(|x - x_i|) = k_i(h),$$

where $h$ is the distance between an arbitrary location $x$ and the kernel center, here the event location $x_i$ (assumed to be at $x = 0$ on the graph).

Typically all $N$ kernels are assumed the same, i.e.,

$$k_i(h) = k(h),$$

for all $i$. 
Kernel characteristics II

- Kernels are (typically symmetric) probability density functions (PDFs), hence non-negative and integrating to 1: \( k(h) \geq 0 \), and \( \int k(h) dh = 1 \)

- As PDFs, kernels have a mean (0, since the abscissa quantizes distance from an event) and positive finite variance: \( \int hk(h) dh = 0 \) and \( 0 < \int h^2 k(h) dh < \infty \)

Relation to density estimation

- Instead of fixing the # of bins and their origin (as done with histograms) we can estimate the local density \( f(x) \) at an arbitrary \( x \)-value as a weighted \textbf{sum} of \( N \) values \( k(x-x_i) \); each such value belongs to a different kernel \( k_i(h) \) centered at a \( x_i \) location/coordinate
Some 1D Kernel Functions I

- **rectangular or uniform or Parzen:**

  \[ k(h) = \begin{cases} 
  1/2 & \text{if } h \in [-1, 1] \\
  0 & \text{if not} 
  \end{cases} \]

- **triangular:**

  \[ k(h) = \begin{cases} 
  1 - |h| & \text{if } h \in [-1, 1] \\
  0 & \text{if not} 
  \end{cases} \]

- For a set of \( P \) values \( \{x_p; p = 1\ldots P\} \) discretizing a 1D segment, and for a particular datum coordinate \( x_i \), the function \( k(x_p-x_i) \) can be evaluated \( P \) times, and the resulting kernel "profile" can be stored in a \((P \times 1)\) array \( k_i = [k(x_p-x_i), p=1\ldots P]^T \)
Some 1D Kernel Functions II

- **Quadratic or Epanechnikov:**

  
  \[ k(h) = \begin{cases} 
  \frac{3}{4} (1 - h^2) & \text{if } h \in [-1, 1] \\
  0 & \text{if not} \end{cases} \]

- **Gaussian:**

  
  \[ k(h) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}h^2\right) \]

Kernels that reach 0 asymptotically, e.g., Gaussian, are called non-compact kernels.
Alternative view of a kernel

A kernel function $k(x-x_i)$ quantifies the "influence" of a particular event at coordinate $x_i$ to its surroundings, i.e., to all other $x$-locations.

Scaling the kernel

The influence of an event at $x_i$ to all $x$-coordinates can be altered by scaling the associated kernel function $k(x-x_i)$; i.e., by dividing the function argument $x-x_i$ by a constant $b$ (called the kernel bandwidth); in order to ensure that the new kernel is a PDF, i.e., integrates to 1, divide the output of this new function by $b$.

E.g.: scaled or non-standard Gaussian PDF: $k(h; b) = \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{h}{b}\right)^2\right]$, where $b$ is the standard deviation.
Scaled 1D Kernels II

Scaled kernel function

- Divide the argument (distance from event) of the kernel function by a scalar b:

\[ k(x - x_i; b) = \frac{1}{b} k \left( \frac{x - x_i}{b} \right) \]

Transformation of PDFs

- Let X be a RV with PDF \( f_X(x) \) and Y be another RV defined as \( Y = \frac{1}{b}X \), i.e., \( y = \frac{x}{b} \).
- The PDF \( f_Y(y) \) of RV Y can be computed as:
  \[ f_Y(y) = (1/b) f_X(x/b); \text{ if the original PDF } f_X(x) \text{ has std deviation 1, the new PDF } f_Y(y) \text{ has std deviation } b \]

For a set of P values \( \{x_p ; p = 1...P\} \) discretizing a 1D segment, and for a particular datum coordinate \( x_i \) the scaled function \( k(x_p - x_i; b) \) can be evaluated P times, and the resulting discrete kernel stored in a (Px1) array \( k_i(b) = [k(x_p - x_i; b); p = 1...P]^T \)
1D Kernel Density Estimation Flowchart

1. Choose a kernel function $k(x-x_i)$, i.e., a PDF, and a bandwidth parameter $b$ controlling kernel extent and consequently the "smoothness" of the final estimated density profile $f(x)$; this amounts to choosing a scaled kernel function $k(x-x_i;b)$

2. Discretize 1D segment, i.e., choose a set of $P$ x-coordinates $\{x_p; p = 1…P\}$ at which the density function $f(x)$ will be estimated

3. For each datum coordinate $x_i$, evaluate the scaled kernel function $k(x_p-x_i;b)$ for all $P$ x-values; this yields $N$ scaled kernel profiles $\{k_i(b); i = 1…N\}$ each one stemming from a particular event coordinate $x_i$

4. For each discretization coordinate $x_p$, compute estimated density $f(x_p)$ as the sum of the $N$ scaled kernel values $k(x_p-x_i; b)$, after weighting each such value by $1/N$:

$$\hat{f}(x_p) = \sum_{i=1}^{N} \frac{1}{N} \frac{1}{b} k \left( \frac{x_p - x_i}{b} \right) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{b} k \left( \frac{x_p - x_i}{b} \right)$$

Output

A $(P\times1)$ vector $k(b)$ with estimated density values $f(x)$ at the specified x-coordinates; the $N$ scaled & weighted kernels $\{(1/N)k_i(b); i = 1…N\}$ can be regarded as $N$ elementary profiles whose super-position builds up the final estimated density profile
1D Kernel Density Estimation Examples

Rules exist for choosing an "optimal" bandwidth parameter, typically based on a presupposed distribution type, e.g., Gaussian, for the N data ... Estimated density profiles are more sensitive to choice of bandwidth parameter b than to choice of kernel type ...
The smaller the bandwidth, the spikier (noisier) the resulting estimated density profile; too large a bandwidth leads to over-smoothed (with no interesting details) density profiles…

Dotted lines depict scaled and weighted (by \(1/N\)) kernel profiles
Separable 2D Kernels

- Two 1D Gaussian kernels
  \[ k_x(x - x_i; b_x) = \frac{1}{b_x \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - x_i}{b_x} \right)^2 \right] \]
  \[ k_y(y - y_i; b_y) = \frac{1}{b_y \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{y - y_i}{b_y} \right)^2 \right] \]
  event location \( u_i = (x_i, y_i) \), arbitrary location \( u = (x, y) \), kernel bandwidths \( b_x \) and \( b_y \)

- 2D composite kernel
  \[ k(x - x_i, y - y_i; b_x, b_y) = \frac{1}{2\pi b_x b_y} \exp \left[ -\frac{1}{2} \left( \frac{x - x_i}{b_x} \right)^2 - \frac{1}{2} \left( \frac{y - y_i}{b_y} \right)^2 \right] \]
  bivariate Gaussian PDF for 2 independent RVs, a product of 2 univariate Gaussian PDFs

- Separability
  - Any (scaled or not) 2D kernel that can be derived as a product of 2 elementary 1D kernels is called separable
Constructing A Separable 2D Kernel

- Two 1D Gaussian kernels for the x- and y-dimensions

![Gaussian PDF](image)

- Replicated 1D Gaussian kernels and 2D separable composite

![Replicated Gaussian PDF](image)

Anisotropic kernel = multidimensional kernel with different bandwidths along different directions
2D Gaussian Kernel Examples

Isotropic kernel = multidimensional kernel with same bandwidth along different directions
Multidimensional separable kernels

K-dimensional kernels can be constructed in $\mathbb{R}^K$ as products of $K \geq 2$ 1D kernels:

$$f(u - u_i; b) = \prod_{k=1}^{K} \frac{1}{b_k} f \left( \frac{u_k - u_{ik}}{b_k} \right)$$

- $u_k$ = k-th coordinate of location $u$; $u_{ik}$ = k-th coordinate of event location $u_i$
- $b_j$ = bandwidth along k-th dimension; $b$ = vector with K bandwidths

Estimated density

- Apply the same computational flowchart as in 1D:

$$\hat{f}(u) = \sum_{i=1}^{N} \frac{1}{N} f(u - u_i; b) = \frac{1}{Nb_1 \cdots b_K} \sum_{i=1}^{N} \left[ \prod_{k=1}^{K} f \left( \frac{u_k - u_{ik}}{b_k} \right) \right]$$

- Similar rules with 1D case exist for choosing “optimal” bandwidths along each dimension
2D Kernel Intensity Estimation I

1. center a circle $C(u; b)$ of radius $b$ at any arbitrary location $u$ in $D$
2. estimate local intensity at $u$ as:
   $$\lambda(u) = N(u; b) / |C(u; b)|$$
   where $N(u; b) = \#$ of events within $C(u; b)$
   $|C(u; b)| = \text{kernel measure, } b^2 \text{ in 2D.}$

Note: steps 1 and 2 amount to choosing a 2D kernel function that plots like a cylinder with base radius $b$ and height $1/(b^2)$
3. repeat estimation for set of points (typically arranged at the grid nodes of a regular raster) in the study region to create an intensity map

Looping over # of grid nodes (P) instead over # of events (N), yields same results as kernel density estimation case
conversion of point events to raster format, used for visualizing spatial patterns in event intensity and for detecting “hot spots"

resulting raster surface reveals large-scale patterns in intensity variation

larger kernel bandwidth $b$ yields smoother intensity maps; reverse true for smaller bandwidths

could define local bandwidth $b(u)$ as a function of presence of events in neighborhood of $u$; this is termed adaptive density estimation

different kernels, e.g., quadratic, give more weight to nearby events when calculating $\lambda(u)$

Ideally, one should use some theoretical basis for selecting an appropriate kernel, e.g., a Gaussian kernel is suitable for diffusion-type processes
Example – Lung & Larynx cancer in Lancashire county

- From Bailey & Gatrell (P. 129-132), images from Tony Smith’s Spatial Data Analysis notebook
  [http://www.seas.upenn.edu/~ese502/notebook](http://www.seas.upenn.edu/~ese502/notebook)
- Based on data collected between 1974-1983
- Blue dots are lung cancer cases
  Red dots are larynx cancer cases
Example – Lung & Larynx cancer in Lancashire county

Just how likely was such a cluster of cancer cases located in the sparsely populated south?
Recap

- **Event intensity of spatial point patterns**
  - $\lambda(u)$: mean # of events over a unit area centered at $u$
  - estimated overall intensity $\lambda = N/|D|$  
  - local intensity via quadrat counts or kernel density estimation

- **Kernel intensity estimation**
  - conversion of point data (events) to raster format (intensity surface)
  - statistical multidimensional (multivariate) density estimation methods are used to estimate local intensity $f(u)$. Note: Density surface integrates to 1, so multiply every such estimate $f(u)$ by $N$ to convert it to an intensity value $f(u)$
  - resulting intensity surface depends on: (i) kernel type, and (ii) bandwidth; the latter is more influential
  - alternative approaches for non-parametric multivariate density estimation include: k-nearest neighbor and mixture of Gaussian densities methods
  - intensity surface can be linked (via regression models) to explanatory variables, e.g., disease occurrence intensity as function of air quality variables
  - An introduction to R - Venables and Smith (on site).