Determining the Optimal Search Area for a Serial Criminal

Towson University
Applied Mathematics Laboratory

Dr. Mike O'Leary

Modeling and Simulation Technical Working Group
Jersey City, NJ
March 2006

Applied Mathematics Laboratory

- Looks for undergraduate research projects in mathematics.
- Established at Towson University in 1980.
- We form teams of 2-6 undergraduate students, led by 1-2 faculty members and potentially one M.S. student.
- Usually, we work on one research problem each year.

Recent Projects

- Carroll Area Transit System (2004-2005)
- Baltimore County Department of Environmental Planning and Resource Management (2003-2005)
- Baltimore City Fire Department (2002-2003)

2005-06 Participants

- Dr. Coy L. May, Dr. Andrew Engel, Dr. Mike O'Leary
- Paul Corbitt
- Brandie Biddy, Brooke Belcher, Greg Emerson, Laurel Mount, Ruozhen Yao, Melissa Zimmerman

The Question

- What is the optimal search area for a serial criminal?

Centrographic Measures

- Centroid
- Center of minimum distance
- Center of the circle
- Harmonic mean
- Geometric mean

Probability Distance Strategies

- Suppose we have a series of linked crimes committed at points $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n$.
- For any point $\vec{z}$, the relative likelihood function $\sigma(\vec{z})$ is

$$\sigma(\vec{z}) = \sum_{i=1}^{n} f(d(\vec{x}_i, \vec{z}))$$

Here f is a probability density function and $d(\vec{x}_i, \vec{z})$ is the distance between $\vec{x}_i$ and $\vec{z}$.

Probability Distance Strategies

- Rossmo

$$\sigma(\vec{z}) = k\sum_{i=1}^{n} \frac{\phi}{\left(\left|x_i^{(1)} - z^{(1)}\right| + \left|x_i^{(2)} - z^{(2)}\right|\right)^f} + k\sum_{i=1}^{n} \frac{(1-\phi)B^{g-f}}{\left(2\mathrm{B} - \left|x_i^{(1)} - z^{(1)}\right| - \left|x_i^{(2)} - z^{(2)}\right|\right)^g}$$

- Here k, B, f, g are all empirical constants.
- $\phi = 0$ if $\vec{z}$ is in a buffer zone of size B around $\vec{x}_i$, while $\phi = 1$ if $\vec{z}$ is outside the buffer zone.

Our Approach

- Postulate: There is an *a priori* function $P(\vec{x};\vec{z},\vec{\beta})$ that gives the probability that an offender with anchor point $\vec{z}$ commits a crime at the point $\vec{x}$.
  - $\vec{\beta}\in\mathbb{R}^k$ represents additional parameters.
  - Both $\vec{z}$ and $\vec{\beta}$ are unknown.
- Given: A series of crimes committed at points $\vec{x}_1, \vec{x}_2, \cdots \vec{x}_n$.
- What is the best estimate of the anchor point $\vec{z}$? What is the best estimate of $\vec{\beta}$?

Maximum Likelihood Estimate

- The maximum likelihood estimate for $\vec{z}$ is

$$\vec{z} = arg_{\vec{z}}\, max \prod_{i=1}^{n} P(\vec{x}_i; \vec{z}, \vec{\beta})$$

- We only assume that $P(\vec{x}; \vec{z}, \vec{\beta})$ has a particular form, and then choose the parameter(s) that best match the data.

Advantages

- We can explicitly incorporate other features into the model by choosing $P(\vec{x};\vec{z},\vec{\beta})$ judiciously.
  - Geography
  - Demographics
  - Jurisdictional boundaries
  - Buffer zones
- The MLE method applies regardless of the precise form of $P(\vec{x};\vec{z},\vec{\beta})$.

Questions

- What is the right form for $P(\vec{x};\vec{z},\vec{\beta})$?
  - Ideally, this should be calculated from empirical data.
  - Are there reasonable choices?
    - Start with the simplifying assumption: geography is homogeneous.

Gaussian Distribution

- If we assume P is Gaussian:

$$P(\vec{x};\vec{z},\sigma)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{|\vec{x}-\vec{z}|^2}{2\sigma^2}\right)$$

  then the maximum likelihood estimate of $\vec{z}$ is exactly the centroid.

- This remains true if we allow P to have variance $\vec{\sigma}=(\sigma_1,\sigma_2)$ and correlation ρ.

Exponential Distribution

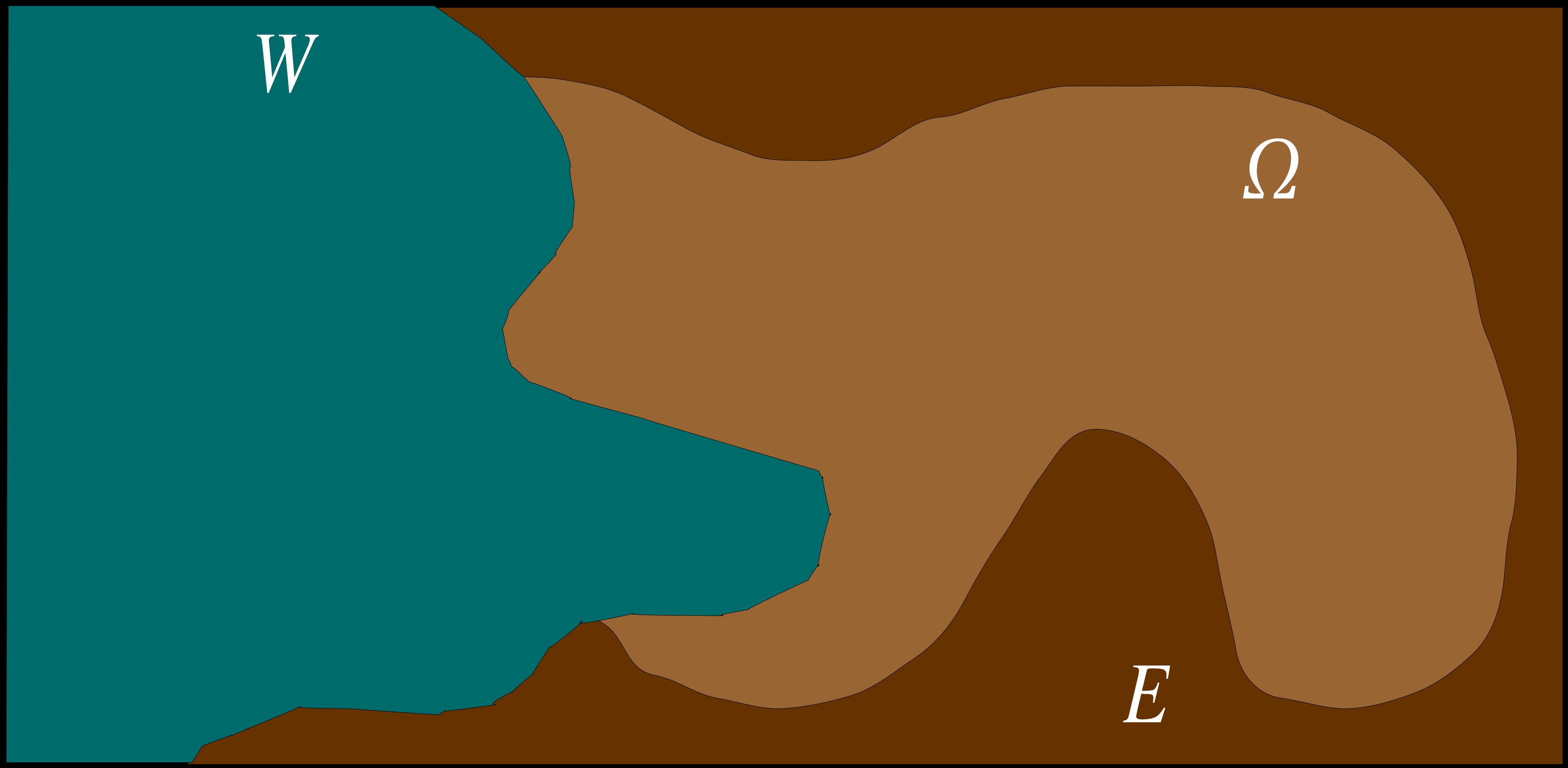
- If we assume that P is exponential

$$P(\vec{x};\vec{z},\beta)=\frac{1}{2\pi\beta^2}\exp\left(-\frac{|\vec{x}-\vec{z}|}{\beta}\right)$$

  then the maximum likelihood estimate for $\vec{z}$ is the center of minimum distance.
- This remains true if $|\vec{x}-\vec{z}|$ is replaced by another distance function $d(\vec{x},\vec{z})$.

Geography

- Add some simple geographic features.



Geography

- Regions
  - Ω: Jurisdiction(s). Crimes and anchor points may be located here.
  - E: "elsewhere". Anchor points may lie here, but we have no data on crimes here.
  - W: "water". Neither anchor points nor crimes may be located here.
- In all other respects, we assume the geography is *homogeneous*.

Geography

- Let π be the distance dependence; for example we can use the Gaussian
$$\pi(s;\beta)=\exp(-s^2/\beta).$$
- We would like to define

$$P(\vec{x};\vec{z})=\pi(|\vec{x}-\vec{z}|)=\exp\left(\frac{-|\vec{x}-\vec{z}|^2}{\beta}\right) \text{ if } \vec{x}\in\Omega,\vec{z}\in\Omega\cup E$$

$$P(\vec{x};\vec{z})=0 \text{ if } \vec{x}\notin\Omega \text{ or } \vec{z}\in W.$$

- This needs to be normalized to become a probability distribution.

Geography

- Thus we have

$$P(\vec{x};\vec{z},\vec{\beta})=\frac{\pi(|\vec{x}-\vec{z}|)}{\iint_{\Omega}\pi(|\vec{y}-\vec{z}|)\,dy_1\,dy_2}=\frac{\exp\left(\frac{-|\vec{x}-\vec{z}|^2}{\beta}\right)}{\iint_{\Omega}\exp\left(\frac{-|\vec{y}-\vec{z}|^2}{\beta}\right)dy_1\,dy_2}$$

- for $x\in\Omega, z\in\Omega\cup E$ and

$$P(\vec{x};\vec{z})=0$$

otherwise.

Implementation

- Given: a sequence of crimes committed at $\vec{x}_1, \vec{x}_2, \cdots \vec{x}_n$.
- Assumption: The distance dependence π is Gaussian.
- Assumption: Any two crime locations in Ω are equiprobable, and no (known) crimes occur outside Ω.
- Assumption: Any two anchor points in $\Omega \cup E$ are equiprobable, and no anchor points occur in W.

Implementation

- Find the choice of $\vec{z}$ that solves

$$\max_{\vec{z}\in\Omega\cup E}\prod_{i=1}^{n}\frac{\pi\left(\left|\vec{x}_i-\vec{z}\right|\right)}{\iint_{\Omega}\pi\left(\left|\vec{y}-\vec{z}\right|\right)dy_1\,dy_2}$$

$$=\max_{\vec{z}\in\Omega\cup E}\frac{\exp\left(-\frac{1}{\beta}\sum_{i=1}^{n}\left|\vec{x}_i-\vec{z}\right|^2\right)}{\left[\iint_{\Omega}\exp\left(\frac{-\left|\vec{y}-\vec{z}\right|^2}{\beta}\right)dy_1\,dy_2\right]^n}$$

Implementation

- The student team is:
  - Implementing this algorithm in Python, and
  - Linking the result to ArcGIS so that they can be used together.

Current Status

- Write geographic data from ArcGIS into a file.
  - Done
- Writing programs that can plot points in ArcGIS.
  - Done
- Reading geographic data from a file into Python.
  - Testing
- Re-writing the double integral as a line integral to ease computation.
  - Done

Current Status

- Numerical evaluation of the line integrals in Python. — Testing
- Python code to determine if a point is in a set. — Testing
- Choosing an algorithm to find the optimum. — In progress
- Implementing the optimization algorithm in Python. — Not yet

Current Status

- Integrating all of the parts into one program.
- Testing.
  - Compare results against empirical data (to be provided by Phil Canter, Baltimore County Police Department)

- Not yet
- Not yet

Pre-preliminary results

Next Steps

- Look at other distance-decay models.
  - We can explicitly model buffer zones with this method.

Next Steps

- When calculating a proposed anchor point, also compare the best estimates of the parameters $\vec{\beta}$ with historical / empirical data.
  - Bad fits of the parameters might suggest times when the model is inappropriate.

Next Steps

- Look directly at empirical data to determine the proper form for $P(\vec{x};\vec{z},\vec{\beta})$.
  - What is the “right” distance-decay function?
- Allow $P(\vec{x};\vec{z},\vec{\beta})$ to depend on more nuanced geographical features?
  - *e.g.* population density

$$P(\vec{x};\vec{z},\vec{\beta})=\pi(|\vec{x}-\vec{z}|)\rho(\vec{x})$$

Next Steps

- Bayesian analysis may let us determine the probability density function for the criminal's anchor point, rather than a point estimate.

Questions?

Contact information:

Dr. Mike O'Leary

Director, Applied Mathematics Laboratory

Towson University

Towson, MD 21252

410-704-7457

moleary@towson.edu