A MATHEMATICAL FRAMEWORK FOR ANALYZING AND REPRESENTING RECUR AND NEAR-RECUR RESULTS IN BURGLARY CRIME DATA

M. VIJAYAKUMAR* AND C. CHENDRA SEKAR

1. INTRODUCTION:

Repeat victimization has recently emerged as a central focus in criminology. Research has demonstrated that individuals who have been victims of personal or property crimes are more likely to be victimized again (Farrell and Pease 2001). In the case of residential burglary, which focuses on in this paper, repeat victimization is described in terms of exact-repeat and near-repeat events (Johnson et al. 2007). Exact-repeat events are defined as consecutive burglaries occurring at the same location, separated by a time interval of any duration; near-repeat burglary events are instead classified as taking place within a set spatial neighborhood of a focal burglary point.

Repeat victimization may be due to a variety of reasons, including persistent spatial heterogeneity of risk and event dependence tied to the specific activities ofburglars (Tseloni and Pease 2003, 2004; Johnson 2008). When considering a complex urban environment, risk heterogeneity implies that some houses (or neighborhoods) are at higher risk than others, and that this difference is persistent throughout time. Some houses may be at higher risk because they are physically soft targets (e.g., easily forced doors or windows) or because the routine activities of inhabitants leave them much less secure than other homes. By contrast, event dependence suggests that some aspect of the burglar’s previous experience victimizing the house increases their preference to return. For example, the burglar may discover an abundance of items that could be targeted in a subsequent burglary, or they may simply prefer to return to a location where they know that their entry methods are guaranteed to work again (Farnet et al. 1997; Townley et al. 2003). It has been suggested that this elevated risk may spread to neighboring homes as well (Johnson et al. 1997; Townley et al. 2003; Sagovsky and Johnson 2007), especially in areas where nearby houses are similar in layout and type of inhabitant. Notice, however, that in the case of event dependence, burglary risk is not persistent throughout time, but may change as the burglar’s preferences, skills, and exposure to other opportunities change (Farrell et al. 1995). The concept of such biased repeat burglary carries strong implications for the dynamics of crime pattern formation and the development of prevention and resource allocation strategies (Bowers et al. 1998; Farrell and Pease 1993, 2001). Models based on the event dependence hypothesis (Johnson et al. 1998; Farell and Pease 1993, 2001) have been successful in explaining observed distributions of repeat burglaries, but they are typically based on the assumption that the burglar has a fixed set of preferences throughout the entire time period. This assumption is not realistic, as the preferences of burglars may change over time as their opportunities and methods of entry evolve. For example, the burglar may discover new targets or learn new techniques that allow them to enter houses more easily. In short, the concept of event dependence suggests that the risk of burglary is not constant, but rather changes as the burglar’s preferences and opportunities change.

The concept of event dependence is crucial for developing effective prevention and resource allocation strategies. By identifying the specific factors that increase the risk of burglary for a particular location, law enforcement agencies can target their efforts more effectively. For example, if the risk of burglary is higher in areas where the routine activities of inhabitants leave homes less secure, then efforts to improve the security of these areas could significantly reduce the risk of burglary. Similarly, if the risk of burglary is higher in areas where the burglar has previously been successful, then efforts to identify and respond to these areas could help to break the cycle of crime.

In this paper, we develop a mathematical framework aimed at analyzing repeat and near-repeat effects in crime data. Parsing burglary data from Promoter Apartments, Chennai according to different counting methods, we determine the probability distribution functions for the time intervals between repeat offenses. This paper compares these observed distributions to theoretically derived distributions in which the repeat effects are due exclusively to persistent risk heterogeneity. We find that risk heterogeneity alone cannot explain the observed distributions, while a form of event belief can. Using this information, we model repeat victimization as a series of random events, the likelihood of which changes each time an offense occurs. We are able to estimate typical time scales for repeat burglary events in Promoter Apartments by fitting our data to this model. Computer simulations of this model using these observed parameters agree with the empirical data.
Repeat burglary effects are often observed via the distribution of victimization order within a population of homes, where the victimization order is here defined as the number of times a home is burgled within some fixed temporal window. This distribution is typically inconsistent with a Poisson distribution, which is what would be expected if all homes had the same, persistent risk of burglary. In order to see whether event dependence may be responsible for this inconsistency, one often focuses on the distribution of time intervals $\tau$’s between successive events that occurred at the same location, a procedure that has been performed using burglary data from a variety of cities worldwide (Johnson et al. 2007). In general, it is observed that the distribution of time intervals between burglary events is a rapidly decaying function, with short time intervals much more likely to occur than longer ones. This observation has been taken as evidence for the existence of event dependence, and that a house will exhibit an increased risk of being burgled after being victimized once. However, there has not yet been a rigorous discussion as to why exactly these decaying time interval distributions support this hypothesis. In fact, as we will show in this paper, this observation alone does not necessarily support the event dependence hypothesis at all, and the method of counting the time intervals is of critical importance when interpreting the distribution of $\tau$.

Throughout the remainder of this paper, we will be performing analyses on a dataset which includes the geographic location and day on which each reported residential burglary for the years 2003–2005 occurred in Promoter Apartments, Chennai. This paper consider only those burglaries that occurred at single family homes, since we do not possess data that is detailed enough to pinpoint specific units within multi-family housing. Here, the term “single family homes” refers to stand-alone housing units (i.e., detached houses) with unique physical addresses as opposed to “multi-family housing”, which could be an apartment complex or condominium building where many separate units share the same physical address that belongs to the entire structure. In our analyses, we have ignored the influence of seasonality (Farrell and Pease 1994), specifically because the climate of Promoter Apartments minimizes such variations. However, all of the results and formulas can be modified to include seasonal effects in a straightforward manner. The dataset contains 9,042 events, and the distribution of victimization order across the homes is: 7,002 order one, 819 order two, 98 order three, 19 order four, 5 order five, and 1 order seven. According to the 2000 Indian census, there are between 70,000 and 80,000 occupied single-family homes in Promoter Apartments. Using this fact, and the use order distribution, the authors find that a simple Poisson distribution does not fit our Promoter Apartment data as well, indicating that something is indeed causing repeat victimization there.

The goal of this paper is to present a theory in which the repeat victimization is modeled to the environment; we refer to this as the random event hypothesis (REH; see also Nelson 1980). We then show, following from some very reasonable assumptions, that the distribution of time intervals $\tau$’s for exact-repeats in the REH is that of a sum of decaying exponentials, and that we should observe just this when using a affecting-window counting method on our data (which will be describe later). Using our burglary data, we illustrate that the observed distribution of $\tau$’s is completely compatible with that predicted by the REH, with the parameters of the fit interpreted as measures of risk heterogeneity. This compatibility, however, is not sufficient to prove the validity of the REH since other possible mechanisms of burglary dynamics might be equally compatible with the observed results. In fact, the parameters of the fit lead to a predicted distribution of home orders that is wildly different from that observed, indicating that the REH is insufficient to explain exact-repeat effects in our data. We then introduce a different method of counting exact repeat time intervals that allows us to unequivocally differentiate data sets generated via the REH from those in which burglary events are in fact related via event dependence, using only the time interval distribution. Applying this novel analysis method to our data set, we find that there is, in fact, event dependence in Promoter Apartments. We present a simple mathematical model with a straightforward criminological interpretation that explains the observations under both counting methods, and which reproduces both the time interval distributions and the home order distributions well in simulation. Finally, we extend some of these results to the measurement of event dependence in the near-repeat effect, finding that it too is present in our data.

2. THE RANDOM EVENT HYPOTHESIS:

The simplest assumption to make about burglary events at house $i$ is that they occur entirely at random, defining a stochastic process where each burglary event is independent of all others. We will call this model the REH. In addition, it is obvious that two burglary events at house $i$ cannot occur simultaneously, as they would then simply be thought of as one event. Mathematically, such a phenomenon can be modelled as a Poisson process characterized by a rate parameter $\lambda$, representing the expected number of burglary events per unit time. The characteristics of Poisson processes are long-established (Feller 1968), but we present now a brief summary of the main results which are relevant to our analysis. For a Poisson process with rate parameter $\lambda$, the probability that one burglary occurs within a time interval $t$ to $t + \Delta t$ is given by

$$P_1(\delta t) = e^{-\lambda \delta t} \lambda \delta t$$

(1)

The probability that $k$ burglaries occur is given by the general Poisson distribution

$$P_k = \frac{e^{-\lambda \delta t} (\lambda \delta t)^k}{k!}$$
The probability that no events occur within a time interval $\delta t$, then, is given by

$$P_0(\delta t) = e^{-\lambda \delta t}$$

(3)

a monotonically decreasing function of time.

Consider now the time $T_1$ until the first burglary occurs at house $i$, as measured from a reference point with $t = 0$. In this case, $T_1$ will be greater than a given time $\tau$ only if there have been no events within the time interval from 0 to $\tau$.

Hence,

$$P(T_1 > \tau) = P_0(\tau) = e^{-\lambda \tau}$$

(4)

Extending this result, we see that the probability that the first event occurs between times $\tau$ and $\tau + \delta t$ is

$$p(\tau(T_i); \tau + \delta t) = p(T_i; \tau) - P(T_i; \tau + \delta t) = e^{-\lambda \tau}(1 - e^{-\lambda \delta t})$$

(5)

If we divide this result by $\delta t$ and take the limit as $\delta t \to 0$, we arrive at the standard Poisson process probability density function for the time interval $\tau$ s between events:

$$e^{-\lambda \tau}$$

(6)

Therefore, if the REH is correct, the distribution of time intervals between exact-repeat events at a given home with rate constant $\lambda$ should follow an exponential decay of the type shown in Eq. 5. Since the distribution is a memoryless one, the distribution of events at a time interval of any duration would be the same, with no introduction of any notion of correlation between burglary events. In fact, this distribution will only hold if the events are statistically independent, a notion that is completely contrary to the typical assumptions of the event dependence hypothesis.

We do not, however, expect every home within a city to display the same burglary rate $\lambda$, as it is well known that crime rates may vary spatially (Ratcliffe and McCullagh 1999) and we have already shown that our Promoters Apartments data does not conform to a simple Poisson distribution. We therefore allow our homes to be divided into $N$ groups, each of which is characterized by a particular burglary rate $\lambda_i$ that is persistent in time for that group. If the fraction of homes exhibiting rate $\lambda_i$ is defined to be $\omega_i$, then the composite distribution of waiting times should be given by

$$p(\tau) = \sum_{i=1}^{N} w_i \lambda_i e^{-\lambda_i \tau}$$

(7)

which is just a weighted sum of the individual distributions for each group. Equation 7 might also be read as the mean waiting time for all houses, since the $w_i \lambda_i$ sum to 1. Equation 7 is therefore a mathematical representation of pure risk heterogeneity, in terms of the time intervals between exact-repeat events modelled via a compound Poisson process.

3. THE MOVING-WINDOW METHOD:

In order to test the distribution predicted by the REH, one must first develop a proper counting scheme for the time intervals $\tau$ s between exact-repeat events. Ideally, one would watch each burgled house within the city of interest until it is burgled again, and simply mark the time to repeat. However, this is clearly infeasible, as many homes will not be burgled again during a reasonable observation period. In fact, for our Promoter Apartments data set, out of the 7,944 unique locations burgled, only 942 of them were burgled more than once. If we were to only use the time intervals from these relatively few locations, we would likely introduce a bias into our count because we would be systematically discarding many time intervals which were at longer timescales and were, therefore, never observed.

To count properly, then, we use a method that we will call here the moving-window method. The basic idea behind this method is to first choose a time window of interest, $\tau_{\text{max}}$, and then to observe after each burglary event whether or not
another event occurs at that same location within this time window; let us use as an example a \( \tau_{\text{max}} \) of 727 days for our promoters apartment data set. If an event does indeed occur, the time interval \( \tau \) between the initial and secondary event is noted. Of course, any event which occurs within the last \( \tau_{\text{max}} \) days for which we have data cannot be subsequently watched over the full \( \tau_{\text{max}} \) window, as some of the window would clearly lie outside of the dates for which we have data. Therefore, we do not perform our observation following these events. We call this final \( \tau_{\text{max}} \) period within our data the “buffer interval”, which corresponds to the years 2004 and 2005 in our example. The final output of the count consists of the number of events \( N_0 \) for which an observation was performed (this is equal to the total number of events in our dataset minus the number of events that occur within the buffer interval) and a list of time intervals observed. Note that the number of time intervals recorded will in general not be \( N_0 \), since not every home that is observed will be subject to another burglary within our \( \tau_{\text{max}} \) window, as discussed above. Finally, we make a histogram of the observed \( \tau \), dividing the frequency for each histogram bin by \( N_0 \) to arrive at a probability distribution that we can directly compare to Eq. 7. It is in this way that the homes burgled only once affect our count—they contribute no time intervals, but they do increase \( N_0 \) and thereby influence the probabilities.

The results of such a moving-window count can be seen in Fig. 1, using our promoters apartments data. Here we have, as in our example above, chosen a \( \tau_{\text{max}} \) of 727 days, making the buffer interval roughly the years 2004 and 2005. The choice of 727 days is arbitrary, but was used because it is evenly divisible by our desired histogram bin width of 14 days. By adding together the probabilities for each histogram bin, we find that 10.8% of the events in our data set we follow to an exact repeat within 727 days.

Along with the observed \( \tau \) distribution, we have plotted in Fig. 1 (the solid line) a curve of the type shown in Eq. 7 with parameters chosen to give the best fit to our data. Using \( N = 3 \), we find the best fit to be \( w_1 = 0.915, \tau_1 = 5.32 \times 10^{-5}, w_2 = 0.066, \tau_2 = 2.45 \times 10^{-3}, w_3 = 0.019, \tau_3 = 8.41 \times 10^{-2} \). The choice of \( N = 3 \) was made simply because this was the smallest value for which a good fit of the curve to our data could be found. Both the \( N = 1 \) and \( N = 2 \) curves deviate too substantially from our data. For this choice, though, the REH curve fits our data rather well.

**THE FLAT-WINDOW METHOD:**

Although the Flat-window counting method is a valid approach, its corresponding null hypothesis curve as derived through the REH contains a large number of parameters, making it difficult to compare to observations in a meaningful way. In addition, as shown above, even if parameters can be chosen such that the REH curve fits the data very well,
further calculations are needed to interpret these results. In order to more easily determine the validity of the REH, we develop a counting method for which the null hypothesis curve is completely parameter free and that can by itself definitively confirm or deny the REH; we term this the Flat-window method. We first remind the reader that each home appearing within our data set can be classified by the number of times it was burgled in total over the D days of data available; we refer to this as the order of the house. The probability of any given home with burglary rate $\tau_i$ being of order k is given by Eq. 2, replacing $\delta$ with D. Note, however, that Eq. 2 is independent of the particular times at which the home was burgled, so long as there were a total of k events. This means, for example, that for order one homes, each of the D days is equally likely to be the day on which the one event occurred, assuming that $\lambda_i$ is persistent in time (i.e., seasonality is ignored and there is no event dependence). Similarly, for order two homes, each possible pair of days that can be made from our D day interval is equally likely to be the observed pair.

Suppose that we isolate all order two homes from our data set and ask how the time intervals $\tau_i$ between the two events at each of these houses ought to be distributed, assuming validity of the REH. Although all pairs of days are equally likely to occur within our trimmed data set, all time intervals are not. For example, with a fixed window of 1 year ($D = 365$), there are very many pairs of days that will lead to a s of just 1 day (Jan 1–Jan 2, Jan 2–Jan 3, etc.), while there is only one pair that exhibits the maximum time interval of 364 days (Jan 1–Dec 31). More generally, the number of pairs that will exhibit a time interval of $\tau$ s is given by

$$N_{\text{pairs}}(\tau) = D - \tau$$

one for each day that has at least $\tau$ days following it in the time window. We normalize these counts to one, so that the probability distribution of time intervals for order two homes is

$$P_2(\tau) = \frac{D - \tau}{D(D + 1)}$$

This particular distribution is specifically for order two homes. However, it can be shown that the probability distribution for order k homes of any order is given by using similar arguments:

$$P(k)(\tau) = \frac{K}{D+K-1} \prod_{l=0}^{k-2} \frac{D - \tau + l}{D + l}$$

using similar arguments.

There are two important points to make concerning the fixed-window method of counting. First, as seen in the moving-window method, if the REH is true and houses do not experience any increased risk of burglary after an initial event (i.e., there is no event dependence), the distribution of time intervals that we count will still be heavily weighted toward the short end of the spectrum, and will disappear at the long end (Fig. 2). Therefore, we must be mindful of this combinatorial effect when interpreting the results of a fixed window counting procedure. Second, and more positively, we see that the fixed-window method is an excellent way to test the REH, and thus the validity of risk heterogeneity for explaining repeat burglaries, because the expected probability distribution given by Eq. 10 has no free parameters. The fixed-window method therefore represents a simple and unambiguous way to determine whether or not the REH is in fact true, or if event dependence must be invoked.

To illustrate the usefulness of the fixed-window counting method in testing the REH, we present here the results of such a count, using our Long Beach data (Fig. 3). To perform this count, we first break up our data into six non-overlapping 364 day sets (we use $D = 364$ because it is a multiple of 14). For each set, we isolate the order two homes

![Graph showing probability distribution](image-url)
within that set, then count the time interval between the two burglaries for each of these homes. The resulting six \( \tau \) lists are then combined into one, a histogram is made of this master list, and the histogram bins are each divided by the total number of order two homes used in the count to convert to a probability that will add up to 1. When this histogram is compared with the null hypothesis curve as derived above in Eq. 9, we see that the disagreement is very substantial, with many more events occurring at short time intervals than predicted by the REH, and fewer, therefore, at long time intervals. Thus, the REH is disproven in our data without the need for any further analysis, and it is clear that event dependence must be responsible for at least some of the repeat victimization effect.

**NEAR-REPEATS:**

A near-repeat event occurs whenever two “nearby” houses are burgled within some period of time. Like exact-repeats, we can measure the time interval that lies between each event in a near-repeat pair, but in this case we must also make note of the physical separation of the two homes. This procedure allows us to examine separately the time interval histogram for near-repeat pairs that lie at varying physical distances from each other. It has been noted in previous studies that those near-repeat events that are relatively close in space tend to occur more closely in time as well, like exact-repeats, whereas those that are far apart seem to exhibit no temporal correlation. These previous studies use Monte Carlo algorithms to find the likelihood of the observed patterns happening if there were no correlation between the spatial and temporal distributions (Johnson et al. 2007; Ratcliffe and Rengert 2008), determining that this is highly unlikely. In this section, we instead test explicitly for near-repeat event dependence by extending our finite-window counting method used above for exact-repeats to the case of near-repeat events in our Promoters Apartments data.
The first step in our derivation is actually a fact that we notice from our data, which is supported by our model. We observe, when each year of our dataset is examined separately, that order one homes exhibit approximately equal probability of being burgled on all of the 365 days available. This is illustrated in Fig. 5, where we see the flat nature of the distribution of order one burglary days. This observation is consistent with the theoretical curve predicted from our exact-repeat model (also plotted in Fig. 5) which fits the data well when using our previously estimated parameters. The observed pattern arises from the fact that about 91% of homes that are burgled only once in \( D = 365 \) days will enter into the time window with the lowest burglary rate, \( \lambda_1 \), and still exhibit that rate after the event. These houses should all have a completely flat line in this count, because their rates are not changing and so each day is equally likely. The remaining 9% of homes consists almost entirely of those that either entered the year with the middle rate, \( \lambda_2 \), and then switched to \( \lambda_1 \) or those that did the opposite. These homes explain the slight increases in the theoretical curve near the beginning and end of the year. In addition, Fig. 5 shows the lack of seasonality in our dataset.

To test for the presence of near-repeat event dependence, we first isolate in our data all order one homes. We then perform a fixed-window count on these events in a pair-wise fashion, measuring both the temporal separation and physical distance between the burglaries comprising each possible pair of events. Note that this is essentially the same procedure as was performed for the exact-repeats earlier, except that in that case the

Fig. 5 Fraction of order one events occurring per day in Promoter Apartments, counting each year separately (white dots). The curve predicted by our model is also shown (white line). The distribution is essentially flat, with only slight increases at the beginning and end of the interval (shown magnified on the right). Most homes that are burgled only once will exhibit a constant rate of \( \lambda_1 \), leading to overall flatness. A small number of order one homes will transition from \( \lambda_2 \) to \( \lambda_1 \), or vice versa, which happens most often near the beginning and end of the interval; this explains the deviations from flatness in these regions.

The fact that order one homes are approximately equally likely to be burgled on any day of our fixed interval means that the time intervals for near-repeat events should be distributed exactly as in Eq. 10 if no correlation between the two burglaries making up a pair exists (i.e., if there is no event dependence). This is because, since each of the homes is equally likely to be burgled on any given day, each of the possible pairs of days making up a near repeat event ought to be equally likely as well, which is the condition that leads directly to Eq. 10.1
CONCLUSIONS:

The results here reinforce the view that repeat and near-repeat victimization may play a role in the nucleation of crime patterns in space and time and, as a consequence, may be an appropriate basis for designing crime prevention strategies.

However, we also note that there are a number of challenges yet to meet in designing optimized responses to repeat crimes. In particular, our results from analyses of exact-repeat burglaries in Promoter Apartments suggests that at any given time only about 0.002% of houses exhibit the highest excited state (Eqs. 11 and 12) with an expected time to a repeat event of approximately 12 days. This corresponds to only a single family residence from a total of approximately 70,000 units. The challenge is to determine which house(s) belong to this very small set, which would allow preferential targeting of resources at these locations.

REFERENCES:


A mathematical framework for analyzing and representing recur and near-recur results in burglary crime data / IJMA- 2(7), July-2011, Page: 1061-1069


See M.B. Short, et al., JQC 25 (2009)