Patterns in Offender Distance Decay and the Geographic Profiling Problem.

Mike O'Leary

Towson University

2011 Crime Mapping Conference
Miami, FL
October 19-21, 2011

Acknowledgments

- This work has been supported by the National Institute of Justice, through grant 2009-SQ-B9-K014
- We thank the Baltimore County Police Department for providing the data used in this study
- We thank Phil Canter from the Baltimore County Police Department for his assistance.

The Geographic Profiling Problem

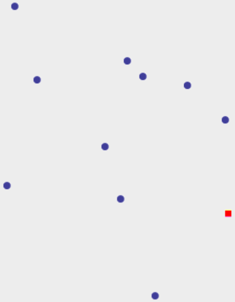
- The geographic profiling problem is to estimate the location of the home base of a serial criminal from the known locations of the elements of the offender's crimes.
  - The home base is also called the anchor point of the offender. It may be the offenders home, the home of a relative, a place of work, or even a favorite bar.
- We have developed a new tool for the geographic profiling problem.
  - It is free for download and use, and is entirely open source.
    - `http://pages.towson.edu/moleary/Profiler.html`
  - It is still in the prototype stage.

Circle Theory

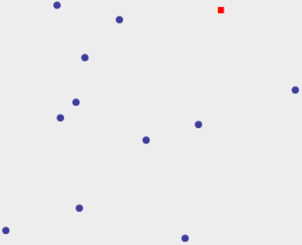
- Canter's Circle hypotheses[1]: Given a series of crimes, construct the circle whose diameter is the segment connecting the two crimes that are farthest apart.
  - If the offender is a *marauder*, then their anchor point will lie in this circle.
  - If the offender is a *commuter*, then their anchor point will lie outside this circle.
    - Note that all of the crimes are not necessarily within the circle.
- For crimes like rape and arson, there is evidence that most offenders are marauders; for crimes like residential burglary the evidence shows a mixture of marauders and commuters.
- This is a binary approach- either someone is a commuter or they are a marauder.
  - This binary approach may not be suitable in many cases.

[1] Canter D. & Larkin, P. (1993). The environmental range of serial rapists. Journal of Environmental Psychology, 13, 63-69.

Which is the Commuter?



or



- Here the crime locations are blue points, and the offender's anchor point is a red square.

Commuters & Marauders

- We have created a different way to differentiate between commuters and marauders.
- Suppose that:
  - The crimes are at $x_1, x_2, \ldots, x_n$;
  - The offender's anchor point is z.
- For $1 \leqslant p < \infty$ define

$$\mu_p = \left[\frac{\min_{\mathbf{y}} \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{y})^p}{\sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{z})^p} \right]^{1/p}$$

- Note that $0 \leqslant \mu_p \leqslant 1$.
- Offenders with small μ_p correspond to μ_p-commuters, while offenders with large μ_p correspond to μ_p-marauders.

- If $p = 1$, then

$$\min_{\mathbf{y}} \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{y}) = d(\mathbf{x}_i, \mathbf{y}_{\text{cmd}})$$

and

$$\mu_1 = \frac{\sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{y}_{\text{cmd}})}{\sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{z})}$$

where $\mathbf{y}_{\text{cmd}}$ is the center of minimum distance of the crime series.

Commuters & Marauders

- If $p = 2$, and d is Euclidean distance, then

$$\min_{\mathbf{y}} \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{y})^2 = d(\mathbf{x}_i, \mathbf{y}_{\text{centroid}})^2$$

and

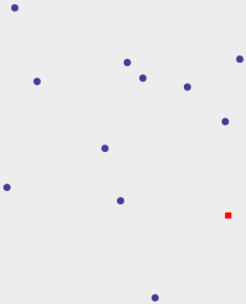
$$\mu_2 = \sqrt{\frac{\sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{y}_{\text{centroid}})^2}{\sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{z}_{\text{anchor}})^2}}$$

where

$$\mathbf{y}_{\text{centroid}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$$

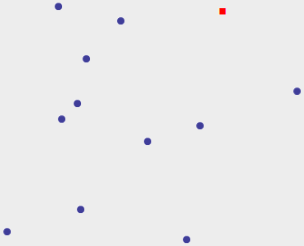
is the centroid of the crime series

Which is the Commuter?



$\mu_2 = 0.58$
(Canter Marauder)

or



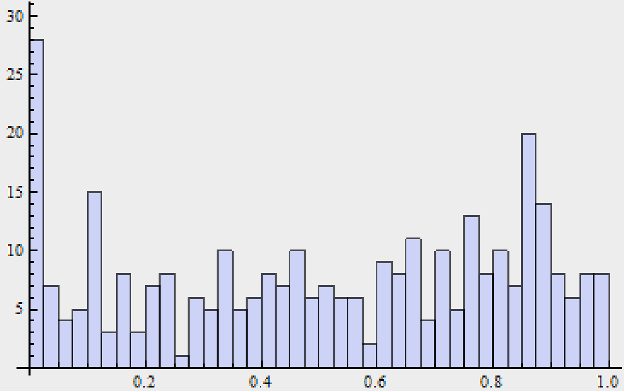
$\mu_2 = 0.56$
(Canter Commuter)

Data

- We have data for residential burglaries in Baltimore County
  - 5863 solved offenses from 1990-2008
  - We have 324 crime series with at least four crimes
    - A series is a set of crimes for which the Age, Sex, Race, DOB and home location of the offender agree.
  - The average number of elements in a series is 8.1, the largest series has 54 elements.
- We have data for non-residential burglaries in Baltimore County
  - 2643 solved offenses from 1990-2008
  - We have 167 crime series with at least three crimes.
  - The average number of elements in a series is 7.87, the largest series has 111 elements.
- We have data for bank robberies in Baltimore County
  - 602 solved offenses from 1993-2009.
  - We have 70 crime series with at least three crimes.
  - The average number of elements in a series is 4.51, the largest series has 15 elements.

Commuters & Marauders

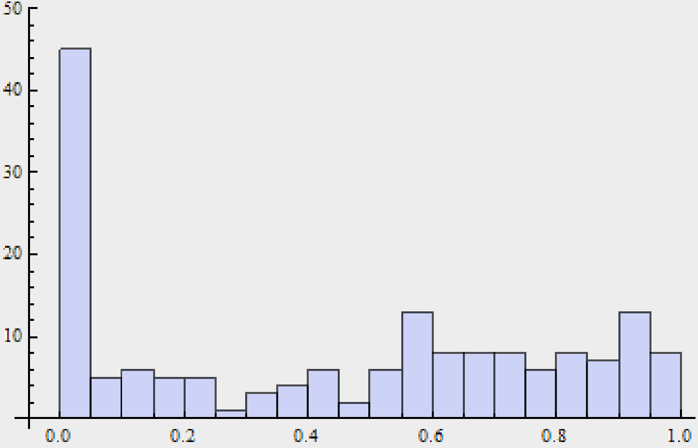
- What is the distribution of μ_2 commuters and marauders for residential burglary?



- There does not appear to be a sharp distinction between commuters and marauders in this data

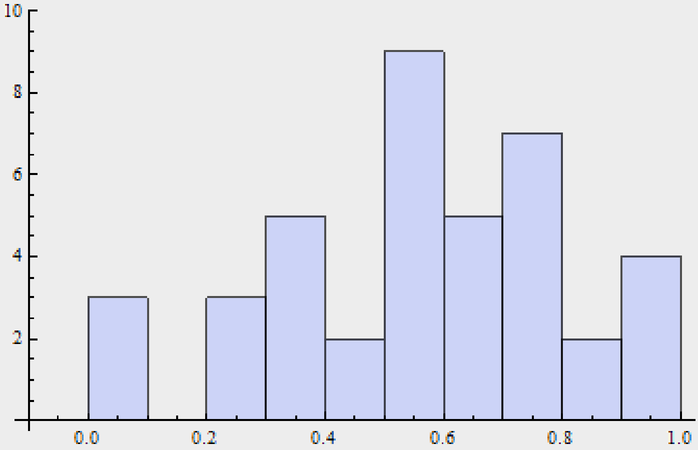
Commuters & Marauders

- Non-residential burglary shows a decided preference for commuters.



Commuters & Marauders

- Bank robberies show a slight preference towards marauders.

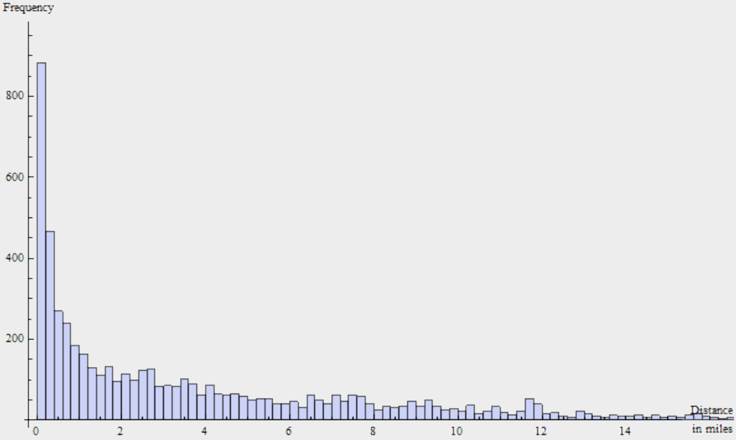


Distance Decay

- The distance decay patterns of offenders are of fundamental importance in the geographic profiling problem.
- Though we have data for the distance from the offenders home to the offense site for a large number of solved crimes, we cannot directly use this information to draw inferences about the behavior of any individual offender.
    - To do so is to commit the *ecological fallacy*.
- There are two sources of variation- the variation within each individual, and the variation between individuals.
    - If all of the individuals behaved in the same fashion, then the aggregate data can be used to draw inference about the (common) underlying behavior.

Distance Decay

- Distance from home to offense site for residential burglary



Distance Decay

- If the only quantity that varies between offenders is the average offense distance, then the resulting scaled distances should exhibit the same behavior regardless of the offender.
  - In particular, this will allow us to aggregate the data across offenders and draw valid inference about the (assumed) universal behavior.
- For each serial offender with crime sites $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ and home $\mathbf{z}$, estimate the average offense distance α by

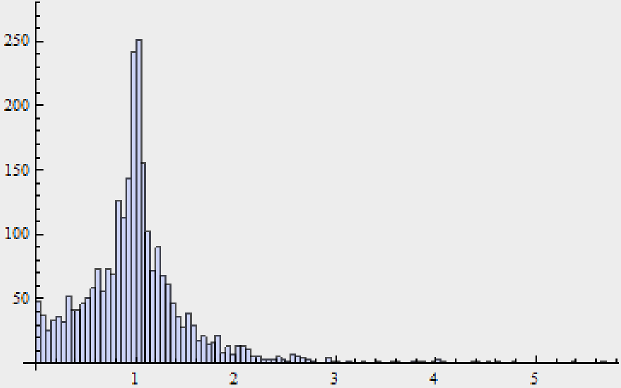
$$\hat{\alpha}_h = \frac{1}{n} \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{z})$$

and now consider the set of scaled distances

$$\rho_i = \frac{d(\mathbf{x}_i, \mathbf{z})}{\hat{\alpha}_h}$$

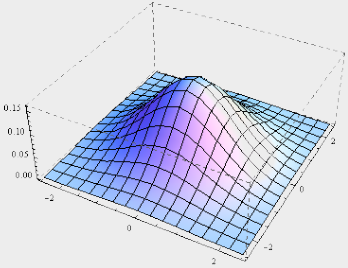
Distance Decay

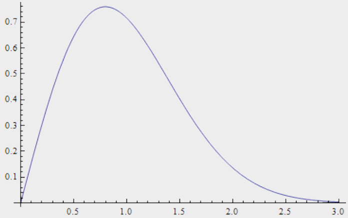
- What do we obtain when we plot a histogram not for offense distance, but for scaled distance?



Distance Decay

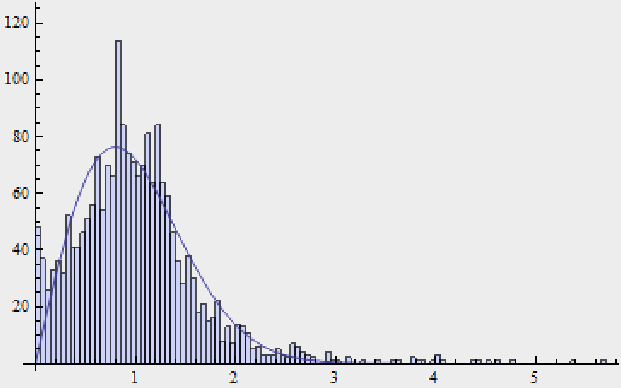
- When considering distance, it is important to realize that it is a *derived* quantity.
  - Offenders do not select a distance- they select a target.
- For example, if the offender selects a target from a two-dimensional normal distribution; then the distribution of distances is a Rayleigh distribution.





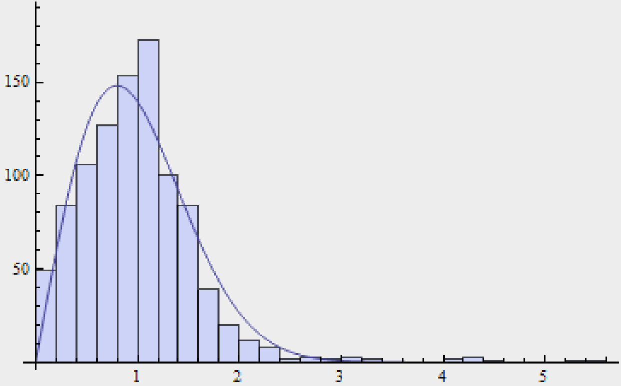
Distance Decay

- Focus our attention only on non-commuters- say $\mu_2 \geqslant 0.25$, and compare the result to a Rayleigh distribution



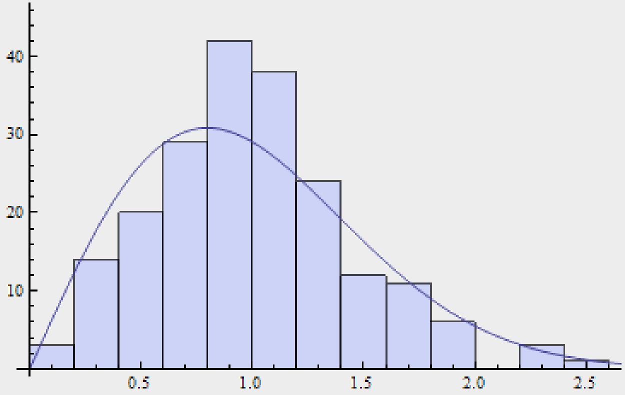
Distance Decay

- The agreement with the Rayleigh distribution does not appear to be happenstance. Here is what occurs for non-residential burglaries with $\mu_2 \geqslant 0.25$



Distance Decay

- Here is what occurs for bank robberies with $\mu_2 \geqslant 0.25$



Distance Decay

- It is possible that these fits are caused by something peculiar to the geography of Baltimore County.
- However, we are not the first to examine scaled distances.
  - Warren, Reboussin, Hazelwood, Cummings, Gibbs, and Trumbetta (1998). *Crime Scene and Distance Correlates of Serial Rape*, Journal of Quantitative Criminology **14** (1998), no. 1, 3559.
  - In that paper, they graphed scaled distances for serial rape:

Distance Decay

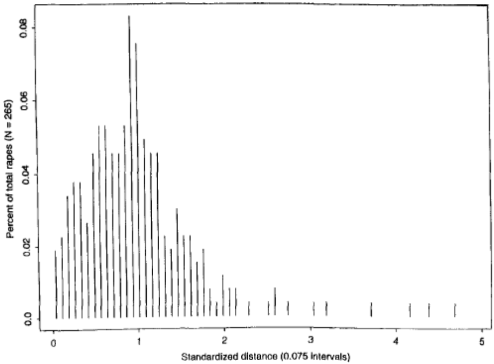
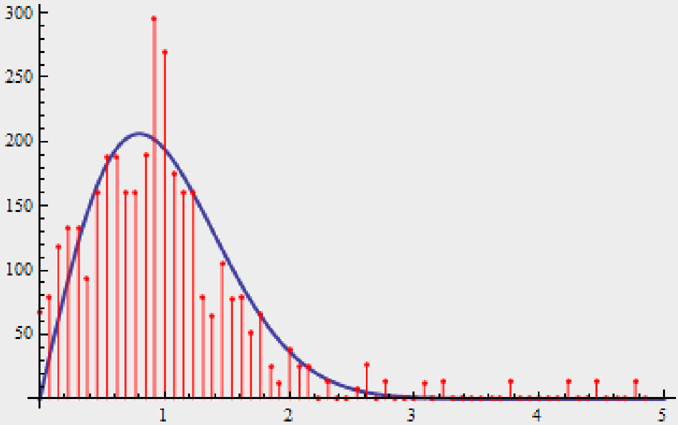


Fig. 2. Proportion of rapes by standardized distance from residence to rape location. Cases with five or more rapes.

two reasons. First, the nonrepresentative nature of the data diminishes the meaningfulness of significance levels. Second, the applied purpose of the paper heightened the need to present the data in a visually clear and practically interpretable form. Distance was found to vary with the demographic characteristics of the offender as well as certain "signature" and "modus

Distance Decay

- Our Rayleigh distribution with mean 1 appears to fit this data as well:

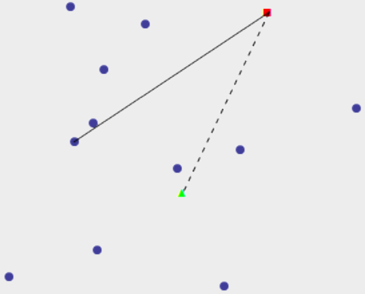


Distance Decay- Caveats

- It is important to note that, though compelling, these graphs do not provide justification that offenders follow a bivariate normal distribution.
  - Agreement is necessary, but not sufficient for this conclusion.
  - There are other two dimensional distributions whose distribution of distances also is Rayleigh.
- We still do not understand the situation yet with commuters.
  - The Warren *et. al.* data is for serial rape, which is known to be well approximated by circle theory- suggesting that this data set may be weighted away from commuters, which our theory does not yet handle.

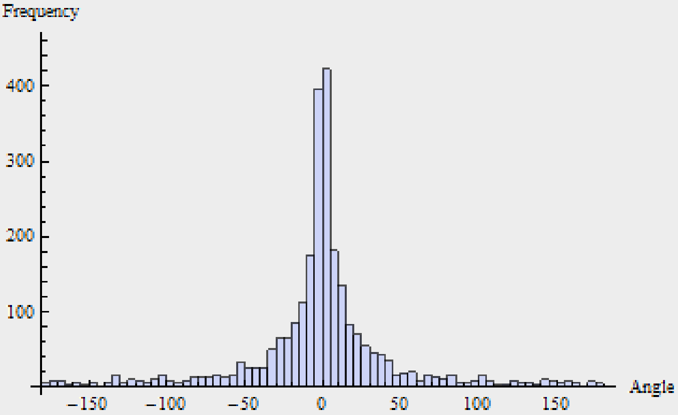
Angular Dependence

- If our idea that the underlying distribution is bivariate normal is correct, then there should be no angular dependence in the results.
- To measure angles, let the blue dots represent crime locations, the red square the anchor point, and the green triangle the centroid of the crime series.
- Then measure the angle between the ray from the anchor point to the crime site and the ray from the anchor point to the centroid.



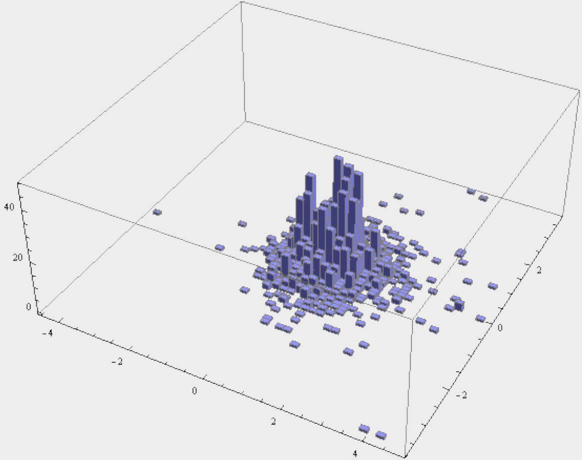
Angular Dependence

- The residential burglary data shows a striking relationship- nearly all of the crime sites lie in the same direction as the centroid.



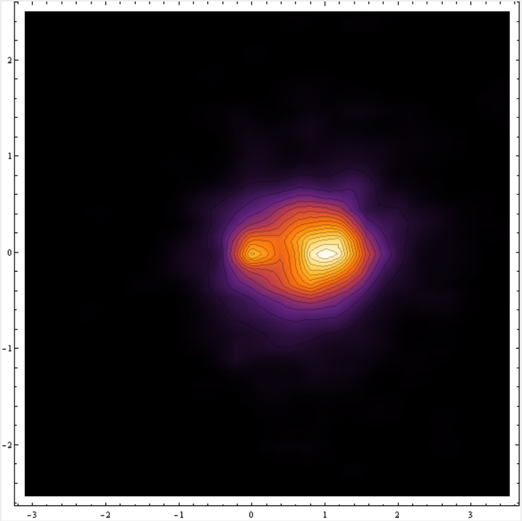
Two-dimensional Distribution

- Plot the histogram of the scaled two dimensional data set; here the offender's home is at the origin, and the centroid of the crime series is at $(x, y) = (1, 0)$.



Two-dimensional Distribution

- Here is another view as a two-dimensional density; note that it is not centered at the origin.



Alternative Hypotheses

- One one hand, the distribution of distances from the offender's home to crime site appears to follow a Rayleigh distribution- at least for marauders.
- On the other hand, it is just as clear that the bivariate distribution is not bivariate normal.
- Indeed, it is clear that there are significant correlations between the locations of the different crime site locations.
    - As evidence, we have the fact that the scaled bivariate distribution clusters not around the offender's home, but around the centroid of the crime series.
- Perhaps we should consider a two stage mixture model:
    - Offenders select a target area
    - Within that target area, offenders select a target.
- Can we model these processes separately?

Scaling

- If we want to model the selection of crime sites within a hunting area, we should not use as the length scale the distance from the offender's home to the crime site.
- A reasonable choice for the length scale is the distance from the individual crime sites to the centroid of the crime series.
- For each serial offender with crime sites $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, let

$$\mathbf{c} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$$

be the centroid, select the length scale

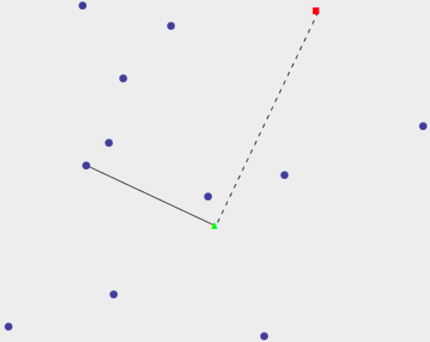
$$\hat{\alpha}_c = \frac{1}{n} \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{c})$$

and consider the set of scaled distances

$$\rho_i = \frac{d(\mathbf{x}_i, \mathbf{c})}{\hat{\alpha}_c}$$

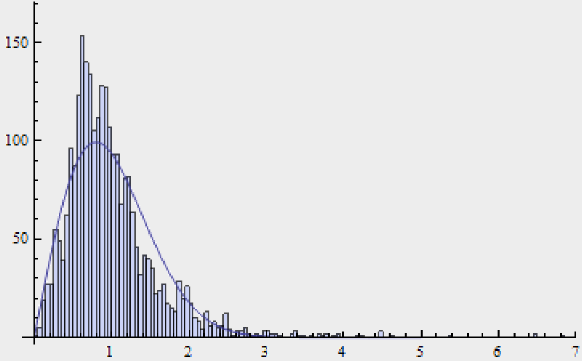
Site Distribution

- The center of our coordinate system is on the centroid of the crime series, angles are measured from the ray from the centroid (green triangle) to the offender's home base (red square).



Site Distribution

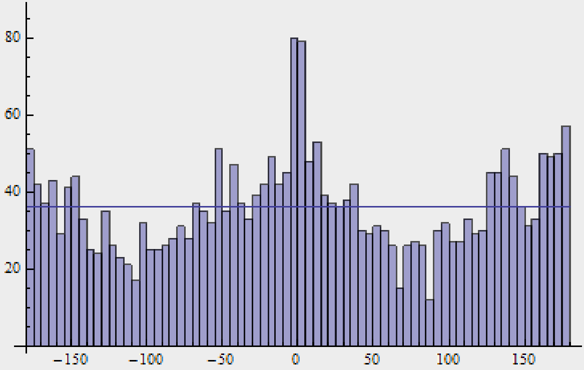
- If we plot the scaled distances from the crime site to the centroid, we again obtain a good match with the Rayleigh distribution.



- This includes all offenders- commuters and marauders.

Site Distribution

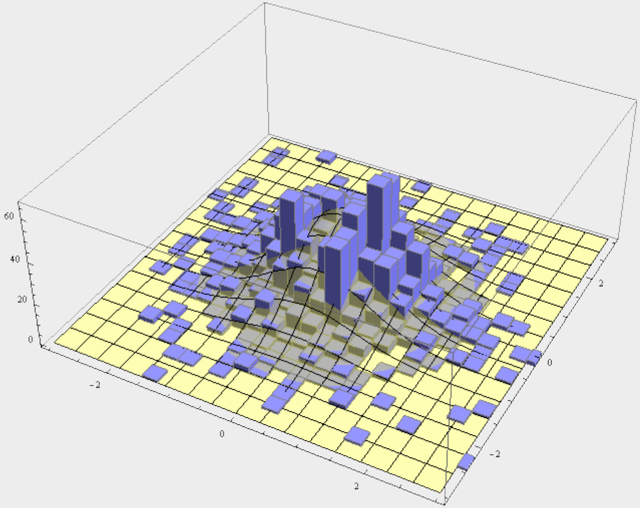
- Now if we plot the angles, we see that the uniform distribution is a much better fit.



- Clearly some anisotropy remains.

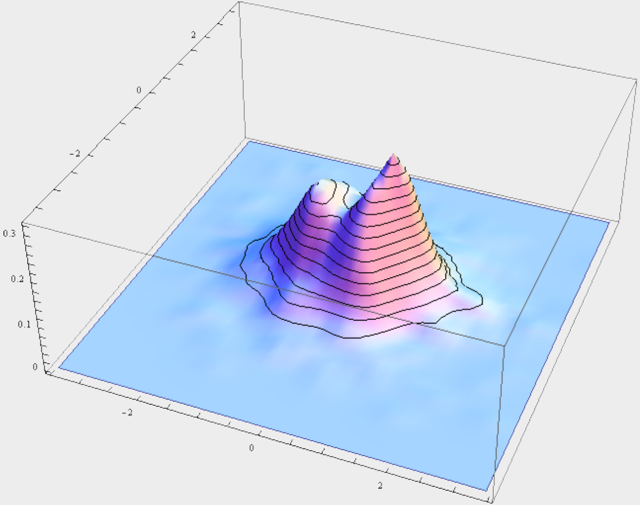
Site Distribution

- We can now directly compare the bivariate normal to the scaled two-dimensional distribution, to see a reasonable fit.



Site Distribution

- The deviation of the scaled distribution from a bivariate normal is more obvious when we smooth the histogram.



Site Distribution: Conclusions

- We have evidence that the distribution of crime site locations is roughly bivariate normal, and centered around the centroid of the crime series.
- No distinction needs to be drawn between commuters and marauders.
- There are noticeable deviations from normality:
  - Directions in line with the offender's home address are preferred to perpendicular directions.
  - There is a preference for crime sites closer to the offender's home address than locations farther away.
  - There appears to be weak evidence for the existence of a buffer zone around the offender's home.
- These hypotheses have only been tested on residential burglaries in Baltimore County.

Hunting Area Distribiution

- How do we get information about the location of the offender's home?
- Consider the distribution of the centroid of the crime series (the hunting area) around the offender's home- can we find a good match for that distribution?

Hunting Area Distribution: Scaling

- When we examined the distance from the crimes to the home, we used as the scale the distance from the centroid of the series to home. Clearly we cannot use that scaling now.
- Instead, we use the same scaling as for the distribution of the crimes around the centroid;
- For each serial offender with crime sites $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, let $\mathbf{c}$ be the centroid, select the length scale

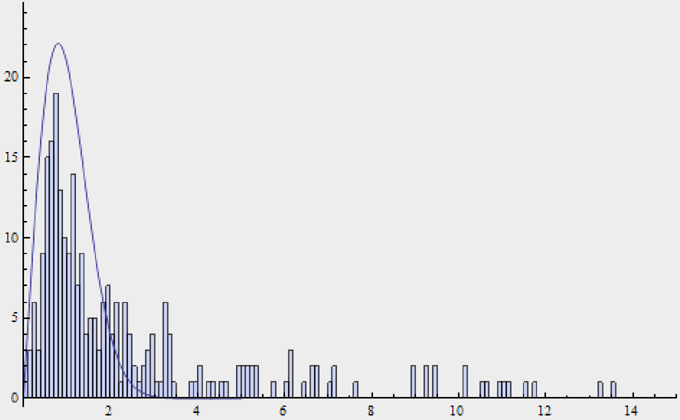
$$\hat{\alpha}_c = \frac{1}{n} \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{c})$$

and consider the set of scaled distances from the centroid $\mathbf{c}$ to the home base $\mathbf{z}$

$$\rho = \frac{d(\mathbf{z}, \mathbf{c})}{\hat{\alpha}_c}$$

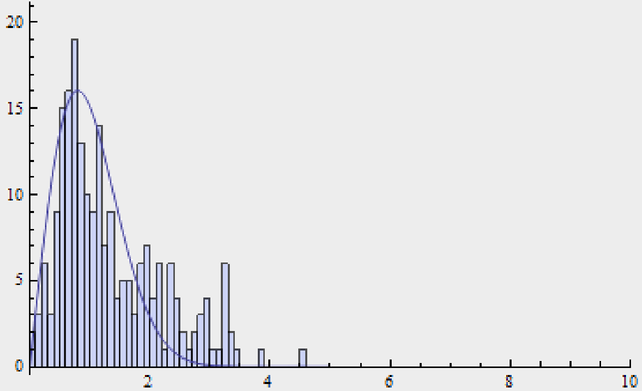
Hunting Area Distribution

- Plot the scaled distances from the offender's home to the centroid of the crime series:



Hunting Area Distribution

- Rayleigh is not a *bad* match.
- Clearly the tails of the distribution are far too heavy.
- Better fits can be obtained by including only marauders:



- The data suggests that there are other unknown factors in play for commuters; perhaps a mixture of behaviors.

Hunting Area Distribution

- We would like to study the distribution of angles, as well as the full two-dimensional distribution.
- But- What is the angle referent?
  - We can't use the ray from the centroid to the offender's home as we have done.
  - There is no natural replacement.
  - Principal component axes of the crime series?

Conclusions

- We investigated the use of a mixture distribution for offender distance behavior, where the select a single hunting area, and then select individual crime sites from within that hunting area.

Conclusions

- We investigated the use of a mixture distribution for offender distance behavior, where the select a single hunting area, and then select individual crime sites from within that hunting area.
- The distribution of crime sites is well modeled by a bivariate normal distribution centered at the centroid of the crime series, with one length scale parameter that varies across offenders.

Conclusions

- We investigated the use of a mixture distribution for offender distance behavior, where the select a single hunting area, and then select individual crime sites from within that hunting area.
- The distribution of crime sites is well modeled by a bivariate normal distribution centered at the centroid of the crime series, with one length scale parameter that varies across offenders.
- There is weak evidence for a Rayleigh type model for the scaled distance from the offender's home to the centroid of the crime series.
    - It is clear that this does not model commuters well.
    - No two-dimensional evidence is available.

Conclusions

- We investigated the use of a mixture distribution for offender distance behavior, where the select a single hunting area, and then select individual crime sites from within that hunting area.
- The distribution of crime sites is well modeled by a bivariate normal distribution centered at the centroid of the crime series, with one length scale parameter that varies across offenders.
- There is weak evidence for a Rayleigh type model for the scaled distance from the offender's home to the centroid of the crime series.
    - It is clear that this does not model commuters well.
    - No two-dimensional evidence is available.
- If offenders do follow a mixture model, then additional crimes in a series do not improve predictive accuracy of the offender's home, save through an improved estimate of the hunting area.

Questions?

Mike O'Leary
Department of Mathematics
Towson University
moleary@towson.edu

`http://pages.towson.edu/moleary/Profiler.html`