The Application of Geographical Analysis in Serial Crimes

Wang Xin1*, Qi Xinshe1, Li Guo1, GaoCuicui1, Wang Na1, Huang Ruiping1
Liu Rong2

1 Institute of information communication National defense university of science and technology, Xi’an, China
2 Air Force Engineering University, Xi’an, China

Email of all the authors:
*Corresponding Author: Wang Xin email: 39652970@qq.com; phone: 13772062967

Abstract: In this paper, we use bayes' theorem to generate two mathematical models for a series of crimes. This series of crimes is a dynamic process, so we regard it as JTC model. In this model, there are some major factors, such as the individual's vulnerability to crime, the attractiveness of the place of the crime and the utilization of the crime from the place of origin to the place of destination. We combine the bayesian theorem with the JTC model to obtain the mathematical model of geographical distribution. We made full use of the model to predict the location of the next perpetrator in the series.

1. Executive summary
In order to analyze the geographical distribution of serial crimes, this paper uses Bayes theorem to establish two mathematical models of serial crimes. Combining Bayes theorem with JTC model, the mathematical model of geographical section is obtained:

\[ P(\theta | X) \propto P(X | \theta) P(\theta) \]  

(1)

In order to apply the method, separate estimates have to be made for the prior probability \( P(\theta) \), the likelihood function \( P(X | \theta) \), and the data function \( P(X) \). Next, for the application to JTC analysis, \( P(\theta) \) is the probability that the offender lives at any location \( \theta \), based on the JTC distance decay function.

When we get more additional information that can be added to improve the estimate, our likelihood function \( P(X | \theta) \) will be more accurate. So administrators should collect as much crime location information and residence records as possible, which can increase the possibility of arresting criminals. And the data function \( P(X) \), that of obtaining the data, is the most difficult to estimate because there is no simple way of estimating the probability of obtaining the information under all possible scenarios.

Therefore, this method applies only to the molecules of the bayesian method, and the results are rescaled so that the sum of all probability estimates is equal to 1.0. Rough approximations can be obtained by estimating the likely sources of all perpetrators, regardless of the distribution of incidents committed by the serial perpetrators of concern. This is a general probability that applies to all serial offenders being evaluated. Assume that all of these probabilities apply to similar events, in this case, crimes committed by the perpetrator. A key research question concerns whether it needs to be applied to subsets of crimes such as robbery or burglary, or whether it can be applied more broadly to a variety
of crimes.

Then, we consider the problem in terms of other factors, such as where the criminal is hiding and how far the criminal is willing to travel. We still use Bayes' theorem to build another mathematical model of geographical distribution:

$$P(z | x_1, ..., x_n) \propto P(x_1 | z, \alpha) \cdots P(x_n | z, \alpha) \ H(z) \pi(\alpha) d\alpha$$  \hspace{1cm} (2)

Where the expression $P(z | x_1, ..., x_n)$ gives us the probability density that the offender has anchor point $z$ given that they have committed crimes at the locations $x_1, ..., x_n$, and $P(x_1 | z, \alpha) \cdots P(x_n | z, \alpha)$ means that all crime locations are mathematically independent because they have fixed points. Meanwhile, U.S. Census data provides community-level population counts and the land area of the community. So we can use this data and density parameter estimation technology to generate $H(z)$. Considering that our framework does not require a specific parameter derived from the prior distribution $\pi(\alpha)$, but can be directly constructed using appropriate empirical data. In addition, there is no need to make the same choice for $\pi(\alpha)$ for different types of crime. Again, we can choose which historical data to use when generating $\pi(\alpha)$.

Finally, we will make full use of the above model to predict where the perpetrator will commit a series of crimes. Therefore, the prediction function can be obtained:

$$P(x_{next} | x_1, x_2, ..., x_n) \propto \prod_{i=1}^{n} P(x_i | z, \alpha) P(x_1 | z, \alpha) \cdots P(x_n | z, \alpha) H(z) \pi(\alpha) dz^{(1)} dz^{(2)} d\alpha$$  \hspace{1cm} (3)

Therefore, we can see that we can calculate the probability that the criminal chooses the next potential location, based on the previous location of the crime that the criminal has committed.

2. Building the model

The bayesian approach is a good way to do this because it provides a system framework for incorporating other information. The model uses information about where other offenders have committed crimes to improve estimates of where particular serial offenders live. The way to do this is through an empirical bayesian (EB) approach.

In the statistical interpretation of bayes' theorem, probability is an estimate of a random variable. Let $\theta$ be a parameter of interest and let $X$ be some data. For example, in this special issue, $\theta$ is a location where an offender might live when multiple locations are being considered. Thus, we can express Bayesian Theorem as:

$$P(\theta | X) = \frac{P(X | \theta) \times P(\theta)}{P(X)}$$  \hspace{1cm} (4)

$P(\theta)$ is the probability of $\theta$ at all positions, usually called a priori probability. $P(X | \theta)$ is the probability of getting data when the given minimum probability is true, usually called likelihood function. It is the probability of getting data when the distribution of the given minimum probability is true. $P(X)$ is the marginal probability of the data, the probability of getting the data under all possible circumstances; Essentially, it's data. Finally, $P(X | \theta)$ is the probability of $\theta$ given the data, $X$ and is called the posterior probability. This is the result of combining the new information $X$ with the previous probability estimate. Since $P(\theta)$ is the prior probability of position $\theta$, for some new data $X$, Bayes' theorem can be used to update the probability estimate of $\theta$. The prior probability of $\theta$ can come from previous research, assuming no distinction between conditions that have any effect on $\theta$ (sometimes called “non-information prior”) or a mathematical distribution of the assumption. However interpreted, the result is a posterior probability (or posterior distribution) $x$ of the parameter $\theta$.

When the prior probability is completely estimated through data, the posterior probability is often called empirical Bayesian (EB) estimation, to distinguish it from bayesian analysis defined
subjectively or through a series of mathematical functions. (Carlin & Louis, 2000; Lee, 2004).

In general, the prior probability (denominator of the above formula) of the obtained data is unknown or difficult to calculate. Therefore, only the numerator is used to estimate the posterior probability, because

\[ P(\theta | X) \propto P(X | \theta) P(\theta) \]  

Where "\( \propto \)" means "in proportion to". The result is not necessarily equal to 1.0, which is a necessary condition for making it a probability, but it can be rescaled to produce a true probability estimate.

3. Explanation

Bayesian Theorem can be applied to the JTC methodology. To apply the method, separate estimates have to be made for the prior probability \( P(\theta) \), the likelihood function \( P(X | \theta) \), and the data function \( P(X) \).

(1) Prior probability

The JTC method generates estimates based on the assumed cost function. This feature can be derived from previous studies (Canter, 2003; Canter & Gregory (1994) estimates this by creating a sample of known offenders (Levine, 2004, chapter 10) or by assuming that each offender follows a specific mathematical function (Rossmo, 1995, 2000). Essentially, it is a prior probability that the offender lives at a particular location, \( \theta \), and is the equivalent of \( P(\theta) \) in equations (4) and (5). For the application to JTC analysis, \( P(\theta) \) is the probability that the offender lives at any location, \( \theta \), based on the JTC/distance decay function and is written as \( P(JTC) \).

(2) Likelihood function

However, estimates can be improved by adding additional information. In this article, we will investigate whether the residence information of other offenders has been incorporated, and these offenders have committed crimes at the same location as the serial criminals of interest. If you assign these positions to a group of areas, you can create a matrix that associates the original area with the target area. Generally, when implementing this method, a grid is overlaid on the study area, and these areas are grid cells. However, other zone types can be used (e.g. census tracts, traffic analysis zones). The result is an origin–destination matrix (also called O–D matrix or trip distribution matrix) (Levine, 2004, ch. 14). The number of crimes committed in all destination regions must equal the number of crimes committed in all source regions.

By concentrating only on the columns that correspond to the locations where the particular serial offender committed crimes, a sub-set of the O–D matrix can be extracted and a probability estimate made. This is a conditional JTC estimate and is the equivalent of \( P(X | \theta) \) in equations (4) and (5). That is to say, instead of using all target areas (all columns of the O-D matrix), only areas related to the location of the crime committed by the specific offender are used. The distribution of the origins from only those zones produces a probability distribution that is dependent (conditional) on the location where the one serial offender committed crimes.

Then, the marginal total number of conditional O-D matrices can be converted into probability by dividing the total number of each row by the sum of the marginal total numbers. In this application, the probability is written as \( P(O | JTC) \), the probability that other criminals living in the area commit crimes in the same location as serial criminals.

(3) Data probability

Finally, the third probability, the probability of obtaining data, is the most difficult to estimate because there is no easy way to estimate the probability of obtaining information in all possible situations. Therefore, this method is only applicable to the numerator of the Bayesian method, and the results are rescaled so that the sum of all probability estimates equals 1.0. A rough approximation can be obtained by estimating the possible sources of all offenders, regardless of the distribution of incidents committed by the consecutive offenders of interest. This is a common probability that applies
to all serial offenders who are estimating it. It is roughly similar to \( P(X) \) in equation (4), written as \( P(O) \)

By changing symbols, the Bayesian formula can be written as follows:

\[
P(JTC \mid O) \propto P(O \mid JTC) \times P(JTC)
\]

(6)

Among them, \( P(JTC \mid O) \) is the updated probability estimate value of the offender living in a certain place, \( P(JTC) \) is the probability obtained from JTC's own estimation, and \( P(O \mid JTC) \) is the probability of life of other offenders in the same place. \( P(O) \) is a rough approximation to the denominator of equation (4). Therefore, the product term in equation (6) can be divided by \( P(O) \) to obtain an approximation to the full Bayesian probability (Equation 6):

\[
P(JTC \mid O) \approx \frac{P(O \mid JTC) \times P(JTC)}{P(O)}
\]

(7)

This is called a Bayesian risk measure because it relates the posterior probability estimate (the "product" term in Equation 6) to the general probability. It is assumed that all these probabilities apply to similar events, in which case the crime is committed by the offender. A key research question involves whether it needs to be applied to a subset of crimes such as robbery or theft, or whether it can be applied more widely to various crimes.

4. Conclusion

The model introduced in this paper have a general supposition in consideration the difference of probability of the different areas.

We should modify the models according to the type of the serial criminal, the work or the address of the victims, the environment around the potential victims. We need more relational data on further study to get a model more applicable to the real situations, the more data there is, and the more accurate the result is. We need also take account of the effect of the criminal to the next one, actually this indeed have something with the probability of the area.

Reference:
[1] Mike O’Leary, (2009). The Mathematics of Geographic Profiling. Department of Mathematics, Towson University, Towson, MD, USA.
[2] Chun Gong & Zhenglin Wang, (2009). Proficient in MATLAB Optimization Calculation. Electronics Industry Press
[3] Ned Levine & Associates, (2009), Houston, TX, USA. Introduction to the Special Issue on Bayesian, J. Investig. Psych. Offender Profi l. 6: 167–185 (2009)
[4] Rengert, G. F., Piquero, A. R., & Jones, P.R. (1999). Distance decay re-examined. Criminology, 37(2), 425–427. Reprinted in D. Canter, & D. Youngs (Eds.), Principles of geographical offender profiling (pp. 177–193). Aldershot, Hampshire, UK: Ashgate Publishing Ltd.
[5] Hua Liu & Donald E. Brown, (2003). Criminal Incident Prediction Using a Point-pattern-based Density mode. CSG Systems, Inc., One Main Street, Cambridge, MA 02142, USA & Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA 22903, USA
[6] Zhonggeng Han, (2007). Practical Operational Research. Tsinghua University Press