IMPERIAL

Spatio-Temporal Analysis of Drug Crime Nearby, Using Kernel Density Estimation

Haochen Gu

Introduction

Drug crime is a social problem all over the world. In this project, we will analyze the spatio-temporal point process of drug crime near Imperial. Where intensity function is estimated by the kernel density estimation [1].

Spatio-Temporal Point Process and Intensity Function

Definition 2.1: Spatio-Temporal Point Process [7]

Spatio-temporal point process (STPP) is a stochastic collection of points, where each point denotes an event $x = (t, \boldsymbol{s})$ associated with time t and location $\boldsymbol{s}$.

In this poster, $\boldsymbol{s} = (x, y)$ is a two-dimensional vector representing event's location, where x, y are Longitude and Latitude. And t is the twelve months in 2023.

Definition 2.2: Intensity Function [2]

Intensity is the expected number of points per unit area per unit time. When the intensity function $\lambda(\cdot)$ exists, the expected number of points in a region $S \times T$ is:

$$E[N(S \times T)] = \int_{S \times T} \lambda(\boldsymbol{s}, t) d\boldsymbol{s} dt \tag{1}$$

where $N(X)$ denotes the number of points in region X.

For different kinds of point processes, the intensity function is defined in different ways. Here's some examples:

- Homogeneous Poisson Process: $\lambda(\boldsymbol{s}, t) = \lambda$ where λ is a constant.
- Renewal Process : $\lambda = u(t - \tau_{<t->})$ if the waiting time distribution has hazard function u. [4]

Theorem 2.1: Separability of Intensity Funtion [6]

Assuming the independence between when and where the drug crime happened, the intensity function has the following property:

$$\lambda(\boldsymbol{s}, t) = \lambda(\boldsymbol{s})\lambda(t) \tag{2}$$

Generally speaking, the intensity fully characterizes the spatio-temporal distribution of the point process. In this project, we will introduce a method called Kernal Desnity Estimation to estimate the intensity function of the crime data, which is a common method to estimate the intensity function where no assumption of the underlying point process is made.

Kernal Density Estimation and its relationship with Intensity Function

Definition 3.1: Kernal Density Estimation [8]

Let the series $\{x_1, x_2, \ldots, x_n\}$ be an independent and identically distributed sample of n observations taken from a population X with an unknown probability distribution function $f(x)$. The probability density function(PDF) is estimated by:

$$\hat{f}(t) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - x_i}{h}) \tag{3}$$

where h is a parameter to be determined called *bandwidth*, and K is a *kernel function*. The kernel function is usually a symmetry probability distribution function, such as Gaussian, Normal, etc.

3.1 How KDE works

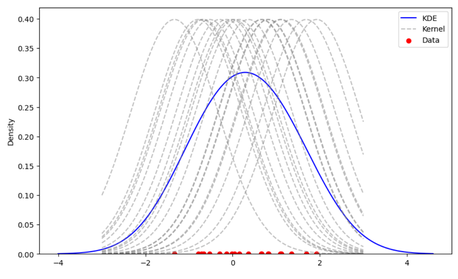

Figure 1: KDE with high bandwidth

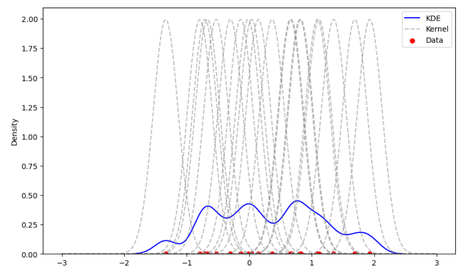

Figure 2: KDE with low bandwidth

Take the Figure 1 as an example. Dotted lines are the kernel function centered at each points, and the solid line is the estimated density function. By adding all the kernel functions together and divide it by nh, which make sure the area under the estimated density function is 1, we get the estimated density function. Such that there are two parameters determining the shape and performance of the estimated density function: the kernel function and the bandwidth.
In this poster, we don't talk to much about the kernel function, choosing it to be Gaussian. In this case, the bandwidth is the standard deviation of Gaussian kernel. Figure 1 and Figure 2 shows the KDE with high and low bandwidth respectively. Figure 2 with bandwidth of 0.2 has a lot of noise and we can call it *under-smoothed*. Figure 1 with bandwidth of 1.0 is *over-smoothed* and has higher bias.

3.2 Relationship with Intensity Function

KDE estimates the density, not intensity. However, there's a proportional relationship between them and it's easy to transform.
At a point $\boldsymbol{s}$, the value of PDF defines how likely an event occurs at $\boldsymbol{s}$. The intensity function $\lambda(\boldsymbol{s})$ defines the expected number of events in $\lambda(\boldsymbol{s})$ per unit area per unit time. As a result [5]:

$$\lambda(\boldsymbol{s}) = f(\boldsymbol{s}) \cdot E[N(S \times T)] \tag{4}$$

Application and Results

In the following section, I'm going to model the drug crime data in 2023, and show the results by explaining the intensity function.

4.1 Processing Data

- By Theo 2.1, the intensity function is estimated by KDE on time and place separately. The results are also analyzed separately since properties of $\lambda(\boldsymbol{s}, t)$ are fully captured by $\lambda(\boldsymbol{s})$ and $\lambda(t)$.
- $E[N(S \times T)]$ in Eq 4 is taken to be the sum of the data.
- Both bandwidths are chosen by cross-validation. The optimal bandwidths are 0.04 for place and 3.8 for time.
- Data source was from data.police.uk. The data were 12 months in 2023, and the region is a circle with a radius of 5 kilometers around Imperial College London.
- KDE and Cross-validation are implemented by cuml, a package supporting GPU acceleration.

4.2 Results

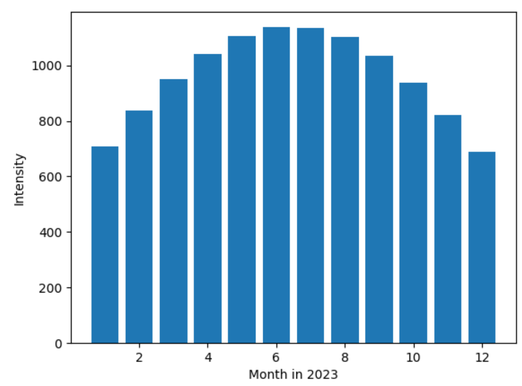

Figure 3: Intensity on time.

Fig 3 shows the intensity function of drug crime on time. The intensity is centered around the middle year, showing that the drug crime is more likely to happen in the middle of the year. However, this estimation has two potential worries, which are over-smoothed and edge-effect. The bandwidth chosen here is relatively large, which may lead to over-smoothed. Also, the intensity tends to be lower at the beginning and end of the year, which may be caused by "edge-effect" [3]. It means that near the region's boundary, there are fewer data points contributing to the estimation of intensity than the center of the region. Large bandwidths increase the influence among data points, intensifying the edge-effect.

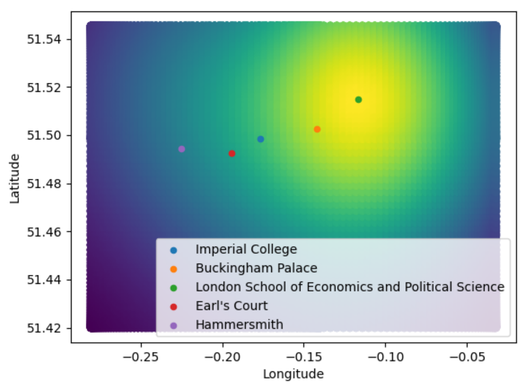

Figure 4: Intensity on place.

Fig 4 shows the intensity function of drug crime on place and location of some important buildings in each region are marked. It's so dramatical that, the intensity is centered at the London School of Economics and Political Science, which is the most likely place for drug crime to happen. Imperial College London isn't in the "hottest" place but still close to the yellow region. The closer to the southwest corner, the lower the intensity. For example, Hammersmith has relatively lower intensity of drug crime in this graph.

Conclusion and Limitaion

5.1 Conclusion

Based on the result of analysis, it's concluded that the drug crime doesn't occur uniformly in both space and time. It's more likely to occur in the northeast of Imperial. And middle of the year tends to have more drug crime.

5.2 Evaluation

In this project, due to the limited time available and in order to focus more on point process, I simply choose bandwidth by cross-validation, resulting in bandwidth of 3.8 for time. Mixing with "edge-effect", such a large bandwidth makes Fig 3 over-smoothed and unreliable. In the future, I will use other methods to choose bandwidth, such as Silverman's rule of thumb, which may give a more reliable result. In comparison, the conclusion on place is more reliable.
Moreover, if I get more time, I will do more elementary analysis on the raw data, hoping of finding more characteristics of the data, which will make it possible to model it using some specific type of point process, such as clustering point process.

References

[1] WTP Sarojinie Fernando and Martin L Hazelton. Generalizing the spatial relative risk function. *Spatial and spatio-temporal epidemiology*, 8:1–10, 2014.

[2] Jonatan A González and Paula Moraga. An adaptive kernel estimator for the intensity function of spatio-temporal point processes. *arXiv preprint arXiv:2208.12026*, 2022.

[3] Mark S. Handcock. Lecture 7: Problems and solutions for density estimation. Accessed: 2024-06-13.

[4] Martin Jacobsen and Joseph Gani. Point process theory and applications: marked point and piecewise deterministic processes. 2006.

[5] Paula Moraga. *Spatial Statistics for Data Science: Theory and Practice with R*. CRC Press, 2023.

[6] Dennis Sun. Lecture 10: Spatio-temporal point processes, 2014. Accessed: 2024-06-09.

[7] Yuan Yuan, Jingtao Ding, Chenyang Shao, Depeng Jin, and Yong Li. Spatio-temporal diffusion point processes. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3173–3184, 2023.

[8] Adriano Z Zambom and Ronaldo Dias. A review of kernel density estimation with applications to econometrics. *International Econometric Review*, 5(1):20–42, 2013.