Testing for causal clustering in point processes

Ian McGovern, P. Jeffrey Brantingham, and Frederic Schoenberg

University of California, Los Angeles, United States of America

November 8, 2023

Abstract

Discriminating causal clustering in point processes is of high interest for a variety of applications. We propose a simulation-based test evaluating the relative fit of Hawkes models, Neyman-Scott models, and inhomogeneous Poisson models, and compare with the time reversal test of Cordi et al. (2017). Under general conditions, causal clustering can be distinguished from inhomogeneity with high accuracy using these tests. The methods are applied to crime data on reported shootings in Boston from 2015-2021, where strong evidence of retaliatory triggering of events is seen in certain areas.

1. Introduction

Discriminating causal clustering from inhomogeneity is one of the key problems in point process analysis. Indeed, Diggle(2014) [4] describes this problem as one of the most important challenges currently facing point process analysts. Both causal clustering and inhomogeneity can lead to aggregation of points in certain locations, though the mechanisms for this aggregation have very different implications. Causal clustering, or triggering, refers to the situation where the occurrence of a point causes other points to be more likely to occur in the near vicinity of space-time. Inhomogeneity refers to the case where, because of differences in the background spatial-temporal environment, points are simply more likely to occur at certain locations of space-time than others. Discriminating between these two phenomena can be very difficult in practice.

For example, suppose one is analyzing catalogues of reported gang-related violent crimes, and many occurrences are present in one spatial-temporal area. Is this aggregation of points due to the socioeconomic and geographical circumstances in the particular location, in which case the explanation is inhomogeneity? Or is the aggregation due at least partly to retaliation, where one such crime that just happens to occur in the region may spark several retaliations, each of which might yield further retaliations, and so on, in which the explanation is triggering?

One approach suggested in Diggle(2014) [4] is to observe the point process repeatedly and inspect whether the clustering of points appears to occur predominantly in the same spatial-temporal locations, in which case inhomogeneity is the dominant paradigm. However, very often in practice, obtaining repeatedly observed point process data is not feasible. Furthermore, it is possible for a process, such as a Neyman-Scott process, to exhibit aggregation of points at certain random hot-spots, which may be different in each realization, yet the aggregation of points is still not causal in the sense of individual points triggering others as in a Hawkes process.

An approach taken in several papers including Park et al. (2020) [8] is to attempt to fit a Hawkes model with both inhomogeneity and triggering, and where the spatially and temporally varying background rate is modeled as accurately as possible, for instance using kernel smoothing of previously occurring points as well as covariate information that may influence the background rate, such as demographic information on each census tract. The idea is that if the inhomogeneity is accurately modeled via the background rate, then any triggering estimated in the resulting Hawkes model may be attributed to causal clustering. While this idea is sensible, it can be difficult to assess whether the background rate indeed accurately captures all the inhomogeneity in the spatial-temporal environment, and any inhomogeneity not adequately captured by the model will leak into the triggering function and be incorrectly characterized as causal clustering.

An idea explored in Cordi et al. (2017) [2] is to fit a Hawkes model to the data and another to the data with the times reversed. If the model fits significantly better to the forward time data, then this suggests the aggregation of points may indeed be causal, whereas if the model fits equally well with the times reversed, then the aggregation is most likely due to inhomogeneity. The idea is that typically in applications it would make no sense for points to trigger the occurrence of *prior* points, so the observation that a Hawkes model fits as well to the time-reversed data as it does to the forward-time data is incompatible with actual causal clustering but is consistent with the notion that the Hawkes model's background rate term is not accurately describing the inhomogeneity in the process and thus incorrectly classifying some of the inhomogeneity as triggering. This is a clever idea but it is not obvious how to extend it to the case where the inhomogeneity may vary over space instead of (or in addition to) time. Also, in some cases, it can happen that a process with truly causal clustering might result in a Hawkes model happening to fit well to the data with times reversed.

Previous research in causality in point processes has focused on different elements of causal inference compared to the causal ideas that this paper explores. Xu et al(2016) studied learning Granger causality within Hawkes processes [13]. This paper, however, studies ways in which to determine whether the events of a certain dimension(the result dimension) can be predicted by knowing the history of a different dimension(the cause dimension). This method, however, still suffers the same limitations of most Hawkes process analysis in that the "causality" that is more accurately described as correlations between different dimensions within the point process.

An alternative idea explored here is to fit both a Hawkes model with causal clustering as well as non-causal models as similar as possible to the Hawkes model but without causal clustering, such as Neyman-Scott and inhomogeneous Poisson models, and compare how these models fit. This allows one to quantify the degree of causal clustering in the data, and, provided the models are sensible and fit well, if the Hawkes model fits significantly better than the alternative models without causal clustering, then this is potentially strong evidence that the causal triggering identified by the Hawkes model is real.

We consider formal hypothesis tests using the information gain statistic to determine if a Hawkes model fits significantly better than Neyman-Scott or inhomogeneous Poisson alternatives, and a similar test is performed using the time-reversal method. Both the time reversal test and the model comparison tests are applied to simulated spatial-temporal point processes to quantify the accuracy in identifying causal clustering models. Following this, these methods are applied to reported shooting data from Boston. The analysis for the simulated data show that accuracy in correctly identifying causal structures can be very high under certain conditions. The application to the reported Boston shootings data suggest that there is indeed causal triggering in certain locations, perhaps due to retaliatory crime activity.

2. Background on inhomogeoeus Poisson, Hawkes, and Neyman-Scott processes

A spatial-temporal point process N on $S = \mathbb{R}^2 \times R$ is a random collection of points such that for any bounded Borel subset B of S, the number of points that are within B is some finite number, which is denoted as $N(B)$ [12]. A point process is *simple* if with probability one, all its points are distinct. Such point processes can be defined by their conditional intensity, $\lambda(x, y, t|\mathcal{H}_t)$,

$$\lambda(x, y, t|\mathcal{H}_t) = \lim_{\delta_x,\delta_y,\delta_t \to 0} \frac{\mathbb{E}[N((x, x+\delta_x) \times (y, y+\delta_y) \times (t, t+\delta_t)|\mathcal{H}_t)]}{\delta_x\delta_y\delta_t}$$

where $\mathcal{H}_t$ is defined as the history of the process up to time t.

If λ varies with x, y, and t but $\lambda(x, y, t)$ does not depend on what points have occurred previously, then N is an inhomogeneous Poisson process. Such processes embody the notion that aggregation of points is due to inhomogeneity only. In the context of crimes, an inhomogeneous Poisson model may allow the rate of points at any particular location and time to depend on the socio-economic features of the location in question, but would not incorporate retaliatory behavior in the model.

A Hawkes process is referred to as a "self-exciting" process, in that a point will trigger future points in its spatial vicinity. This type of model has often used to describe clustered phenomena such as earthquakes and infectious diseases (see e.g. Ogata 1988 [14], Meyer et al. 2012, Reinhart 2018 [1]). Parent points occur according to a background inhomogeneous Poisson process, $\mu(x, y, t)$. These parent points then produce offspring according to some triggering density h and some productivity value κ, the latter of which represents the expected number of points triggered by any given point. Once the parent points have produced offspring, those offspring trigger further offspring, and so on. The conditional intensity is thus given by

$$\lambda(x, y, t) = \mu(x, y, t) + \kappa \sum_{i:t_i<t} h(x - x_i, y - y_i, t - t_i).$$

A Neyman-Scott process is a clustering model defined in a two-part process. First, "parent" points are distributed throughout the spatial-temporal domain. Each parent point creates a random number with mean A of offspring points according to a specified triggering distribution, and the final process consists only of the offspring points. [11] Neyman-Scott processes have been used to describe clustered spatial processes such as tree stands (Penttinen et al. 1992), rainfall (Guttorp 1996), and galaxies (Snethlage et al. 2002), and typically the triggering density is symmetric so that offspring are distributed around their parents according to some isotropic density. Here, we consider the spatial-temporal context where the offspring points are distributed around their parents isotropically in space and time, meaning the parent points generate offspring occurring both before and after their parents. Thus the aggregation in such a Neyman-Scott process is causal but is not physically sensible for applications where a point cannot trigger prior points, and we will be using such models not for their physical plausibility but purely for purposes of comparison with Hawkes processes. The conditional intensity of a Neyman-Scott process is difficult to write in condensed form (Møller and Waagepetersen 2004, Zhang 2018), but can readily be estimated via maximum likelihood, minimum constrast, or other methods, despite occasional difficulties with convergence failure or numerical instability (Baddeley et al. 2022).

3. Methods

For a point process with conditional intensity $\lambda(x, y, t)$ and with points, denoted as $p_1 = (x_1, y_1, t_1), ..., p_n = (x_n, y_n, t_n)$, the likelihood can be expressed as

$$\prod_{1 \le i \le n} \lambda(p_i) \exp\left(- \int_S \lambda(x, y, t) dxdydt \right)$$

where S is the spatial-temporal observation region. Therefore, estimating the likelihood becomes a process of calculating the intensity at each observed point, then calculating the integral of the conditional intensity over the state space. Calculating the conditional intensity at each point is very simple, however the integral can often be too complicated to compute analytically. Approximation is typically necessary to calculate this integral.

For formal comparison of models, a hypothesis test method based on the expected information gain per trial is performed. The expected information gain per trial, described in Dayle(2016) [3], is a measure of the change in entropy scores from a null model and an alternate model. This information gain is a measure of the predictive performance of a model in terms of predicting the next occurring point within the point process, and is closely approximated by the mean log-likelihood ratio (Harte 2007 [5]),

$$\hat{G}_N = \frac{1}{N} \log\left(\frac{L_1}{L_0}\right)$$

where L_1 is the likelihood of the alternate model, L_0 is the likelihood for the null model, and N is the total number of points observed.

We consider a test with the following assumptions.

H_0: The data are generated from a Neyman-Scott process.

H_1: The data are generated from a Hawkes process.

Suppose the significance level $\alpha = .05$. By design, if the data truly arise from a Neyman-Scott model, then the test will reject the null hypothesis H_0 with probability 5%. The idea, however, is that such a formal test may also be useful when the data may arise from an inhomogeneous Poisson model, since in such cases the Hawkes model would not be expected to fit significantly better than the Neyman-Scott model and thus the test may be expected often to fail to reject the null hypothesis that the Neyman-Scott process is the generating mechanism. Using simulations, we consider the case where the data are generated via an inhomogeneous Poisson process and the above hypotheses are tested, as well as when the data are generated according to a Hawkes process.

4. Simulations

Hawkes processes are simulated in order to find the power of the test, i.e. the fraction of times the test correctly rejects the null hypothesis of a Neyman-Scott process in favor of the Hawkes model. The triggering density, $h()$ is spatially a two dimensional Gaussian distribution, and temporally a truncated Gaussian distribution with a lower bound of 0. The μ parameter is the mean number of first generation points in the spatiotemporal region and the σ parameter is the standard deviation for the Gaussian triggering density. The spatial region is $[0, 1] \times [0, 1]$ and the temporal region is $[0, 1]$.

For each simulation, three likelihood values were calculated: the likelihood for the standard Hawkes process model, the likelihood for a Neyman-Scott model, and then all the times were reversed and the likelihood was found for the "backwards" or "reversed" Hawkes process.

In order to obtain an approximate sampling distribution for the information gain statistic under the null hypothesis, the following Monte Carlo technique is used. First, a Hawkes process is simulated, and a Neyman-Scott model is fit to this Hawkes process via maximum likelihood. Then, using the parameters estimated for the Neyman-Scott process, 100 Neyman-Scott models are simulated. Then, the likelihoods for the forward Hawkes process, reversed Hawkes process, and Neyman-Scott process were calculated for each simulated Neyman-Scott process. Finally, the information gain statistic for both the Neyman-Scott comparison test and time reversal comparison test were calculated for each simulated Neyman-Scott process. These values were then used to create estimate the sampling distribution of the information gain statistic.

A hypothesis test was performed on simulated Hawkes processes using the information gain statistic. By design, this test will fail to reject the null hypothesis with probability 95% when the simulated data come from a Neyman-Scott model. When the simulated data come instead from an inhomogeneous Poisson model, the information gain test failed to reject the null hypothesis approximately 95% of the time as well. A variety of different Hawkes processes were simulated in order to investigate the power of the test. The three parameters that were altered were the μ parameter, the κ parameter, and the σ parameter. Each of these were tested over a range of values, and the other parameters were kept constant at $\mu = 18$, $\kappa = .81$, and $\sigma = .0002$. Both the value of σ in the temporal triggering distribution(denoted as σ_t) and the value of σ in the spatial triggering distribution(denoted as σ_{xy}) were allowed to vary. The range of tested values was chosen to be similar to the fitted Hawkes model for the application to crime data in Section 5 of this paper.

First, the κ values were altered in the range of $(.75, 9)$. The results can be seen in Figure 1. The

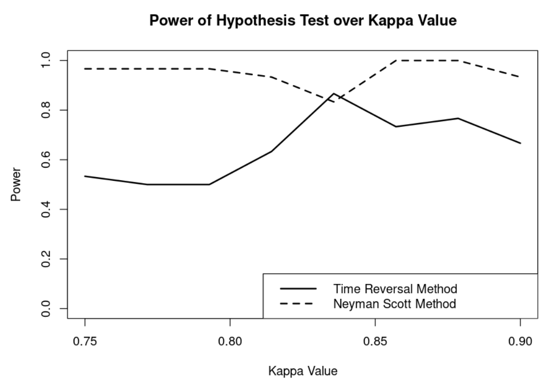

Figure 1

Neyman-Scott test has much higher power than the time reversal test for nearly all values of κ. The power for the time reversal method generally gets higher as the value of κ increases, although this pattern is not fully consistent.

When the σ_{xy} and σ_t parameters were allowed to vary on a logarithmic scale from .000005 to .1, the resulting power of the tests is shown in Figures 2 and 3.

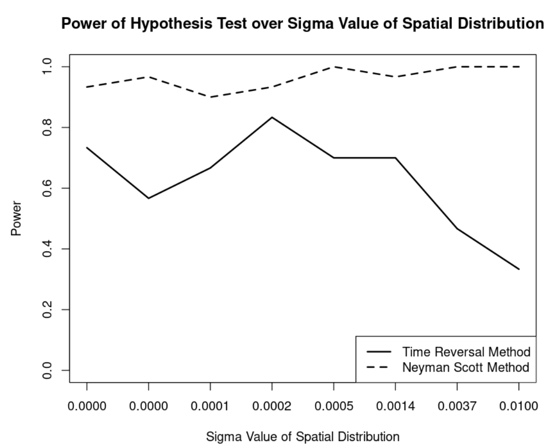

Figure 2

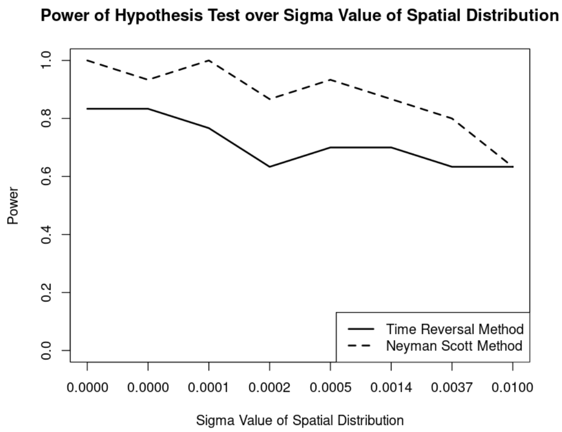

Figure 3

The Neyman-Scott test appears to have very high power for all values of σ_t, while the Neyman-Scott test has lower power for higher values of σ_{xy}. For the time reversal tests, increasing σ_t or σ_{xy} tends to reduce power, most likely because as the variance values increase, it is increasingly difficult to discern any meaningful clusters, and the points simply appear random within the space. This potentially makes distinguishing between different types of clustering more difficult.

Finally, the background rate parameter μ was permitted to vary from 10 to 30, and the results of these simulations are summarized in Figure 4. The power of the tests does not appear to change

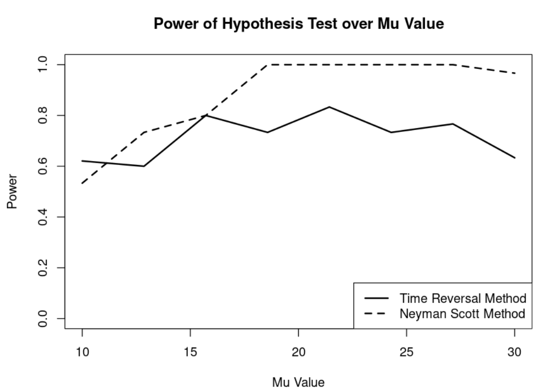

Figure 4

very significantly as μ varies. The two tests have nearly equal power over all values of μ, with the Neyman-Scott test having slightly higher power overall. Overall, the power of the Neyman-Scott test is 84.4% and the power of the reversal test is 74.4%.

5. Application to Crime Data

5.1. Data

Recorded data on 8,862 reported illegal shootings in Boston between 2015 and 2021 were collected from the public data source for the Boston government(https://data.boston.gov/dataset/shootings). Figure 5 shows a kernel smoothing of the locations of these reported crimes.

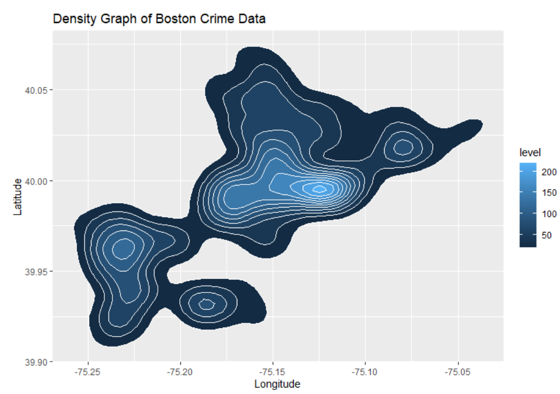

Figure 5

5.2. Methods

The data were divided into a 10×10 grid to analyze individual sections more clearly. Grid cells including less than 5 points were excluded from this analysis due to insufficient data. For each remaining grid cell, a Neyman-Scott model was fit by maximum likelihood estimation to the data within the cell, and then realizations of Neyman-Scott models were simulated repeatedly with parameters equal to these maximum likelihood estimates, to create a sampling distribution for the information gain statistic. For each simulation, the likelihood, L_1, for a Hawkes model, and the likelihood L_0, for a Neyman-Scott model, were calculated, and used to calculate $\hat{G}_n$. This creates a sampling distribution for the value of the information gain statistic, and the value of $\hat{G}_n$ for the actual data is then compared to this sampling distribution. If the value of $\hat{G}_n$ is above the 95% percentile for the simulated sampling distribution, then we say the test rejected the null hypothesis. For the time-reversal test this procedure was repeated, but with L_1 as the likelihood of the Hawkes model given the data and L_0 as the likelihood of the Hawkes model with the times reversed.

5.3. Results

Figure 6 shows the results of the Neyman-Scott hypothesis test. The results suggest that for the majority of locations in Boston, there is significant causal clustering present in the data on recorded shootings. Of the grid sections that were included within the analysis, 84.6% resulted in the test rejecting the null hypothesis, and these sections contained 83.7% of the total reported shootings. At the same time, there are several locations, especially on the Northwest borders of the dataset, where the test fails to reject the null hypothesis and suggests that the local aggregation of points in these locations may be entirely due to inhomogeneity.

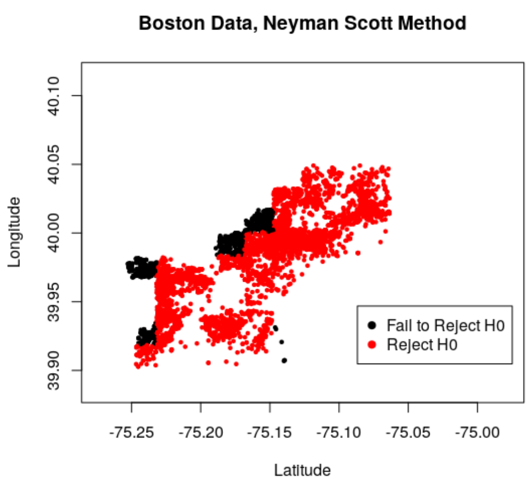

Figure 6

The time reversal test results, shown in Figure 7, shows far more grid cells where the test fails to reject the null hypothesis. The time reversal test only rejected the null hypothesis in 44.2% of the grid cells, corresponding to a total of 46.8% of the reported shootings. The majority of grid cells where the Neyman-Scott test failed to reject the null hypothesis also had the time reversal test fail to reject the null hypothesis, again suggesting inhomogeneity as the dominant cause of aggregation of points in these areas.

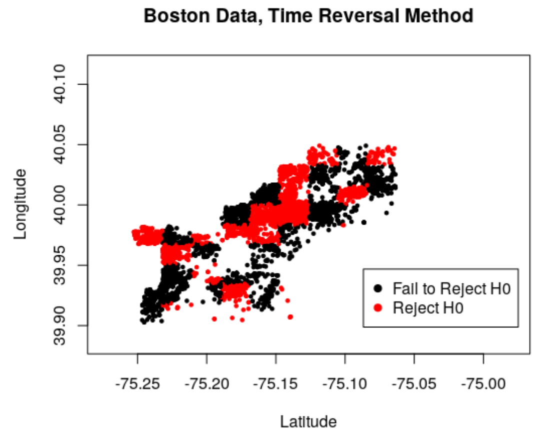

Figure 7

5.4. Analysis

The results of the Neyman-Scott test indicate that, in the vast majority of locations within Boston, the reported shooting data from 2015-2021 are significantly better fit by a Hawkes model with causal clustering than by a Neyman-Scott model. This suggests that the clustering in these points is truly causal, i.e. is not well explained by inhomogeneity alone, and may instead be at least partly explained by triggering, perhaps via retaliatory criminal behavior.

The estimated values of κ for the Hawkes models fit to each grid cell had a mean of 0.4, suggesting that according to the fitted model, about 40% of the reported shootings were triggered by prior shootings.

The results provide evidence that a Hawkes model with causal clustering may be appropriate for certain crime data. However, there are still some areas, especially near the Northwestern borders of the observation region, where causal clustering is not indicated. This could possibly be due to spatially varying covariates, differences in gang territory, or other factors resulting in more causal clustering in certain locations rather than others.

The Neyman-Scott and time reversal tests resulted in substantially different classifications. One possible explanation for this can be seen in the power analysis indicated by the simulations, since the Neyman-Scott test had higher power than the time reversal test in most cases. Therefore, the reason that so many more sections failed to reject the null hypothesis using the time reversal test could be because the power of this test was too low.

6. Conclusion

Distinguishing between causal clustering and inhomogeneity in point processes is still a problem requiring much further study. Simulations show that under certain conditions, a simulated Hawkes model can be correctly distinguished from a Neyman-Scott model using the information gain statistic, and furthermore, the test appears to have high power in distinguishing a Hawkes model from an inhomogeneous Poisson model as well. The time reversal test, by contrast, has somewhat lower power. This power is affected by the parameters of the simulation, with larger data sets and more intense clustering resulting in higher power for both tests.

Hawkes models have been used extensively in crime data analysis, typically without much investigation into whether or not the assumption of causal clustering is indicated. Models without causal clustering, such as inhomogeneous Poisson models or Neyman-Scott models, may fit just as well to the data in some situations. However, with regard to the application to the recorded shooting data in Boston, our results do suggest strong evidence of causal clustering in most areas of the city.

Future research should investigate this evidence of causal clustering further. Here, we considered Gaussian triggering functions for both the Neyman-Scott and Hawkes model, but alternative triggering functions could be considered. In addition, we allowed each spatial grid cell to have its own background rate, perhaps due to spatially varying covariates such as poverty levels or education levels. Future work could alternatively model the background crime rate more more explicitly as a function of such socio-economic covariates, as in Park et al. (2021). In addition, other types of reported crime data should be analyzed and the relationship between different types of crimes and the strength of evidence of causal clustering should be studied.

References

[1] Reinhart A. A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, 33(3):299–318, 2018.

[2] Marcus Cordi, Damien Challet, and Ioane-Muni Toke. Testing the causality of hawkes processes with time reversal. *Statistical Science*, 2017.

[3] D. J. Daley and D. Vere-Jones. Scoring probability forecasts for point processes: the entropy score and information gain. *Journal of Applied Probability*, 2016.

[4] Peter J. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. CRC Press, 2014.

[5] David Harte and David Vere-Jones. The entropy score and its uses in earthquake forecasting. *Pure Applied Geophysics*, 162, 2005.

[6] Alan Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society*, 33(3):438–443, 1971.

[7] Alan Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[8] Andrea L. Bertozzi Junhyung Park, Frederic Paik Schoenberg and P. Jeffrey Brantingham. Investigating clustering and violence interruption in gang-related violent crime data using spatial–temporal point processes with covariates. *Journal of the American Statistical Association*, 2021.

[9] Y. Kagan and L. Knopoff. Earthquake risk prediction as a stochastic process. *Physics of the Earth and PLanetary Interiors*, 14, 1977.

[10] Jerzy Neyman. On a new class of contagious distributions, applicable in entomology and bacteriology. *The Annals of Mathematical Statistics*, 10(1), 1939.

[11] Jerzy Neyman and Elizabeth L. Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society*, 20(1):1–43, 1958.

[12] M.N.M van Lieshout. *Theory of Spatial Statistics: A Concise Introduction*. CRC Press, 2019.

[13] Hongteng Xu, Mehrdad Farajtabar, and Hongyuan Zha. Learning granger causality for hawkes processes, 2016.

[14] Ogata Y. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[15] Jiancang Zhuang. Likelihood-based detection of cluster centers for neyman–scott point processes. *Journal of Environmental Statistics*, 8(3), 2018.