The Mathematics of Geographic Profiling

Mike O'Leary
Department of Mathematics
Towson University

We begin by describing some of the mathematical foundations of the geographic profiling problem. We then present a new mathematical framework for the geographic profiling problem based on Bayesian statistical methods that make explicit connections between assumptions on offender behavior and the components of the mathematical model. It also can take into account local geographic features that either influence the selection of a crime site or influence the selection of an offender's anchor point.

Keywords: Geographic profiling, Bayesian analysis, mathematical modeling

Introduction

The geographic profiling problem is the problem of constructing an estimate for the location of the anchor point of a serial offender from the locations of the offender's crime sites; see Rossmo (2000, p. 1). In this paper, we shall present a mathematical survey of some of the algorithms that have been used to solve the geographic profiling problem. We will then present a new mathematical framework for the geographic profiling problem based on Bayesian methods that is able to incorporate both geographic features that influence the choice of a crime site as well as geographic features that affect the location of the offender's anchor point.

It has long been recognized that there are important relationships between geography and crime, including serial crime; we mention Brantingham and Brantingham (1993), Canter and Larkin (1993) and Rossmo (2000). Today there are a number of software packages being used to solve the geographic profiling problem. These include CrimeStat, developed by Ned Levine; Dragnet, developed by David Canter; and Rigel, developed by Kim Rossmo. There have been significant disagreements in the literature as to what is the best methodology to evaluate the currently existing geographic profiling software. See the original report prepared for NIJ (Rich & Shively, 2004), the critique of Rossmo (2005a), and the response of Levine (2005). There is also an ongoing lively discussion in the literature as to whether or not computer systems are as effective as simply providing humans with some simple heuristics; see Snook, Canter, and Bennell (2002), Snook, Taylor, and Bennell (2004), Rossmo (2005b) and Snook, Taylor, and Bennell (2005b). We also have the discussion in Snook, Taylor, and Bennell (2005a), Rossmo and Filer (2005), Bennell, Snook, and Taylor (2005), and Rossmo, Filer, and Sesley (2005), as well as the papers of Bennell, Snook, Taylor, Corey, and Keyton (2007) and of Bennell, Taylor, and Snook (2007).

Given these and other controversies, we begin by enumerating the characteristics we feel that a sound mathematical algorithm for the geographic profiling should possess:

- The method should be logically rigorous;
- There should be explicit connections between assumptions on offender behavior and the components of the model;
- The method should be able to take into account local geographic features; in particular, it should be able to account for geographic features that influence the selection of a crime site and geographic features that influence the potential anchor points of offenders;
- The method should be based on data that is available to the jurisdiction(s) where the offenses occur; and
- The method should return a prioritized search area for law enforcement officers.

Ensuring that the algorithms are rigorous and explicit in the connections between the assumptions on offender behavior and components of the model will help in the analysis of the model. In particular, it will give researchers another tool for evaluating a model. As we have noted there is no consensus as to the method(s) that should be used to evaluate the effectiveness of a geographic profiling strategy. It is important that the mathematics explicitly allows for the influence of the local geography and demography. It is well known that there are relationships between the physical environment and crime rates; see for example Brantingham and Brantingham (1993). It is essential that a good mathematical framework has the ability to incorporate this information into the model. However, it is equally important that the model use only data that is available to the appropriate law enforcement agency. Finally, we recognize that simple point estimates of offender anchor points are not very valuable to practicing law enforcement officers. Rather, to be useful to practitioners, a good algorithm must produce prioritized search areas.

This work was supported by National Institute of Justice through grant 2007-DE-BX-K005 and 2005-IJ-CX-K036.

We express our gratitude to Phil Canter, Baltimore County Police Department.

Existing Methods

To begin our review of the current state of geographic profiling, let us agree to adopt some common notation. A point $\mathbf{x}$ will have two components $\mathbf{x} = (x^{(1)}, x^{(2)})$. These can be latitude and longitude, or distances from a fixed pair of perpendicular reference axes. We presume that we are working with a series of n linked crimes, and the crime sites under consideration are labeled $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. We use the symbol $\mathbf{z}$ to denote the offender's anchor point. The anchor point can be the offender's home, place of work, or some other location of importance to the offender.

We shall let $d(\mathbf{x}, \mathbf{y})$ denote the distance metric between the points $\mathbf{x}$ and $\mathbf{y}$. There are many reasonable choices for this metric, including the Euclidean distance, the Manhattan distance, the total street distance following the local road network or the total time to make the trip while following the local road network.

Existing algorithms begin by first making a choice of distance metric d; they then select a decay function f and construct a hit score function $S(\mathbf{y})$ by computing

$$S(\mathbf{y}) = \sum_{i=1}^{n} f(d(\mathbf{x}_i, \mathbf{y})) = f(d(\mathbf{x}_1, \mathbf{y})) + \cdots + f(d(\mathbf{x}_n, \mathbf{y})). \quad (1)$$

Regions with a high hit score are considered to be more likely to contain the offender's anchor point than regions with a low hit score. In practice, the hit score $S(\mathbf{y})$ is not evaluated everywhere, but simply on some rectangular array of points $\mathbf{y}_{jk} = (y_j^{(1)}, y_k^{(2)})$ for $j \in \{1, 2, \ldots J\}$ and $k \in \{1, 2, \ldots, K\}$, giving us the array of values $S_{jk} = S(\mathbf{y}_{jk})$.

Rossmo's method, as described in (Rossmo, 2000, Chapter 10) chooses the Manhattan distance function for d and the decay function

$$f(d) = \begin{cases} \frac{k}{d^h} & \text{if } d < B, \\ \frac{kB^{g-h}}{(2B-d)^g} & \text{if } d \geq B. \end{cases}$$

We remark that Rossmo also considers the possibility of forming hit scores by multiplication; see (Rossmo, 2000, p. 200)

The method described in Canter, Coffey, Huntley, and Missen (2000) is to use a Euclidean distance, and to choose either a decay function in the form

$$f(d) = e^{-\beta d}$$

or functions with a buffer and plateau, with the form

$$f(d) = \begin{cases} 0 & \text{if } d < A, \\ B & \text{if } A \leq d < B, \\ Ce^{-\beta d} & \text{if } d \geq B. \end{cases}$$

The CrimeStat program described in Levine (2009a) uses Euclidean or spherical distance and gives the user a number of choices for the decay function, including

- Linear: $f(d) = A + Bd$,
- Negative exponential: $f(d) = Ae^{-\beta d}$,
- Normal: $f(d) = A(2\pi S^2)^{-1/2} \exp[-(d - \bar{d})^2/2S^2]$,
- Lognormal: $f(d) = A(2\pi d^2 S^2)^{-1/2} \exp[-(\ln d - \bar{d})^2/2S^2]$, and
- Truncated negative exponential: $f(d) = Bd$ if $d < C$ and $f(d) = Ae^{-\beta d}$ if $d \geq C$.

CrimeStat also allows the user to use empirical data to create a different decay function matching a set of provided data as well as the use of indirect distances.

Though each of these approaches are distinct, they share the same underlying mathematical structure; they vary only in the choice of decay function and the choice of distance metric. We remark that the latest version (3.2) of CrimeStat contains a new Bayesian Journey to Crime Module that integrates information on the origin location of other offenders who committed crimes in the same location with the distance decay estimates (Levine, 2009b). Levine and Block tested this method with data from Baltimore County and from Chicago (Levine & Block, 2010). See the introduction to this special issue.

A New Mathematical Approach

We begin by looking for an appropriate model for offender behavior and start with the simplest possible situation– where we know nothing about the offender. Thus, we assume that our offender chooses potential locations to offend randomly according to some unknown probability density function $P(\mathbf{x})$. For any geographic region R, the probability that our offender will choose a crime site in R can be found by adding up the values of P in R, giving us the probability $\iint_R P(\mathbf{x})\, dx^{(1)}\, dx^{(2)}$.

At first glance, it may seem odd to use a probabilistic model to describe human behavior. In fact, probabilistic models are commonly used to describe many kinds of apparently deterministic phenomena. For example, classical models of the diffusion of heat or chemical concentration can be derived probabilistically; they also see application in models of the stock market (Baxter & Rennie, 1996), (Wilmott, 1998), (Wilmott, Howison, & Dewynne, 1995), in models of population genetics (Ewens, 2004), and in many other models (Beltrami, 1993).

More precisely, the probability density function P represents our knowledge of the behavior of the offender. We use a probability distribution, not because the offender's decision has a random component, although it may. Rather, we use a probability density because we lack complete information about the offender. Indeed, consider the following thought experiment. If we want to model the flip of a coin, we use probability and assume that each side of the coin is apt to occur half the time. Now instead of flipping the coin, let us take the coin to a colleague and ask them to choose a side. In this case the outcome is the deliberate result of a decision by an individual. However without knowing more information about our colleague's preferences, the best choice to model the outcome of that experiment is still the use of a probability distribution.

Returning to our model of offender behavior, we begin with a question: upon what sorts of variables should our probability density function P depend? One of the fundamental assumptions of geographic profiling is that the choice of an offender's target locations is influenced by the location of the offender's anchor point $\mathbf{z}$. Therefore, we assume that P depends upon $\mathbf{z}$. Underlying this approach are the requirements that the offender has a single anchor point and that it is stable during the crime series.

A second important factor is the distance our offender is willing to travel to commit a crime. Let α denote the average distance that our offender is willing to travel to offend. We allow for the possibility that this value varies between offenders. Combining these, we assume that there is a probability density function $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$ for the probability that an offender with a single stable anchor point $\mathbf{z}$ and average offense distance α commits a crime at the location $\mathbf{x}$.

We assume that this model is local to the jurisdiction under consideration. In particular, we explicitly allow for the possibility that different models $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$ may need to be chosen for different jursidictions.

The key mathematical point is that the unknown is now the entire distribution $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$, rather than just the anchor point $\mathbf{z}$. On its face, it seems a step backwards, but in fact, it is not. Indeed, let us suppose that the *form* of the distribution P is known, but that the values of the anchor point $\mathbf{z}$ and average offense distance α are unknown. Then the problem can be stated mathematically as, given a sample $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ (the crime site locations) from the distribution $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$ with parameters $\mathbf{z}$ and α to determine the best way to estimate the parameter $\mathbf{z}$ (the anchor point).

For the moment, let us set aside the question of what reasonable choices can be made for the form of the distribution $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$, and focus on how we can estimate the anchor point $\mathbf{z}$ from our knowledge of the crime locations $\mathbf{x}_1, \dots, \mathbf{x}_n$.

It turns out that this is a well studied mathematical problem. One approach is to use the maximum likelihood estimator. To do so, one first forms the likelihood function:

$$L(\mathbf{y}, a) = \prod_{i=1}^{n} P(\mathbf{x}_i \,|\, \mathbf{y}, a) = P(\mathbf{x}_1 \,|\, \mathbf{y}, a) \cdots P(\mathbf{x}_n \,|\, \mathbf{y}, a).$$

Then the maximum likelihood estimates $\hat{\mathbf{z}}_{\mathrm{mle}}$ and $\hat{\alpha}_{\mathrm{mle}}$ are the values of $\mathbf{y}$ and a that make L as large as possible. Equivalently, one can maximize the log-likelihood function

$$\lambda(\mathbf{y}, a) = \sum_{i=1}^{n} \ln P(\mathbf{x}_i \,|\, \mathbf{y}, a) = \ln P(\mathbf{x}_1 \,|\, \mathbf{y}, a) + \cdots + \ln P(\mathbf{x}_n \,|\, \mathbf{y}, a).$$

Though rigorous, this approach is unsuitable as simple point estimates for the offender's anchor point are not operationally useful. Instead, we continue our analysis by using Bayes' Theorem.

Bayesian Analysis

To see how Bayesian methods can be applied to geographic profiling, we begin with the simplest case where the offender has only committed one crime at the location $\mathbf{x}$. We would like to use the information from this crime location to form an estimate for the probability distribution for the anchor point $\mathbf{z}$. Bayes' Theorem gives us the estimate

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}) = \frac{P(\mathbf{x} \,|\, \mathbf{z}, \alpha)\pi(\mathbf{z}, \alpha)}{P(\mathbf{x})} \tag{2}$$

(Carlin & Louis, 2000; Casella & Berger, 2002). Here $P(\mathbf{z}, \alpha \,|\, \mathbf{x})$ is the posterior distribution, which gives the probability density that the offender has anchor point $\mathbf{z}$ and average offense distance α, given that the offender has committed a crime at the location $\mathbf{x}$.

The term $P(\mathbf{x})$ is the marginal distribution. The important thing to note is that it is independent of $\mathbf{z}$ and α, therefore it can be ignored provided we replace the equality in (2) with proportionality.

The term $\pi(\mathbf{z}, \alpha)$ is the prior distribution. It represents our knowledge of the probability density that the offender has anchor point $\mathbf{z}$ and average offense distance α before we incorporate any information about the crime series. One approach to the prior is to assume that the anchor point $\mathbf{z}$ is mathematically independent of the average offense distance α. In this case, we can factor to obtain

$$\pi(\mathbf{z}, \alpha) = H(\mathbf{z})\pi(\alpha) \tag{3}$$

where $H(\mathbf{z})$ is the prior probability density function for the distribution of anchor points before any information from the crime series is included and $\pi(\alpha)$ is the probability density function for the prior distribution of the offender's average offense distance, again before any information from the crime series is included.

Combining these, we then obtain the expression

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}) \propto P(\mathbf{x} \,|\, \mathbf{z}, \alpha)H(\mathbf{z})\pi(\alpha).$$

Of course, we are interested in crime series, and we would like to estimate the probability density for the anchor point $\mathbf{z}$ given our knowledge of all of the crime locations $\mathbf{x}_1, \dots, \mathbf{x}_n$. To do so, we proceed in a similar fashion; now Bayes' Theorem implies

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{P(\mathbf{x}_1, \dots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)\pi(\mathbf{z}, \alpha)}{P(\mathbf{x}_1, \dots, \mathbf{x}_n)}.$$

Here $P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \dots, \mathbf{x}_n)$ is again the posterior distribution, which gives the probability density that the offender has anchor point $\mathbf{z}$ and average offense distance α, given that the offender has committed a crime at each of the locations $\mathbf{x}_1, \dots, \mathbf{x}_n$. The marginal $P(\mathbf{x}_1, \dots, \mathbf{x}_n)$ remains independent of $\mathbf{z}$ and α, and can be ignored; the prior π can be handled by (3). Then

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \dots, \mathbf{x}_n) \propto P(\mathbf{x}_1, \dots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)H(\mathbf{z})\pi(\alpha). \tag{4}$$

The factor $P(\mathbf{x}_1, \dots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha)$ on the right side is the joint probability that the offender committed crimes at all of the locations $\mathbf{x}_1, \dots, \mathbf{x}_n$ given that they had anchor point $\mathbf{z}$ and average offense distance α. The simplest assumption we can make is that all of the offense sites are mathematically independent; then we have the reduction

$$P(\mathbf{x}_1, \ldots, \mathbf{x}_n \,|\, \mathbf{z}, \alpha) = P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) \cdots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha). \tag{5}$$

Substituting this into (4) gives

$$P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) \cdots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha) H(\mathbf{z}) \pi(\alpha).$$

Finally, since we are only interested in the location of the anchor point $\mathbf{z}$, we take the conditional distribution to obtain our fundamental mathematical result:

$$P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto \int P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) \cdots P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha) H(\mathbf{z}) \pi(\alpha)\, d\alpha. \tag{6}$$

The expression $P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n)$ gives us the probability density that the offender has anchor point $\mathbf{z}$ given that they have committed crimes at the location $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Because we are calculating probabilities, this immediately provides us a rigorous search area for the offender. Indeed regions with larger values of $P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n)$ by definition are more likely to contain the offender's anchor point than regions where $P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n)$ is lower.

This is a very general framework for the geographic profiling problem. There are many choices for the model of offender behavior $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$, and we will later examine a number of reasonable choices. Though the preceding used a model for offender behavior with one parameter α other than the anchor point, the mathematics continues to hold with elementary modifications if we either add additional parameters or remove the parameter α.

In addition to an assumption as to the form of $P(\mathbf{x} \,|\, \mathbf{z}, \alpha)$, we have made two other fundamental assumptions. One is that the prior for $\mathbf{z}$ is independent of the prior for α. This is a reasonable first assumption, and it is what allows us the factorization in (3). Its significance is that we are assuming that the average distance that the offender is willing to travel is independent of the offender's anchor point. This is probably most appropriate in urban areas and for regions where offenders travel short distances to offend. On the other hand, the assumption may be less valid for example, in a town surrounded by a less populated rural area. If potential offense locations are concentrated in the town, then offenders with anchor points far from the town will likely have a higher average offense distance than offenders with anchor points in town.

The remaining fundamental assumption is that the offender's choice of crime sites are independent; this is necessary for the factorization in (5). It can be replaced by other assumptions, but would require a different model for the joint distribution than the simple expression in (5). Though reasonable as a first assumption, there is evidence of deviation from independence in the literature. For example Kocsis, Cooksey, Irwin, and Allen (2002) found in their analysis of 58 multiple burglary cases in rural Australia that the crime sites tended to lie in narrow corridors emanating from the offenders anchor point. Meaney (2004) examined 83 burglary series, 23 sexual offense series, and 21 arson series; she found that the first offense occurred closer to the offender's home than the last offense, suggesting that there is a temporal component to offender's site selection. On the other hand, Laukkanen and Santtila (2006) concluded that the distance a robber travels to offend did not increase as the crime series progressed. We also mention Ratcliffe (2006) who examined some of the interrelation between temporal data and routine activity theory.

Simple Models for Offender Behavior

If our fundamental mathematical result is to have any practical or investigative value, we need to be able to construct reasonable choices for our model of offender behavior. One simple model is to assume that the offender chooses a target location based only on the Euclidean distance from the offense location to the offender's anchor point and that this distribution is normal. In this case we obtain

$$P(\mathbf{x} \,|\, \mathbf{z}, \alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2} |\mathbf{x} - \mathbf{z}|^2\right). \tag{7}$$

If we make the prior assumptions that all offenders have the same average offense distance α and that all anchor points are equally likely, then

$$P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \left(\frac{1}{4\alpha^2}\right)^n \exp\left(-\frac{\pi}{4\alpha^2} \sum_{i=1}^{n} |\mathbf{x}_i - \mathbf{z}|^2\right).$$

We see that the posterior anchor point probability distribution is just a product of normal distributions, one centered at each crime site; compare this to sums used in the calculation of hit scores (1). We also mention that in this model of offender behavior, the maximum likelihood estimate for 'the anchor point is simply the mean center of the crime site locations; this is also the mode of the posterior anchor point probability distribution $P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n)$.

Another reasonable choice of a model for offender behavior is to assume that the offender chooses a target location based only on the Euclidean distance from the offense location to the offender's anchor point, but that now the distribution is a negative exponential so that

$$P(\mathbf{x} \,|\, \mathbf{z}, \alpha) = \frac{2}{\pi\alpha^2} \exp\left(-\frac{2}{\alpha} |\mathbf{x} - \mathbf{z}|\right). \tag{8}$$

Once again, if our prior assumptions are that all offenders have the same average offense distance and that all anchor points are equally likely, then

$$P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \left(\frac{2}{\pi\alpha^2}\right)^n \exp\left(-\frac{2}{\alpha} \sum_{i=1}^{n} |\mathbf{x}_i - \mathbf{z}|\right).$$

We see that this is just a product of negative exponentials centered at each crime site. Further, the corresponding maximum likelihood estimate for the offender's anchor point is simply the center of minimum distance for the crime series locations. Finally, if we construct the function $\tilde{S}(\mathbf{z}) =$

$\ln P(\mathbf{z} \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_n)$, then $\tilde{S}$ is a hit score in the same form as (1) with a linear decay function f and Euclidean distance d.

This preceding analysis was predicated on the assumption that all offenders have the same average offense distance α and that this was known in advance. Similarly, the existing hit score methods all rely on decay functions f with one or more parameters that also need to be determined in advance. Unlike the hit score techniques however, our method does not require that we make a choice for the parameter α in advance. For example, if we assume only that the offender has a distance decay in the form (7) (or in the form (8)), with α unknown, then the maximum likelihood technique will estimate both the anchor point $\mathbf{z}$ and the average offense distance α. Our fundamental mathematical result (6) also does not require that the parameter α be determined in advance, though it does require a prior estimate $\pi(\alpha)$ for the distribution of average offense distances.

More Realistic Models for Offender Behavior

These simple models for offender behavior show that our framework recaptures many existing geographic profiling techniques; however, this new method is more general and allows us a simple way to incorporate geographic features into the model. Indeed, let us suppose that offender target selection depends on more than just the distance from the anchor point to the crime site locations, but that it depends on some features in the local geography. One way to account for this is to suppose that the offense probability density is proportional to both a distance decay term and to a function that measures the attractiveness of a particular target location. Doing so, we obtain the following expression

$$P(\mathbf{x} \,|\, \mathbf{z}, \alpha) = D(d(\mathbf{x}, \mathbf{z}), \alpha) G(\mathbf{x}) N(\mathbf{z}). \tag{9}$$

Here the factor D models the effect of distance decay using the distance metric $d(\mathbf{x}, \mathbf{z})$. For example, we can specify a normal decay, so that

$$D(d, \alpha) = \frac{1}{4\alpha^2} \exp\left(-\frac{\pi}{4\alpha^2} d^2\right).$$

We could also specify a negative exponential decay, so

$$D(d, \alpha) = \frac{2}{\pi\alpha^2} \exp\left(-\frac{2}{\alpha} d\right),$$

but of course there are many other reasonable possibilities.

One of the consequences of this approach to distance decay is that it assumes uniformity of travel direction with respect to the given distance metric; in this way it simplifies actual travel behavior. It may be the case that certain directions are preferred by the offender; for example when searching for potential targets, the offender may prefer to move closer to an urban area than farther away. A new approach to the modeling this distance decay effect is the kinetic random walk model of Mohler and Short (2009).

The factor $G(\mathbf{x})$ is used to account for the local geographic features that influence the selection of a crime site. High values for $G(\mathbf{x})$ indicate that $\mathbf{x}$ is a likely target for typical offenders; low values indicate $\mathbf{x}$ is a less likely target.

The remaining factor N is a normalization required to ensure that P is a probability distribution. Its value is completely determined by the choices of D and G and has the form

$$N(\mathbf{z}) = \frac{1}{\iint D(d(\mathbf{y}, \mathbf{z}), \alpha) G(\mathbf{y}) dy^{(1)} dy^{(2)}}.$$

Returning to the influence of geography on target selection, one simple example of $G(\mathbf{x})$ is to account for jurisdictional boundaries. Suppose that all known crimes in the series must occur in a region J. The offender can commit crimes outside J; but these are presumed unknown to the analyst; the offender's anchor point may also reside outside the region J. We can account for this with the simple model

$$G(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in J, \\ 0 & \text{if } \mathbf{x} \notin J. \end{cases}$$

In practice, the region J corresponds to one or more jurisdictions sharing information about the offender's crime series.

The incorporation of this very simple geographic information has some surprising consequences. In particular, the algorithm is able to distinguish between areas where no crimes in the series have occurred (inside J) from areas where there is no information as to whether or not a crime in the series may have occurred (outside J). For example, suppose that the elements of a hypothetical crime series are all near the southern boundary of a jurisdiction J. Then the algorithm will return a search area skewed to the south of the crime series because the algorithm "knows" that no known crimes take place north of the series, but that there may be crimes to the south of the series that are unknown to the analyst; thus the offender is more likely to live south of the series than to the north. As a consequence, this model does not suffer from the convex hull effect described by Levine (2005).

This simple approach to geographic information affecting the selection of the target is primarily illustrative; clearly a better model can be chosen. To do so, one approach would be to use available geographic and demographic data and the correlations between crime rates and these variables that have already been published to construct an appropriate choice for $G(\mathbf{x})$. However, this approach has a number of issues. First, is the fact that different crime types have different etiologies; in particular their relationship to the local geographic and demographic backcloth depends strongly on the particular type of crime. This would limit the method to only those crimes where this relationship has been well studied. Moreover, even for well studied crimes, there are regional differences. Indeed, Tseloni, Wittebrood, Farrel, and Pease (2004) noted that increased household affluence indicated higher burglary rates in Britain, and indicated lower burglary rates in the U.S.

The primary issue here is that this approach posits a method to explain crime rates by looking for explanatory variables. However, from the perspective of geographic profiling, it is unnecessary to explain; instead we can simply acknowledge these differences, and work on measuring the resulting differences. Rather than look at the local geographic variables, we can use historical data to model the geographic target attractiveness.

In particular, let us assume that historical crime rates are reasonable predictors of the likelihood that a particular region will be the site of an offense. Then, given a crime series that we wish to analyze, we require a representative list of historically committed crimes of the same type. Clearly, this process requires the presence of a skilled analyst to determine which historical crimes are of the same type as the series under consideration. As an example, when looking at a series of street robberies, it is likely that the geographic distribution of street robbery rates are different for daylight robberies as opposed to late night robberies. This approach then inserts the crime analyst and their relevant real world experience directly into the modeling process and the algorithm. This approach also lets us handle different crime types within the same mathematical framework, as different crime types will have different historical patterns. Of course, the local analyst will need to have access to the necessary data.

Once we have the historical data, we need to estimate the target density function $G(\mathbf{x})$. Perhaps the simplest method is kernel density parameter estimation. To use this method, let us suppose that we have a representative list of the crimes of a given type and that they have occurred at the points $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N$. Choose a kernel density function $K(\mathbf{y} \mid \lambda)$ with bandwidth λ. There are a number of reasonable choices for the kernel density function K, including normal or truncated quartic. It turns out that the mathematical properties of this method do not depend strongly on the mathematical form of the kernel, but that they do depend on the bandwidth of the kernel (Silverman, 1986). The bandwidth λ of a given kernel is related to the width of the function; as an example the bandwidth of a normal curve is the variance of that normal; when using a truncated quartic, the bandwidth is the size of the interval for which the quartic is nonzero.

We then construct the local target attractiveness function by calculating

$$G(\mathbf{x}) = \sum_{i=1}^{N} K(\mathbf{x} - \mathbf{c}_i \mid \lambda) \tag{10}$$

for a reasonable choice of bandwidth λ, say the mean nearest neighbor distance between historical crime sites. This is essentially the same as one of the methods used to generate crime hot spots described in Chainey (2005). Similar techniques are used in mathematical biology to estimate the home range of an animal species based on observations on individuals in the environment (Worton, 1989).

Once we have selected a model for offender behavior $P(\mathbf{x} \mid \mathbf{z})$, we also need to make a choice for the prior probability density for offender anchor points $H(\mathbf{z})$ and the prior distribution of the offender's average offense distance $\pi(\alpha)$ before we can use our fundamental result (6).

The prior probability density for offender anchor points $H(\mathbf{z})$ represents our knowledge of the offender's anchor point before we use any of the information from the crime series itself. There are a number of mathematically and criminologically reasonable choices for this prior distribution. The simplest choice would be to assume all potential anchor points are equally likely; we can do this by simply choosing $H(\mathbf{z}) = 1$.

Before we examine more sophisticated priors, we return to the question of what is an anchor point. If we assume that the anchor point is the offender's home, or more generally that the distribution of anchor points follows local population density, then we can use demographic data to generate an estimate for the prior. In this case, we we can choose $H(\mathbf{z})$ so that it is proportional to local population density. U.S. Census data gives population counts at the block level together with the land area of the block. We can use this data and kernel density parameter estimation technique to generate $H(\mathbf{z})$ by calculating

$$H(\mathbf{z}) = \sum_{i=1}^{N_{\text{blocks}}} = p_i K(\mathbf{z} - \mathbf{q}_i \mid \sqrt{A_i})$$

where each block has population p_i, center $\mathbf{q}_i$ and for each block we have chosen a different bandwidth equal to the side length of a square with the same area A_i as the block. We mention that U.S. Census population data at the block level is also available sorted by age, sex, and race/ethnic group. Thus, if demographic information is available about the offender, then this information can be incorporated when the prior distribution of anchor points $H(\mathbf{z})$ is calculated.

Our framework does not require that the anchor point be the offender's home or that the distribution of anchor points follows local population density. Another reasonable approach to calculating $H(\mathbf{z})$ would be to begin with the anchor points of previous offenders who have committed similar crimes. Then the same kernel density process used to generate $G(\mathbf{x})$ in (10) can be used to generate $H(\mathbf{z})$. These historical anchor points can be determined on an offender-by-offender basis; they can be homes, places of work or even the offender's favorite bar. Recall however that one of our assumptions is that each offender has a unique stable anchor point.

The last element needed to implement our fundamental mathematical result is some estimate of the prior distribution of the average distance to crime. Estimates of these types of distance to crime distributions are commonly performed by choosing a common statistical function and using best fit estimates; see Levine (2009a, Chapter 10) for an example of the process. However, our framework does not require a particular parametrized form for the prior distribution $\pi(\alpha)$; we can instead directly use appropriate empirical data in the construction. Further, there is no requirement that the same choice of $\pi(\alpha)$ needs to be made for different crime types. Again, an analyst can choose which historical data to use when generating $\pi(\alpha)$.

Prototype software that implements this framework has been developed and released to for public use and evaluation. Empirical tests are being arranged to evaluate the accuracy and precision of this approach.

Future Offense Prediction

The focus of our attention so far has been on the traditional geographic profiling problem of estimating the location of the offender's anchor point by using the geograhphic information contained in the crime series. However, this is not the only question of interest to law enforcement. Indeed, another question of nearly equal importance is to estimate the location of the serial offender's next target.

This question can be posed in the following mathematical form. Given a series of crimes at the locations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ committed by a single serial offender, estimate the probability density $P(\mathbf{x}_{\text{next}} \,|\, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, that $\mathbf{x}_{\text{next}}$ will be the location of the next offense. The Bayesian approach to this problem is to calculate the posterior predictive distribution

$$\begin{aligned} &P(\mathbf{x}_{\text{next}} \,|\, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \\ &= \iiint P(\mathbf{x}_{\text{next}} \,|\, \mathbf{z}, \alpha) P(\mathbf{z}, \alpha \,|\, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\, dz^{(1)}\, dz^{(2)}\, d\alpha. \end{aligned}$$

Once again, we can use (4) and (5) to simplify, and so obtain the expression

$$\begin{aligned} &P(\mathbf{x}_{\text{next}} \,|\, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \\ &\quad \propto \iiint P(\mathbf{x}_{\text{next}} \,|\, \mathbf{z}, \alpha) P(\mathbf{x}_1 \,|\, \mathbf{z}, \alpha) P(\mathbf{x}_2 \,|\, \mathbf{z}, \alpha) \dots \\ &\qquad\qquad P(\mathbf{x}_n \,|\, \mathbf{z}, \alpha) H(\mathbf{z}) \pi(\alpha)\, dz^{(1)}\, dz^{(2)}\, d\alpha \end{aligned}$$

This approach makes the same independence assumptions about offender behavior as our fundamental result (6).

References

Baxter, M., & Rennie, A. (1996). *Financial calculus*. Cambridge University Press.

Beltrami, E. J. (1993). *Mathematical models in the social and biological sciences*. Jones and Bartlett Publishers.

Bennell, C., Snook, B., & Taylor, P. (2005, October). Geographic profiling- The debate continues. *Blue Line Magazine*, 34–36.

Bennell, C., Snook, B., Taylor, P., Corey, S., & Keyton, J. (2007). It's no riddle, choose the middle. *Criminal Justice and Behavior*, *34*(1), 119–132.

Bennell, C., Taylor, P., & Snook, B. (2007). Clinical versus actuarial geographic profiling strategies: A review of the research. *Police Practice and Research*, *8*(4), 335–345.

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*, 3–28.

Canter, D., Coffey, T., Huntley, M., & Missen, C. (2000). Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, *16*(4), 457–478.

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

Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis* (2nd ed.). Chapman & Hall / CRC.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury.

Chainey, S. (2005, August). Methods and techniques for understanding crime hot spots. In *Mapping crime: Understanding hot spots* (Tech. Rep., pp. 15–34). The National Institute of Justice, Washington DC.

Ewens, W. J. (2004). *Mathematical population genetics*. Springer Verlag.

Kocsis, R. N., Cooksey, R. W., Irwin, H. J., & Allen, G. (2002). A further assessment of circle theory for geographic psychological profiling. *The Australian and New Zealand Journal of Criminology*, *35*(1), 43–62.

Laukkanen, M., & Santtila, P. (2006). Predicting the residential location of a serial commercial robber. *Forensic Science International*, *157*(1), 71–82.

Levine, N. (2005). *The evaluation of geographic profiling software: Response to Kim Rossmo's critique of the NIJ methodology.* Retrieved June 2008, from `http://www.nedlevine.com/Response to Kim Rossmo Critique of the GP Evaluation Methodology.May 8 2005.doc`

Levine, N. (2009a). *CrimeStat: A spatial statistics program for the analysis of crime incident locations (v. 3.2).* Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. Retrieved July 2009, from `http://www.icpsr.umich.edu/crimestat`

Levine, N. (2009b). *Update notes to version 3.2. CrimeStat: A spatial statistics program for the analysis of crime incident locations.* Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. Retrieved July 2009, from `http://www.icpsr.umich.edu/crimestat`

Levine, N., & Block, R. (2010). Bayesian journey to crime estimation: An improvement in geographic profiling methodology. *Professional Geographer*, *to appear*.

Meaney, R. (2004). Commuters and marauders: An examination of the spatial behavior of serial criminals. *Journal of Investigative Psychology and Offender Profiling*, 121–137.

Mohler, G. O., & Short, M. B. (2009). *Geographic profiling from kinetic models of criminal behavior.* Retrieved July 2009, from `http://www.math.ucla.edu/ gmohler/geopro2.pdf`

Ratcliffe, J. H. (2006). A temporal constraint theory to explain opportunity-based spatial offending patterns. *Journal of Research in Crime and Delinquency*, *43*(3), 261–291.

Rich, T., & Shively, M. (2004). *A methodology for evaluating geographic profiling software* (Tech. Rep.). Cambridge, MA: Abt Associates.

Rossmo, K. (2000). *Geographic profiling*. CRC Press.

Rossmo, K. (2005a). *An evaluation of NIJ's evaluation methodology for geographic profiling software.* Retrieved June 2008, from `http://www.txstate.edu/gii/documents/Response to NIJ GP Evaluation Methodology.doc`

Rossmo, K. (2005b). Geographic heuristics or shortcuts to failure?: Response to Snook *et. al.* *Applied Cognitive Psychology*, *19*, 651–654.

Rossmo, K., & Filer, S. (2005, August / September). Analysis versus guesswork. *Blue Line Magazine*, 24,26.

Rossmo, K., Filer, S., & Sesley, C. (2005, November). Geographic profiling debate- round four. The big problem with Bennell, Snook and Taylor's research. *Blue Line Magazine*, 28–29.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. CRC Press.

Snook, B., Canter, D., & Bennell, C. (2002). Predicting the home location of serial offenders: A preliminary comparison of the accuracy of human judges with a geographic profiling system. *Behavioral Sciences & the Law*, *20*, 109–118.

Snook, B., Taylor, P., & Bennell, C. (2004). Geographic profiling:

The fast, frugal, and accurate way. *Applied Cognitive Psychology*, *18*(1), 105–121.

Snook, B., Taylor, P., & Bennell, C. (2005a, April). Man versus machine. The case of geographic profiling. *Blue Line Magazine*, 56.

Snook, B., Taylor, P., & Bennell, C. (2005b). Shortcuts to geographic profiling success: A reply to Rossmo. *Applied Cognitive Psychology*, *19*, 655–661.

Tseloni, A., Wittebrood, K., Farrel, G., & Pease, K. (2004). Burglary victimization in England and Wales, the United States, and the Netherlands. *British Journal of Criminology*, *44*(1), 66–91.

Wilmott, P. (1998). *Derivatives*. John Wiley and Sons.

Wilmott, P., Howison, S., & Dewynne, J. (1995). *The mathematics of financial derivatives*. Cambridge University Press.

Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, *70*(1), 164–168.