Reducing bias in estimates for the law of crime concentration

George Mohler* P. Jeffrey Brantingham † Jeremy Carter ‡ Martin Short§

Abstract

Objectives. The law of crime concentration states that half of the cumulative crime in a city will occur within approximately 4 percent of the city's geography. The law is demonstrated by counting the number of incidents in each of N spatial areas (street segments or grid cells) and then computing a parameter based on the counts, such as a point estimate on the Lorenz curve or the Gini index. Here we show that estimators commonly used in the literature for these statistics are biased when the number of incidents is low (several thousand or less). Our objective is to significantly reduce bias in estimators for the law of crime concentration.

Methods. By modeling crime counts as a negative binomial, we show how to compute an improved estimate of the law of crime concentration at low event counts that significantly reduces bias. In particular, we use the Poisson-Gamma representation of the negative binomial and compute the concentration statistic via integrals for the Lorenz curve and Gini index of the inferred continuous Gamma distribution.

Results. We illustrate the Poisson-Gamma method with synthetic data along with homicide data from Chicago. We show that our estimator significantly reduces bias and is able to recover the true law of crime concentration with only several hundred events.

Conclusions. The Poisson-Gamma method has applications to measuring the concentration of rare events, comparisons of concentration across cities of different sizes, and improving time series estimates of crime concentration.

1 Introduction

The law of crime concentration states that a large percentage of crime falls within a small fraction of all the locations across the urban landscape [36]. The law of crime concentration at place is a derivative of three decades of scholarly attention to the spatial patterning of crime [9, 34, 39, 41]. The law is a quantitative estimation based on historical events. In his discussion of the law and through the use of crime data from cities of various sizes, Weisburd [36] established what he refers to as "spatial bandwidths" that account for disproportionately high volumes of cumulative crime within a city's geography. Specifically, he found that 50 percent of crime is confined to only $4.2-6.0$ percent of the street segments, or a bandwidth of roughly four percent of city geography. He found that 25 percent of crime was confined to $0.8-1.6$ percent of street segments, or a bandwidth of less than one and a half percent. Empirical patterning consistent with the law has been documented in a number of studies [4, 15, 19, 25, 36], though the specific bandwidths identified vary based on whether the spatial units of analysis are street segments [35, 42], addresses [34], or city grid cells (ranging 250 to 1,000 square feet) [19, 25]. From a theoretical perspective the law lends guidance to better understand the nature of crime and disorder across environmental [7, 8] and community structural factors [29, 30]. From an intervention and policy perspective, the law can be leveraged to allocate police resources [6] and

*Computer and Information Science, Indiana University - Purdue University Indianapolis, gmohler@iupui.edu

†Anthropology, University of California Los Angeles, branting@ucla.edu

‡School of Public and Environmental Affairs, Indiana University - Purdue University Indianapolis, carterjg@iupui.edu

§School of Mathematics, Georgia Institute of Technology, mbshort@math.gatech.edu

serves as a foundation to forecast crime [21, 24]. Indeed, the law provides a model on how crime concentration can be realized into crime prevention benefits.

Despite a wealth of promising evidence, refined methods to improve the validity and reliability of estimates of crime concentration are needed to maximize the intervention potential of this law, and to generalize the law to other measures of social problems across varying environmental, social, and political contexts. Weisburd [36, 37] articulates this demand in his solicitation of further research and improved analytical methods to identify where the law of crime concentration at place does, and does not, apply. Moreover, Bernasco and Steenbeek [4] draw attention to challenges of method fidelity when determining crime concentration. Indeed, the law of crime concentration at place is an axiom, and not a prescriptive set of guidelines for measurement outside of the expected cumulative crime and cumulative place bandwidths.

In its most simple form, the law of crime concentration at place is demonstrated by calculating the percentage of total crimes falling within $x\%$ of the city over an observation window (typically $x = 25\%$ or 50%) [36]. Commonly used estimators of crime concentration produce valid estimates when the number of crimes is large and exceeds the number of places. However, when crime counts are low and are exceeded by the number of places, biased estimates are produced. Specifically, crime concentration is estimated as higher, given low crime counts compared to high crime counts, when we expect them to be equivalent. This estimation bias is discussed in detail below, though this concern is broached here as this issue has gone largely unaddressed in research to date and is the focus of the present study.

In this article we introduce an estimator for the Lorenz curve and Gini index with significantly reduced bias for all data set sizes when the crime counts follow a negative binomial distribution. We first model crime counts as a negative binomial random variable, which has the benefit of having a Poisson-Gamma representation. Secondly, we provide analytic formulas for the Lorenz curve and the Gini index given a continuous probability density. We then combine these facts to provide an inference methodology where 1) the parameters of a negative binomial are estimated from the empirical count data and 2) the Lorenz curve and Gini index are then computed via numerical integration or through simulation from the fitted Gamma density. We illustrate the methodology with synthetic data generated by a stationary Poisson process, a self-exciting point process, and real homicide data from Chicago. We show that our estimator can correctly identify the law of crime concentration with just a few hundred events, whereas the empirical estimator commonly used in practice requires several thousand. This has implications for crime concentration analyses over time where event counts may be low in each time window. In our example, we show that yearly homicide concentrates at less than half the level of the estimator used in [36].

2 Need for Improved Concentration Estimation Methods for Low Event Counts

Many studies to date have avoided the estimate bias issue because their event counts far exceed the number of places given their unit of analysis. However, a few notable studies have made substantive contributions to the crime concentration literature based upon low count event data. Recently, Andresen et al. [2] examined assault, burglary, robbery, theft from vehicle, theft, motor vehicle theft, and other crime across 16 years in Vancouver. Street segments and intersections, their spatial units of analysis (18,445), exceeded the number of crime incidents, with the exception of theft, theft from vehicle, and burglary within beginning years of data. In another study using Vancouver data from years 1991, 1996, and 2001, Andresen and Malleson [3] estimate the spatial concentration of assault, burglary, robbery, sexual assault, theft, theft from vehicle, and motor vehicle theft across census tracts (110), dissemination areas (1,011), and street segments (11,730). Their reported crime counts indicate that assaults (in 1996 and 2001), theft (in 2001), and for all years robbery, sexual assault, and theft of vehicle all had lower frequencies than the total number of street segments. In some crime cases the number of incidents was quite small, such as sexual assault (between 440-672) and robbery (between 1,251-1,893).

Macbeth and Ariel [23] investigated cumulative crime and crime harm concentration in Northern Ireland from 2012-2014 using street segments. Though they found 50 percent of cumulative crime to concentrate in 2.5 percent of street segments, the total number of street segments (19,217) exceeded the number of total crime events in all study years; 2012 (18,269), 2013 (16,183), and 2014 (15,769). In their test of the spatial concentration of drug activity at Seattle street segments (a total of 24,023), Hibdon and Groff [18] leveraged call for service data from both emergency medical services (1,706 incidents) and police (3,716 incidents) in 2004. Their overall results suggested half of all drug activity events concentrated within just 1.11 percent of street segments. Lastly, Braga and colleagues [5] estimated the concentration of 7,359 firearm incidents on 28,530 street units across 29 years in Boston. Speaking directly to the issue at hand, the authors note "The fact that each year, on average, there are fewer than 254 ABDW-Firearm incidents among nearly 28,530 street units suggests that even a purely random distribution might produce the observed clustering" (p. 42).

Discussing these studies is not to suggest criticism, but to illustrate a need for methods improvement. This estimation bias issue is also likely to become a more common challenge given the rapid growth of crime concentration studies generally and a growing focus on more crime-specific estimates of spatial crime concentration [17, 33]. Studies of crime concentration within smaller cities are also expected to grow in coming years [36]. Thus, this estimation bias issue has direct implications as smaller cities exhibit lower crime counts than urban areas, and the most common unit of analysis – street segments – are often larger in small cities compared to urban areas. For example, in his study of crime concentration in multiple cities, Weisburd [36] reports descriptive statistics in his Table 2 that show that an average violent crime rate and street segment length of the five large cities is 6.62 violent crimes per 1,000 residents and a street segment length of 364 feet, compared to an average violent crime rate of 3.13 per 1,000 residents and a street segment length of 615 feet for the three small cities. Thus, not only do smaller cities have fewer incidents of crime, but the unit of analyses to capture these incidents are also larger and possibly further exacerbate estimate bias. When the concentration of crime over time is analyzed, event counts in spatial units over shortened time windows may have a larger frequency of zeros that inflate the estimate of crime concentration.

Lastly, an improved method for estimating the spatial concentration of crime for low event counts will help scholars and practitioners apply the law of crime concentration to event types beyond traditional crime counts. Common estimation methods used for crime concentration are also employed by scholars investigating patterns of offending and victimization [4]. Scholars are also beginning to estimate the concentration of events related to public health [40, 43] and social harm [23]. Accurately estimating low count events may also help inform the theoretical understanding of crime concentration as it would enable models to account for environmental factors that shape high-crime places [14]. Perhaps most importantly, accurate estimation of low event concentrations will help tailor more effective interventions. If improved estimates can generate more informative place-identification, police are better positioned to engage in hot spots policing. Moreover, low event counts can be better incorporated into problem-oriented approaches to reduce crime, disorder, and public health challenges – an approach proven to be an effective strategy within hot spots [6].

3 Commonly Used Estimators for the Law of Crime Concentration

The two common estimators used in criminology to empirically test the law are 1) points on the Lorenz curve [19, 25, 36] and 2) the Gini index [4, 15]. For ordered (greatest to least) crime counts $y_{(i)}$ in spatial units $i = 1, ..., N$, the statistic used to estimate a point on the Lorenz curve is,

$$\hat{L}(p) = \frac{\sum_{i=1}^{r} y_{(i)}}{\sum_{i=1}^{N} y_{(i)}}, \tag{1}$$

where $p = r/N$ is the fraction of cells or street segments flagged as hotspots. The estimator for the Gini index is given by,

$$\hat{G} = \frac{1}{N}\left(2\left(\frac{\sum_{i=1}^{N}(N+1-i)y_{(i)}}{\sum_{i=1}^{N} y_{(i)}}\right) - N - 1\right). \tag{2}$$

While both of the above estimators are consistent, they suffer from severe bias for low event counts. In the extreme case of observing one crime event over a short observation window, then 100% of crime is captured in $1/N$ cells on that day ($L(1/N) = 1$ and $G \approx 1$). However, if the observation window is expanded, then other cells will start to contain crime and $L(1/N)$ and G will decrease (for example, they will decrease to $1/N$ and 0 for a spatially uniform Poisson process).

Recent attempts have been made to address this problem. Curiel et al. [11] propose a method for addressing the small sample size problem by modeling the rate of crime as a function of population in each spatial unit. However, population estimates are often not available at the street segment and grid cell level (where the law of crime concentration is typically measured).

Curiel and Bishop [10] suggest modeling crime counts as a Poisson mixture $q_1\lambda_1 + ... + q_k\lambda_k$ where q_i are the mixture weights and λ_i are the Poisson means. They then provide a "rare event concentration coefficient" (RECC) for estimating the Gini index,

$$\hat{G} = \frac{1}{2\sum_{i=1}^{k} q_i\lambda_i}\sum_{i=1}^{k}\sum_{j=1}^{k} q_i q_j |\lambda_i - \lambda_j| \tag{3}$$

A statistic is not given for points on the Lorenz curve in [10], however we show in Section 5.2 how their method can be extended for this purpose. We find the Poisson-Gamma model is a better fit for Chicago homicide concentration, however in other situations the Poisson mixture may be a good alternative. Our point in this paper is that, for low to medium event counts, a model for the data needs to be used instead of the non-parametric estimators given by Equation 1 and Equation 2.

An adjusted Gini coefficient [4] has been proposed for small event counts,

$$G' = \max(\frac{1}{M}, \frac{1}{N})(2\sum_{i=1}^{N}(N+1-i)z_{(i)} - N - 1) - \max(\frac{N}{M}, 1) + 1 \tag{4}$$

where M is the total number of crimes, N is the total number of spatial locations, and $z_{(i)}$ is the rank order (greatest to least) proportion of crimes occurring in place i. We will show that this estimator, like Equation 2, also suffers from bias, albeit in the direction towards under-representing concentration. In particular, for low event counts the adjusted Gini estimator under-estimates the value of the Gini index.

There also has been work outside of criminology to improve estimators of the Gini index. An estimator for the Gini coefficient and its standard error may be obtained through OLS [16] and in [22] a simple method for approximating variance is given. In [22] the authors also review the literature on estimation of the Gini index, highlighting how results have been republished multiple times. Our hope here is to bring to light these issues in criminology, without claim that we are the first to tackle these problems.

4 Estimating Crime Concentration from a Poisson-Gamma Model of Crime Counts

4.1 Poisson-Gamma representation of the Negative Binomial

Let y_i, $i = 1, ..., N$, denote crime counts in each of N spatial units over an observation window of time $[0, t]$, containing a total of M crimes. We assume the rate of crime in each unit i is stationary with Poisson rate λ_i. Therefore, the probability of observing $y_i = y$ is given by:

$$P(y_i = y|\lambda_i) = \frac{(t\lambda_i)^y}{y!}e^{-t\lambda_i} \tag{5}$$

For large sample sizes (large t) when the majority of the y_i are non-zero, the non-parametric estimators in the previous section can be used. For small sample sizes we need a parametric model that can extrapolate into the zero-count cells. Therefore, we assume that λ_i themselves are random variables with continuous probability density $f(\lambda)$ defined on $[0, \infty)$. It is from this density $f(\lambda)$ that the law of crime concentration arises as f captures the heterogeneity of crime risk across space. Our results that follow extend to general $f(\lambda)$, but here we further assume that the density is Gamma:

$$f(\lambda) = \frac{\beta^k}{\Gamma(k)} \lambda^{k-1} e^{-\beta\lambda}. \tag{6}$$

Then it follows [20] that the counts y_i are negative binomial and that

$$P(y_i = y) = \frac{\Gamma(y+k)}{y!\Gamma(k)} \left(\frac{k}{k+\mu}\right)^k \left(\frac{\mu}{k+\mu}\right)^y, \tag{7}$$

where $\beta = kt/\mu$. Here k is the *size* or *dispersion* parameter that controls how over-dispersed the negative binomial counts are and μ is the mean of the negative binomial.

4.2 Integral formulas for the Lorenz curve and Gini index

Let $f(\lambda)$ be a continuous density on $[0, \infty)$ with mean ν and cumulative distribution function $F(\lambda)$. Then the Lorenz curve $L(p)$ is defined by,

$$L(p) = 1 - \frac{1}{\nu} \int_0^{1-p} F^{-1}(s) ds \tag{8}$$

and the Gini index is given by [44],

$$G = 2 \int_0^1 L(p) dp - 1. \tag{9}$$

We argue that the law of crime concentration should be defined in terms of Equations (8) and (9). Note specifically that Equation (8) and (9) are not directly based off of crime counts y_i, but rather the underlying distribution of intensities $f(\lambda)$. However, Equations (1) and (2) provide consistent estimators of (8) and (9) that converge as $t \to \infty$ in Equation (5). This is because, as t gets large, the strong law of large numbers states that almost surely [13],

$$\lim_{t \to \infty} \frac{y_i}{t} = \lambda_i. \tag{10}$$

Thus for large t we may replace Equation (1) and (2) with,

$$\hat{L}(p) = \frac{\sum_{i=1}^r y_{(i)}/t}{\sum_{i=1}^N y_{(i)}/t} \approx \frac{\sum_{i=1}^r \lambda_{(i)}}{\sum_{i=1}^N \lambda_{(i)}} \tag{11}$$

and

$$\hat{G} \approx \frac{1}{N} \left(2 \left(\frac{\sum_{i=1}^N (N+1-i)\lambda_{(i)}}{\sum_{i=1}^N \lambda_{(i)}} \right) - N - 1 \right), \tag{12}$$

which are themselves estimators of (8) and (9). However, new methods of estimation are needed for small t. Specifically, the intensities are not directly observed, and must be estimated from crime counts. The problem we then consider is estimation of (8) and (9) from a finite, and likely small, sample of crime counts that follows a negative binomial distribution.

One important property of the Lorenz curve to note is that in the case of the Gamma distribution the Lorenz curve only depends on the dispersion parameter k [28]. Thus the time parameter t in Equation (5) that also appears in the scale parameter β does not play a role. Also, the Gamma density Equations (8) and (9) do not have analytic solutions, but they can be approximated via simulation or numerical integration.

4.3 Estimation of the law of crime concentration for small t

When t is small enough in Equation (5), the majority of the empirical counts y_i will be zero, yielding a biased estimate of the law of crime concentration if employing Equations (1) and (2) directly. However, we can use Equations (8) and (9) to estimate the Lorenz curve and Gini index for small t, so long as we can accurately estimate the parameters k and β of the Gamma distribution of intensities from our event counts y_i. Given count data defined on grid cells or street segments our general procedure then is as follows:

Estimation Procedure

1. Estimate negative binomial size k and mean μ from the count data.
2. Define a Gamma distribution with shape k and rate kt/μ.
3. Compute the law of concentration using numerical integration or simulation from Equations (8) and (9).

In step 1 above, the parameters of the negative binomial can be estimated via maximum likelihood where the log-likelihood is given by,

$$\mathcal{L} = \sum_{i=1}^{N} \left[\log(\Gamma(y_i + k)) - \log(y_i!) - \log(\Gamma(k)) + k \log\left(\frac{k}{k+\mu}\right) + y_i \log\left(\frac{\mu}{k+\mu}\right) \right]. \tag{13}$$

After taking the partial derivative of $\mathcal{L}$ with respect to μ and setting equal to zero, one finds that the maximum likelihood estimate is $\hat{\mu} = M/N$. A similar procedure for finding $\hat{k}$ unfortunately does not provide an analytical solution, but Newton's method can be used to solve for the parameter (in the MASS library of R the function fitdistr can be used and we provide example code in the appendix).

In step 3 above, either simulation or numerical integration can be used to compute statistics related to the law of crime concentration. In the case of simulation, a Gamma random variable g_i is simulated in each of the N spatial units (the R function rgamma can be used and we provide example code in the appendix). Then Equation (1) or Equation (2) are estimated using the g_i. For example, a point on the Lorenz curve is estimated as,

$$\hat{L}(p) = \frac{\sum_{i=1}^{r} g_{(i)}}{\sum_{i=1}^{N} g_{(i)}}, \tag{14}$$

and the Gini index is estimated as,

$$\hat{G} = \frac{1}{N}\left(2\left(\frac{\sum_{i=1}^{N}(N+1-i)g_{(i)}}{\sum_{i=1}^{N} g_{(i)}}\right) - N - 1\right). \tag{15}$$

It should be noted that the small sample size zero count issue is no longer a problem in Equation 14. This is because the Gamma distribution is continuous and the g_i are all non-zero.

Alternatively, numerical integration can be used to estimate statistics related to the law of crime concentration. For example, the integral in Equation 8 can be approximated as,

$$L(p) = 1 - \frac{1}{\nu}\int_0^{1-p} F^{-1}(s)ds \approx 1 - \frac{1}{\nu}h \sum_{l=1}^{(1-p)/h} F^{-1}(l \cdot h), \tag{16}$$

and the integral in Equation 9 can be approximated as,

$$G = 2\int_0^1 L(p)dp - 1 \approx 2h \sum_{l=1}^{1/h} L(l \cdot h) - 1, \tag{17}$$

for some small parameter h.

Note that in step 2 we technically need the value of t. But, as we pointed out above, the shape parameter k is what determines the Lorenz curve for a Gamma distribution, and β plays

no role. Hence, in practice we simply use a value $\beta = 1$ without loss of generality. In the Appendix at the end of the paper, we provide R code for estimating the Lorenz curve and Gini index from negative binomial count data using simulation (as used in Figure 7) and numerical integration (as used in Figure 8); we find both work equally well in practice. We note that Equation (8) and (9) are general and do not apply only to the Gamma distribution (thus this methodology would apply to other Poisson mixture models). In the experiments that follow, we will show that even for small t corresponding to a few hundred events this procedure works well in recovering the correct law of crime concentration.

5 Results

5.1 Simulation study

In our first experiment we simulate a Poisson-Gamma random variable with shape $k = 0.82$ and rate $\beta = 1/7.28$ for $N = 1000$ cells and varying t. To generate a synthetic data set we first simulate $N = 1000$ Gamma random variables λ_i representing the Poisson parameter in each cell i. Next we generate a Poisson process in cell i on $t \in [0, 4]$ by first drawing a Poisson random number E_i with mean $4\lambda_i$ and then distributing E_i event times uniformly on $[0, 4]$. We then compute each estimator on the data set for increasing t (using all data up to time t). We repeat this simulation for 50 synthetic data sets.

In Figure 1 we plot our Poisson-Gamma estimator (using simulation) against the estimator given by Equation (1) for 5% of cells flagged ($L(0.05) \approx 0.22$ on the Lorenz curve) as the data set size increases. We see that the empirical concentration estimator approaches 1 at small event sizes and only approaches the correct value for several thousand events. For even a few hundred events the Poisson-Gamma estimator is close to the true value of ≈ 0.22. We point out here that the inaccuracy of the empirical non-parametric estimators given by Equation 1 and Equation 2 are due to bias rather than variance, as every curve is above the true value of the statistic being estimated.

In Figure 1 we repeat the same experiment for the Gini index. We also include the adjusted Gini estimator proposed in [4]. We note that the adjusted Gini index is biased downward, under-estimating crime concentration at low events, then eventually converges to the empirical Gini estimator and over-estimates crime concentration before converging to the correct value.

5.2 Study of homicide counts in Chicago (2001-2017)

Next we apply our methodology to homicide count data in Chicago. The data is an open data set that can be downloaded at the Chicago open data portal [1]. There are 8911 homicides in the data set in the years spanning 2001 to 2017. We divide Chicago into a grid of 6524 cells, each of size 1000x1000 ft^2.

First we generate multiple homicide data sets by shuffling the event times in the original data set (this step allows for the quantification of uncertainty and also enforces stationarity of the process). Note that this procedure will not change the number of events in each cell, only when those events occurred. Further, if the intensity λ_i of each cell is stationary, then all events are uniformly distributed in time, and shuffling the times of events will retain this characteristic. Hence, each shuffled dataset will retain the same distribution of intensities $f(\lambda)$, and should yield the same crime concentration. Next we repeat the same exercise as above, computing the estimators for the law of crime concentration as the data set size increases. In Figure 2 we plot the estimates for $L(0.05)$ and G as we did for the simulation study in the previous section. We note that the Poisson-Gamma estimator is stable, predicting concentration at about 15% of crime in 5% of the city, even for a few hundred events. After 6000 events the empirical estimator qualitatively appears to match the Poisson-Gamma estimator in Figure 2. However, for the Gini index the empirical and adjusted Gini estimators still have not converged to our Gini estimate that is consistently less than 0.75.

To understand when the empirical estimator would converge, we perform a simulation analogous to that of the previous section, but with parameters chosen to match the Chicago dataset.

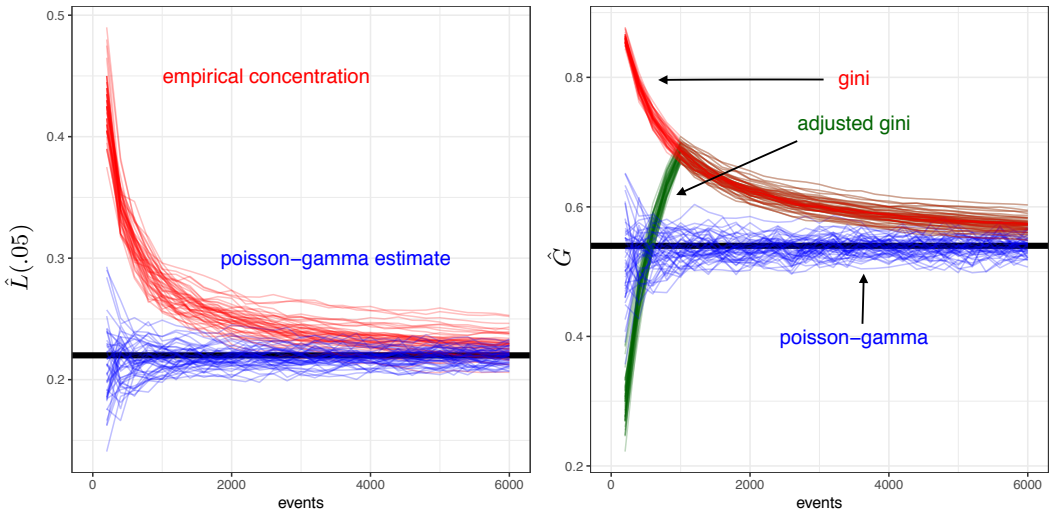

Figure 1: Simulated Poisson-Gamma process on 1000 cells. Left: Empirical concentration (fraction of events in top 5% of cells) vs number of events included in the analysis in red. Poisson-gamma estimator in blue. True value in black. Right: Empirical Gini index vs number of events included in the analysis in red. Adjusted Gini index in green and Poisson-gamma estimator in blue. True value in black.

In particular, we estimate the Poisson-Gamma model on the entire Chicago data set, yielding estimates of .334 and 4.1256 for the shape and scale. In Figure 3 we see that over 20,000 events are needed for the empirical Gini index simulation curves to match those of the Poisson-Gamma estimator.

In Figure 4 we plot the best fit negative binomial curve against the empirical density (normalized histogram) of homicide counts to verify that the Poisson-Gamma model is providing a good fit. The Chicago data is not perfectly fit by a negative binomial, however the fit is sufficient to accurately estimate parameters related to the law of crime concentration.

5.2.1 Comparison to a Poisson mixture

We also compare the Poisson-Gamma method to the Poisson mixture model proposed in [10]. As suggested in [10] we use the CAMAN R package [32] to estimate the Poisson mixture weights q_i and means λ_i. To estimate a point on the Lorenz curve given these parameters, we again assume crime is stationary and that the number of events in a grid cell is given by a Poisson process with intensity λ over time t, where λ is generated by the Poisson mixture model. To estimate the Lorenz curve we use simulation rather than numerical integration. In particular, we use a sufficiently large t to avoid the zero-count cell issue and simulate the estimated Poisson mixture process until the total number of events is as large as necessary to reach convergence in the empirical Lorenz estimator; this is very similar to the simulation approach for the Poisson-gamma model.

In Table 1 we compare the mean absolute percent error (MAPE) for estimation of Chicago homicide concentration in the top 5% of cells vs. number of events used to estimate the Poisson-mixture model (CAMAN) and the Poisson-Gamma model. The empirical concentration of Equation 1 using all $M = 8911$ events is used as ground truth, and the MAPE is calculated

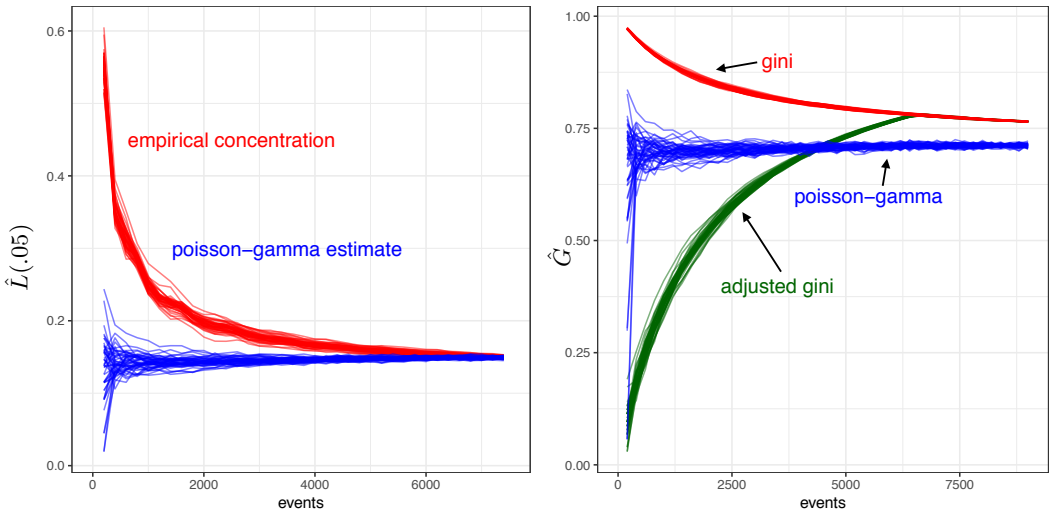

Figure 2: (Left) empirical concentration of homicides in Chicago vs. number of sampled homicides in red. Poisson-gamma estimator in blue. (Right) empirical Gini index vs number of sampled homicides in red, adjusted Gini index in green, Poisson-gamma estimator in blue.

for 200 bootstrap samples of the Chicago data wherein event times are randomly shuffled as described above. Here we see that for low event counts (less than 200), the Poisson-Gamma model yields a better estimator in terms of MAPE, however the Poisson-mixture and Poisson-Gamma are both significantly more accurate than using Equation 1 when M is low. The performance of a method will likely depend on the dataset in question (and sample size), so model selection, for example via the AIC, could be used.

5.2.2 Analysis of concentration over time

Next we illustrate our methodology on Chicago homicide concentration time series. This type of analysis is typically conducted to illustrate the stability of crime concentration over time [19, 25, 36]. However, in such an analysis the time window is shortened (i.e. small t) and thus there is greater risk of over-estimating crime concentration. In Figure 5 we plot the Lorenz and Gini estimators applied to homicide counts for each year from 2001 to 2017. Here we see that the Poisson-Gamma estimator is stable, not only in time, but in comparison to the estimator on the entire data set (both around 15%). However the empirical estimator puts crime concentration per year at around 35%, a significant over-estimate. Similarly, the adjusted Gini index under-estimates crime concentration by a factor of 2.

5.3 Estimating crime concentration for a Hawkes process

The method established in this work has very specific assumptions underlying it - that the process is stationary and that the crime counts are distributed according to a negative binomial. However, there is abundant empirical evidence that the space-time patterns of many types of crime can be described by self-exciting point processes (see for example [24] and the references within). Here, we relax the assumptions used in Section 4 by performing a simulation study in which the individual location's rates are not stationary Poisson processes, but are rather Hawkes

Table 1: Mean absolute percent error (MAPE) for estimation of Chicago homicide concentration in the top 5% cells vs. number of events in the data set. Empirical concentration with all $M = 8911$ events used as ground truth. MAPE calculated for 200 bootstrap samples.

M (# events)	Empirical	CAMAN	Poisson-Gamma
50	585.42	213.80	72.27
100	585.42	108.12	38.23
150	384.52	65.57	25.05
200	279.23	40.57	20.72
250	219.01	25.49	18.06
300	180.85	13.87	15.34
350	156.05	9.33	13.07
400	139.36	7.58	11.36
450	127.70	8.49	9.87
500	120.12	10.93	9.12
550	114.80	11.51	8.54
600	111.40	12.09	7.98
650	109.74	11.27	7.57
700	108.18	11.54	7.14
750	103.27	11.61	6.69
800	95.14	10.88	6.61
850	87.43	10.81	6.13
900	80.89	10.35	6.10
950	75.38	9.94	5.75
1000	70.70	10.03	5.39

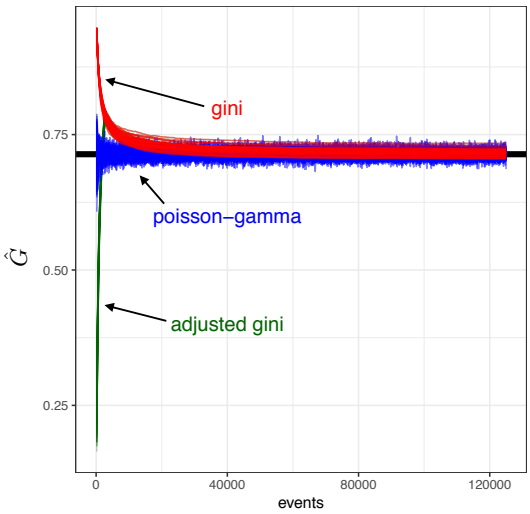

Figure 3: Simulated Poisson-Gamma process with MLE Poisson-Gamma parameters (shape .334 and scale 4.1256) estimated from Chicago homicide data (with 1000x1000 ft^2 cells). Empirical Gini index vs number of events in red. Adjusted Gini index in green and Poisson-gamma estimator in blue. True value in black.

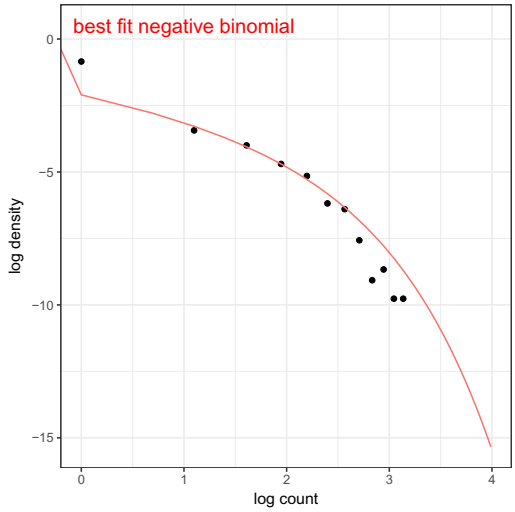

Figure 4: Log-log plot of the empirical density (normalized histogram) of Chicago homicide counts (black circles) and best fit negative binomial (red).

processes [24]. The Hawkes process is self-exciting, in that every event may excite further events to occur, such that all events can be classified as either background events, which are not caused by any prior event, or daughter events, which are events that are directly caused by some other

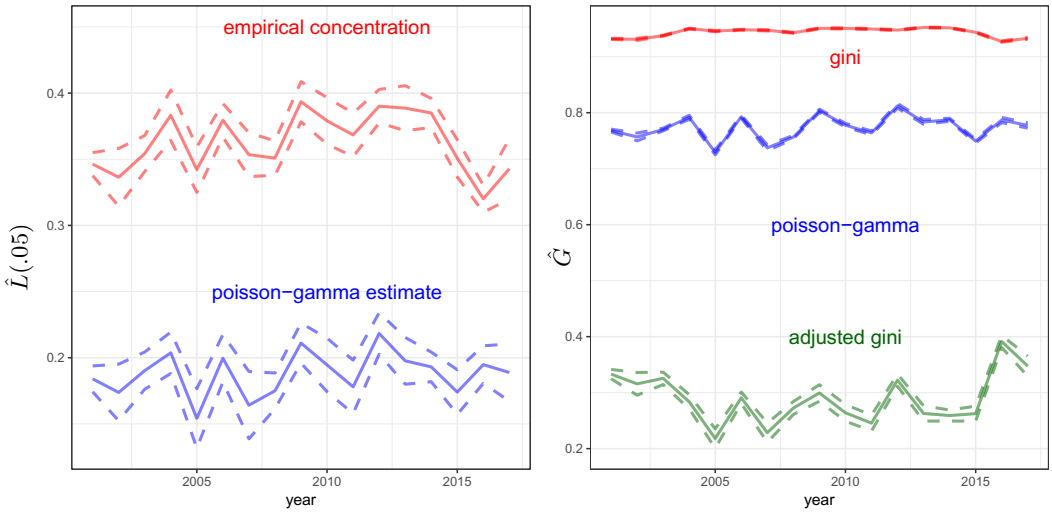

Figure 5: (Left) Concentration of homicides in Chicago in top 5% of hotspots in each year for the empirical estimator (red) and Poisson-Gamma estimator (blue). (Right) Gini index of homicides in Chicago in each year for the empirical Gini estimator (red), adjusted Gini estimator (green) and Poisson-Gamma estimator (blue).

event, referred to as its parent. Here, the intensity at location i, $\lambda_i(t)$ is given by

$$\lambda_i(t) = \mu_i + \sum_{t_{ij}<t} \theta\omega e^{-\omega(t-t_{ij})} \ , \tag{18}$$

where t_{ij} is the time of the jth event at location i, μ_i is the rate of background events at location i, θ is the excitation parameter that measures how many daughter events in expectation are triggered by any other event, and ω sets the timescale over which these excited events will occur following their parent. We point out that here we consider specifically a case in which each location may have its own background intensity μ_i, but that all locations share the same θ and ω parameters.

While certainly Eq. 18 generally causes nonstationary intensities at each location, it is also true [12,31] that the steady-state expected intensity for the Hawkes process is given by $\mu_i/(1-\theta)$, while the steady-state variance of the intensity is given by $\omega\theta^2\mu_i/2(1-\theta)^2$. This fact motivates our numerical experiment. Specifically, we simulate 50 realizations of a Hawkes process with $\theta = 0.7$ and a chosen ω, and where the background rates μ_i are Gamma distributed with shape $k = 0.82$ and rate $\beta = 1/7.28$ for $N = 1000$ locations; these parameters are identical to those used in Section 5.1. Then, the steady-state expected intensities of the Hawkes processes will also be Gamma distributed, and the variance of that distribution will be proportional to ω. We would therefore expect that for small values of ω, our estimation procedure should still work quite well for the Hawkes process, as each location will have an approximately stationary intensity that is drawn from a Gamma distribution. In fact, for $\omega = 0.5$, the results are nearly identical to those of Fig. 1, and the estimation procedure presented here gives a significantly less biased estimator than the alternatives considered for the Hawkes process.

However, for $\omega = 50$, none of the methods presented here gives a very good estimate until the total number of events M is quite large; see Fig. 6. This is to be expected, as the Hawkes process with this ω value has a very large variance, and so is not well approximated by a stationary process, so that many events must be considered before the asymptotic true value is achieved for the various estimators. However, this particular obstacle is surmountable in this case by pre-processing the event data before attempting the estimation via the methods above. Specifically, we can use stochastic declustering [45] to probabilistically identify which of the events in the

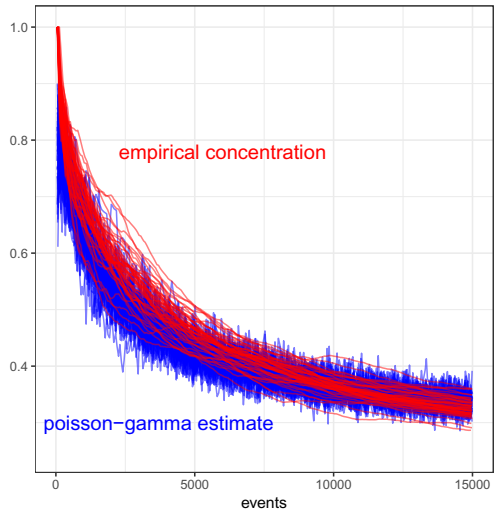

Figure 6: Simulated Hawkes process on 1000 cells. Empirical concentration (fraction of events in top 5% of cells) vs number of events in red. Poisson-Gamma estimator in blue.

Table 2: Estimation of $L(.05)$ with Poisson-Gamma estimator along with Declustered Poisson-Gamma Estimator for Hawkes process with 2500 and 5000 events. For comparison we include the empirical concentration estimator applied to the Hawkes events along with the empirical estimator applied to the declustered events.

# Events	True $L(.05)$	Poisson-Gamma	PG Declust.	Empirical	Emp. Declust.
2500	0.219	0.517 ± 0.032	0.286 ± 0.036	0.565 ± 0.034	0.388 ± 0.041
5000	0.219	0.427 ± 0.023	0.203 ± 0.013	0.462 ± 0.025	0.271 ± 0.013

dataset are background events, whose intensities are indeed the stationary μ_i, then run the concentration estimators on those background events alone. More specifically, the stochastic declustering technique involves estimating the parameters of the Hawkes process along with a probabilistic indicator p_b for each event as to whether it is a background event (and probabilistic indicators as to which event each event may be the daughter of); for our purposes we simply threshold these p_b so that any event with $p_b \geq 1/2$ is considered background, while all others are not. This process is computationally intensive, so we only present results for a small number of total event values M in Tables 2 and 3. The results show that the declustering technique allows for estimates that are significantly closer to the true value than for the non-declustered data, and that the Poisson-Gamma estimator in this case shows less biased estimates than the other methods considered.

Table 3: Estimation of Gini index with Poisson-Gamma estimator along with Declustered Poisson-Gamma Estimator for Hawkes process with 2500 and 5000 events. For comparison we include the empirical Gini estimator applied to the Hawkes events along with the empirical estimator applied to the declustered events.

# Events	True Gini	Poisson-Gamma	PG Declust.	Empirical	Emp. Declust.
2500	0.544	0.842 ± 0.014	0.643 ± 0.044	0.864 ± 0.012	0.778 ± 0.021
5000	0.544	0.789 ± 0.013	0.511 ± 0.027	0.809 ± 0.011	0.673 ± 0.014

6 Discussion and Conclusion

That a small fraction of locations in an urban landscape typically accounts for a large fraction of crime is widely observed and potentially warrants a claim for law-like status. A law of crime concentration implies that universal causal mechanisms are at play and general crime prevention measures might work across temporal or geographic contexts. Thus a law of crime concentration is of great theoretical and practical importance [36].

While crime concentration is widely observed, the two most common measurement methods fail to correctly estimate crime concentration for low crime counts. The concentration of crime is variously overestimated (Lorenz, Gini) or underestimated (adjusted Gini), depending upon the estimator in question, given small M (crime) and large N (segments or cells). Such estimation errors can significantly impede the ability to conduct comparative research. Crime concentration estimated from small M large N settings cannot be compared without bias. Thus, crime concentration from small cities, or low crime rate sub-areas of larger cities, cannot be readily compared with that from large cities. This is a problem given that of the 9,579 cities reporting to the FBI in the U.S. in 2016 for Part I crimes, 88.2% (8448) of them reported <1000 combined property and violent crimes [26]. Also, crime concentration estimated at fine spatial and temporal scales cannot be readily compared with that estimated at coarser scales.

An inability to compare settings without bias may impact how we model crime causation, as well as what policy decisions are made in response to crime. For example, the importance of micro-spatial crime attractors might be overstated (understated) given an overestimate (underestimate) of crime concentration computed for small temporal windows. Community or policing resources might be overallocated (underallocated) to certain places or times given an overestimate (underestimate) of crime concentration.

To address these concerns we introduced a method for estimation of parameters arising in the law of crime concentration with significantly reduced bias. The method has applications to measuring the concentration of rare events and also to improving time series estimates of crime concentration. We restricted our attention to the negative binomial distribution. In this case the estimation procedure is simplified because of the Poisson-Gamma representation and the law of concentration can be estimated from the inferred Gamma using numerical integration or simulation.

In the existing crime concentration literature, there are often underlying assumptions on the distribution of the data (for example stationarity) that are not explicitly stated. In our work we have made our mathematical assumptions explicit in terms of stationarity of the process generating crime and the distribution of counts (negative binomial). Whether these assumptions hold or are violated in different data sets will have implications for the law of crime concentration. Our method will work for any type of spatial unit (place, segment, grid cell, etc.) if the assumptions of iid counts, stationarity, and negative binomial hold. In situations where the data is not negative binomial, a Poisson mixture model may still be employed with an arbitrary prior $f(\lambda)$. Depending on the form of $f(\lambda)$ parametric methods may still be used, though if $f(\lambda)$ is multi-modal then non-parametric methods may be needed.

We have also not provided a rigorous mathematical analysis of the bias and variance of the estimators detailed in this paper, instead relying on simulation. We believe that the accuracy of each method is a function of the number of events and the number of spatial units and it would be interesting to develop a theoretical understanding for the relationship. The size of the spatial unit also plays a role in the calculation of statistics related to the law of crime concentration [27, 38]. The smaller the spatial unit, the greater the concentration as measured by the empirical estimators given by Equation 1 and 2. However, this higher concentration is at least in part due to the large number of zero-count units of measurement and our methodology will be useful in the analysis of concentration in these small areas.

Another consideration is the near-repeat or self-exciting phenomena observed in certain types of crime events. Here we have only addressed one specific case of this type of process, showing through simulation that our method can still be used directly in some cases with good results, while in other cases it can only be used after a further data pre-processing step has been completed. Of course, many other types of self-exciting or near-repeat processes could also be explored, and more advanced estimators may need to be developed for inference of the law of

crime concentration in these scenarios. These are directions for future research.

7 Acknowledgements

This work was supported in part by NSF grants SCC-1737585, SES-1343123, ATD-1737996, and ATD-1737925.

References

[1] https://data.cityofchicago.org/.

[2] Martin A Andresen, Andrea S Curman, and Shannon J Linning. The trajectories of crime at places: understanding the patterns of disaggregated crime types. *Journal of quantitative criminology*, 33(3):427–449, 2017.

[3] Martin A Andresen and Nicolas Malleson. Testing the stability of crime patterns: implications for theory and policy. *Journal of Research in Crime and Delinquency*, 48(1):58–82, 2011.

[4] Wim Bernasco and Wouter Steenbeek. More places than crimes: implications for evaluating the law of crime concentration at place. *Journal of quantitative criminology*, 33(3):451–467, 2017.

[5] Anthony A Braga, Andrew V Papachristos, and David M Hureau. The concentration and stability of gun violence at micro places in boston, 1980–2008. *Journal of Quantitative Criminology*, 26(1):33–53, 2010.

[6] Anthony A Braga, Andrew V Papachristos, and David M Hureau. The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice quarterly*, 31(4):633–663, 2014.

[7] P Jeffrey Brantingham. Crime diversity. *Criminology*, 54(4):553–586, 2016.

[8] Patricia L Brantingham and Paul J Brantingham. A theoretical model of crime hot spot generation. *Studies on Crime & Crime Prevention*, 1999.

[9] Paul J. Brantingham and Patricia L. Brantingham. *Patterns in Crime*. Mackmillan, New York, 1984.

[10] Rafael Prieto Curiel and Steven Bishop. A measure of the concentration of rare events. *Scientific reports*, 6:32369, 2016.

[11] Rafael Prieto Curiel, Sofía Collignon Delmar, and Steven Richard Bishop. Measuring the distribution of crime and its concentration. *Journal of Quantitative Criminology*, 34(3):775–803, 2018.

[12] José Da Fonseca and Riadh Zaatour. Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 34(6):548–579, 2014.

[13] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume I: Elementary theory and methods*. Springer, 2003.

[14] John E Eck, Ronald V Clarke, and Rob T Guerette. Risky facilities: Crime concentration in homogeneous sets of establishments and facilities. *Crime prevention studies*, 21:225, 2007.

[15] John E Eck, YongJei Lee, O SooHyun, and Natalie Martinez. Compared to what? estimating the relative concentration of crime at places using systematic and other reviews. *Crime Science*, 6(1):8, 2017.

[16] David EA Giles. Calculating a standard error for the gini coefficient: some further results. *Oxford Bulletin of Economics and Statistics*, 66(3):425–433, 2004.

[17] Cory P Haberman. Overlapping hot spots? *Criminology & Public Policy*, 16(2):633–660, 2017.

[18] Julie Hibdon and Elizabeth R Groff. What you find depends on where you look: Using emergency medical services call data to target illicit drug use hot spots. *Journal of contemporary criminal justice*, 30(2):169–185, 2014.

[19] John R Hipp and Young-An Kim. Measuring crime concentration across cities of varying sizes: Complications based on the spatial and temporal scale employed. *Journal of quantitative criminology*, 33(3):595–632, 2017.

[20] Samuel Karlin. *A first course in stochastic processes*. Academic press, 2014.

[21] Leslie W Kennedy, Joel M Caplan, and Eric Piza. Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies. *Journal of Quantitative Criminology*, 27(3):339–362, 2011.

[22] Matti Langel and Yves Tillé. Variance estimation of the gini index: revisiting a result several times published. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2):521–540, 2013.

[23] Elizabeth Macbeth and Barak Ariel. Place-based statistical versus clinical predictions of crime hot spots and harm locations in northern ireland. *Justice Quarterly*, pages 1–34, 2017.

[24] George O Mohler, Martin B Short, Sean Malinowski, Mark Johnson, George E Tita, Andrea L Bertozzi, and P Jeffrey Brantingham. Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110(512):1399–1411, 2015.

[25] GO Mohler, MB Short, and PJ Brantingham. The concentration dynamics tradeoff in crime hot spotting. *Unraveling the Crime-Place Connection*, 22:19–40, 2017.

[26] Federal Bureau of Investigation. Crime in the united states. uniform crime reports online. 2016.

[27] Marcos Oliveira, Carmelo Bastos-Filho, and Ronaldo Menezes. The scaling of crime concentration in cities. *PloS one*, 12(8):e0183110, 2017.

[28] Ali BZ Salem and Timothy D Mount. A convenient descriptive model of income distribution: the gamma density. *Econometrica: journal of the Econometric Society*, pages 1115–1127, 1974.

[29] Robert J Sampson and W Byron Groves. Community structure and crime: Testing social-disorganization theory. *American journal of sociology*, 94(4):774–802, 1989.

[30] Robert J Sampson, Stephen W Raudenbush, and Felton Earls. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328):918–924, 1997.

[31] Naratip Santitissadeekorn, MB Short, and DJB Lloyd. Sequential data assimilation for 1d self-exciting processes with application to urban crime data. *Computational Statistics & Data Analysis*, 128:163–183, 2018.

[32] P Schlattmann, J Hoehne, and M Verba. Caman: finite mixture models and meta-analysis tools based on ca man 2014: R package version 0.71, 2015.

[33] Cory Schnell, Anthony A Braga, and Eric L Piza. The influence of community areas, neighborhood clusters, and street segments on the spatial variability of violent crime in chicago. *Journal of quantitative criminology*, 33(3):469–496, 2017.

[34] Lawrence W Sherman, Patrick R Gartin, and Michael E Buerger. Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27(1):27–56, 1989.

[35] Lawrence W Sherman and David Weisburd. General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice quarterly*, 12(4):625–648, 1995.

[36] David Weisburd. The law of crime concentration and the criminology of place. *Criminology*, 53(2):133–157, 2015.

[37] David Weisburd. Hot spots of crime and place-based prevention. *Criminology & Public Policy*, 17(1):5–25, 2018.

[38] David Weisburd, Gerben JN Bruinsma, and Wim Bernasco. Units of analysis in geographic criminology: historical development, critical issues, and open questions. In *Putting crime in its place*, pages 3–31. Springer, 2009.

[39] David Weisburd, Shawn Bushway, Cynthia Lum, and Sue-Ming Yang. Trajectories of crime at places: A longitudinal study of street segments in the city of seattle. *Criminology*, 42(2):283–322, 2004.

[40] David Weisburd, Breanne Cave, Matthew Nelson, Clair White, Amelia Haviland, Justin Ready, Brian Lawton, and Kathleen Sikkema. Mean streets and mental health: Depression and post-traumatic stress disorder at crime hot spots. *American journal of community psychology*, 2018.

[41] David Weisburd and Lorraine Green. Policing drug hot spots: The jersey city drug market analysis experiment. *Justice Quarterly*, 12(4):711–735, 1995.

[42] David Weisburd, Elizabeth R Groff, and Sue-Ming Yang. *The criminology of place: Street segments and our understanding of the crime problem.* Oxford University Press, 2012.

[43] Clair White and David Weisburd. A co-responder model for policing mental health problems at crime hot spots: Findings from a pilot project. *Policing: a journal of policy and practice*, 12(2):194–209, 2017.

[44] Kuan Xu. How has the literature on gini's index evolved in the past 80 years? *Available at SSRN: https://ssrn.com/abstract=423200*, 2003.

[45] Jiancang Zhuang, Yosihiko Ogata, and David Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380, 2002.

8 Appendix: R Code for estimation of the law of crime concentration

```
# Simulation Rcode for Poisson-Gamma estimator of law of concentration L(p) and G
# inputs: counts (vector of counts in each cell)
#         Ntotal (number of cells)
#         Nflag (number of cells used as the top p fraction)

library(MASS)

pars=fitdistr(counts,"negative binomial")$estimate #estimate parameters mu, k

simulated_gam=rgamma(Ntotal,shape=pars[1],rate=1)

sorted_gam=sort(simulated_gam,decreasing=T)
concentration=sum(sorted_gam[1:Nflag])/sum(sorted_gam)

normalized_gam=sorted_gam/sum(sorted_gam)
gini=(1/Ntotal)*(2*sum(cumsum(normalized_gam))-Ntotal-1)
```

Figure 7: R code for Poisson-Gamma estimator via simulation.

```
# Integration Rcode for Poisson-Gamma estimator of law of concentration L(p) and G
# inputs: counts (vector of counts in each cell)
#         Ntotal (total number of cells)
#         Nflag (number of cells used as the top p fraction)

library(MASS)

pars=fitdistr(counts,"negative binomial")$estimate #estimate parameters mu, k

concentration=1-sum(qgamma(seq(.0001,1-Nflag/Ntotal,by=.0001),
        shape=pars[1],rate=1))*.0001/pars[1]

p=seq(.025,.975,by=.05)
Fp=numeric(length(p))
for(i in 1:length(p)){Fp[i]=qgamma(seq(.0001,1-p[i],by=.0001),
        shape=pars[1],rate=1))*.0001/pars[1]}
gini=2*sum(Fp)*.05-1
```

Figure 8: R code for Poisson-Gamma estimator via numerical integration.