Spatial Analysis of School Shootings

J Steven Raquel

3/10/2022

Background

The prevalence of gun violence in schools in the United States has been referred to both as an "epidemic" and a public health crisis. The debate on how to curb these tragedies is a solidly partisan issue, with calls for more and stronger gun control laws, as well as more mental health care, sometimes as an alternative to gun control. While school shootings have taken place in the US for decades, incidences (and gun control debates) have spiked since the 2000s, notably after the Columbine High school shooting in Littleton, Colorado that took place in 1999. This analysis attempts to trace the prevalence of school shootings in the US over the years, by looking at the spatial correlation of the incidents as well as modeling it as a Poisson point process, in order to ascertain whether the locations and events occur with complete spatial randomness. We also utilized gun control data to see whether the existence or later implementation of these laws had any effect on later incidents.

Data

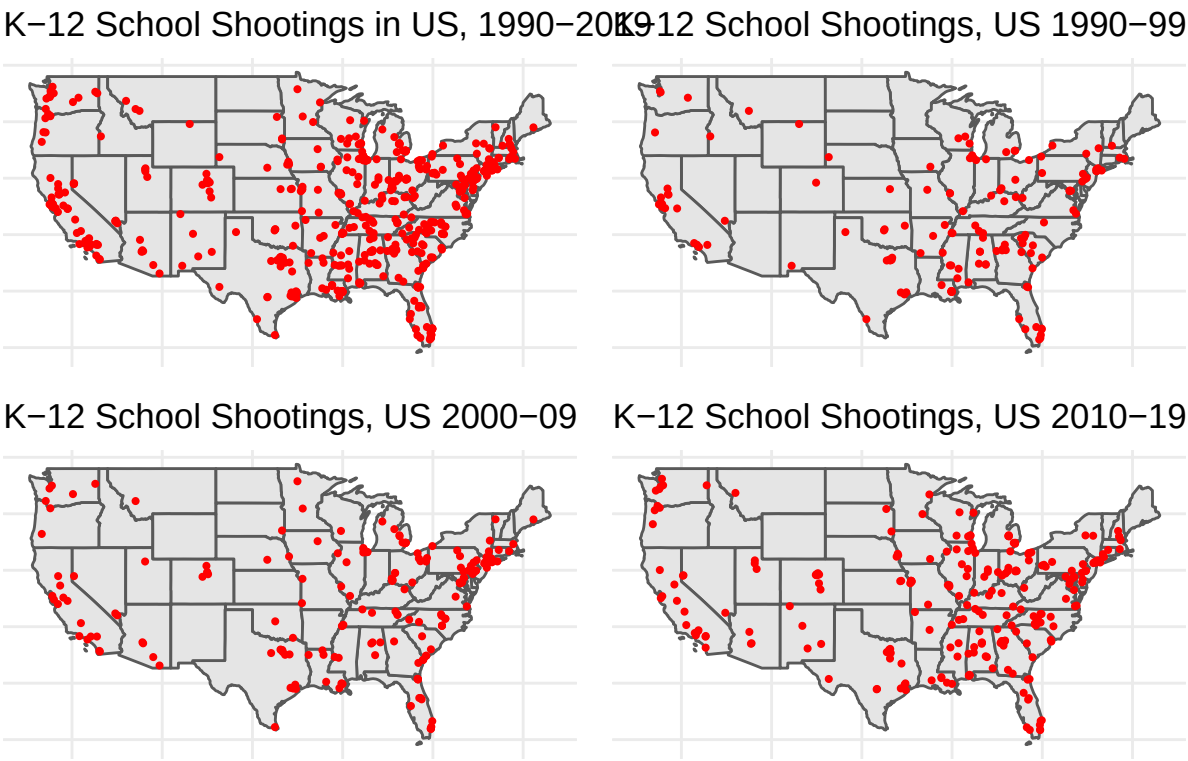
The school shooting data was sourced directly from the "K-12 School Shooting Database" made available by the Center for Homeland Defense and Security (CHDS). The information that comprises the dataset was determined by a specific process which entailed asking what exactly comprises a school shooting, e.g. whether the encounter happened on school property itself or within a classroom, whether or not the perpetrator was involved with the school in any way, e.g. a gang-related shooting between unaffiliated individuals that occurs after hours on school property, or an accidental discharge of a firearm that results in personal injury, would not be considered school shootings in this context. While incidents of this nature occur within the data, it was a deliberate decision on our part to filter out these incidents so as to hone in one the more specific and cultural recognized definition of a school shooting, which is an attack that takes place on school grounds, in order to target either students or faculty with the intent to cause terror and/or inflict harm on specific individuals. The data goes as far back as 1970 all the way to present day (March 2022 at the time of this writing), but it was of particular interest to look at the data after 1990 since this gives us a window into both the pre- and post-Columbine eras. The data originally only contained the name of the school, as well as the city and state where the events took place, but we were able to use the Google Maps API to source the specific latitude/longitude coordinates of these events, effectively giving us point reference data to model as a point process.

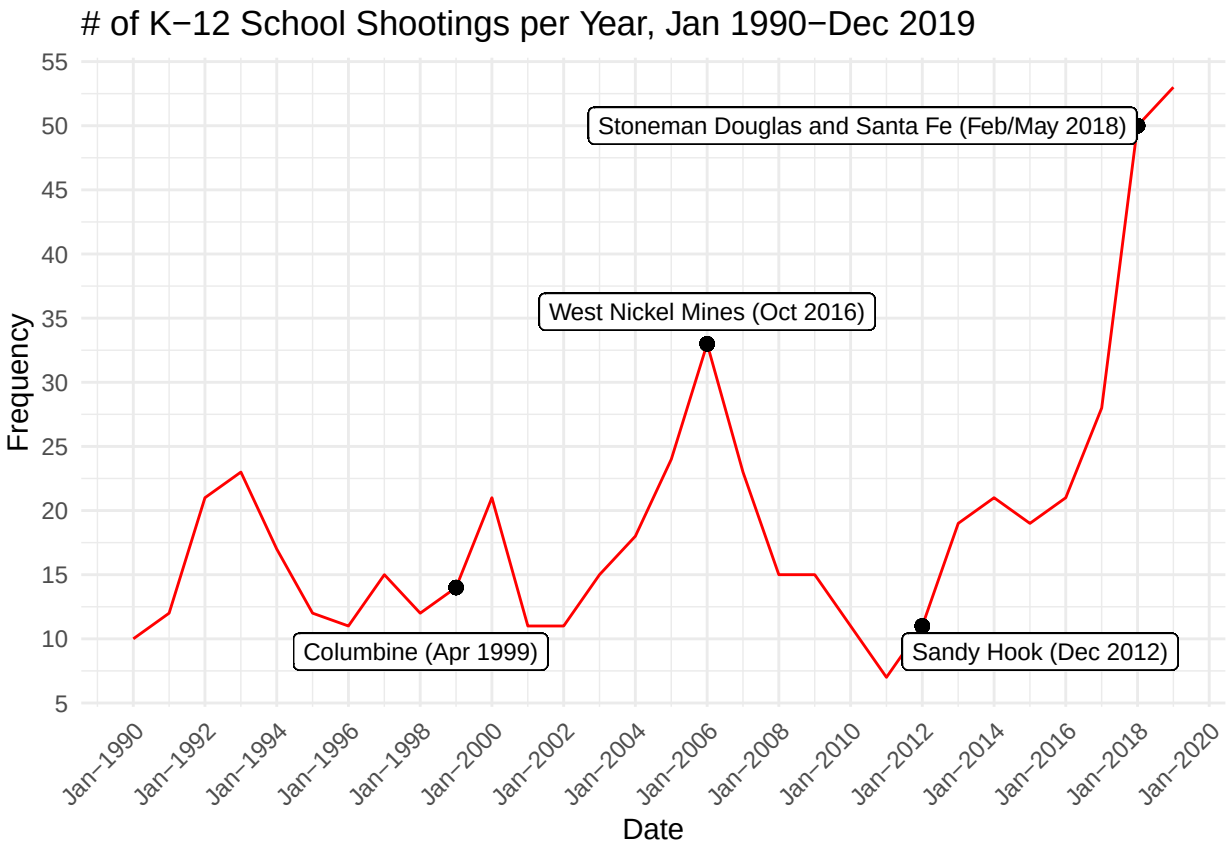
While the Columbine shooting in 1999 and the Virginia Tech shooting in 2007 (which is not recorded in this dataset on account of having taken place at a university) both recorded very high numbers of casualties and made major headlines, few other incidents in their era seemed to replicate this, despite fears at time that the incidents would inspire copycat tragedies at a similar scale. Conversely, half of the 10 deadliest shootings since 1999 all took place in the 2010s, among them the Sandy Hook and Stoneman Douglas shootings.

Methods

Exploratory Data Analysis

Although we have data dating as far back as 1970, we thought it more prudent to look specifically at the specific time period between the 90s all the way to the end of the 2010s, as a way to "book-end" the different eras of gun violence, in terms of the scale and impact.





Point Pattern Analysis

As with any spatial point pattern analysis, we are concerned with the following three questions, 1) whether the points are located at random, 2) whether they are clustered, and 3) whether they are placed regularly. The hypothesis of *complete spatial randomness* (CSR) asserts the following:

- The number of events in any region S with area $|S|$ follows a Poisson distribution with mean $\lambda|S|$, where λ is the intensity, i.e. λ does not change over S
- Given n events in S, the points s_i are independently located according to the uniform distribution on S, i.e. there is no interaction amongst events.

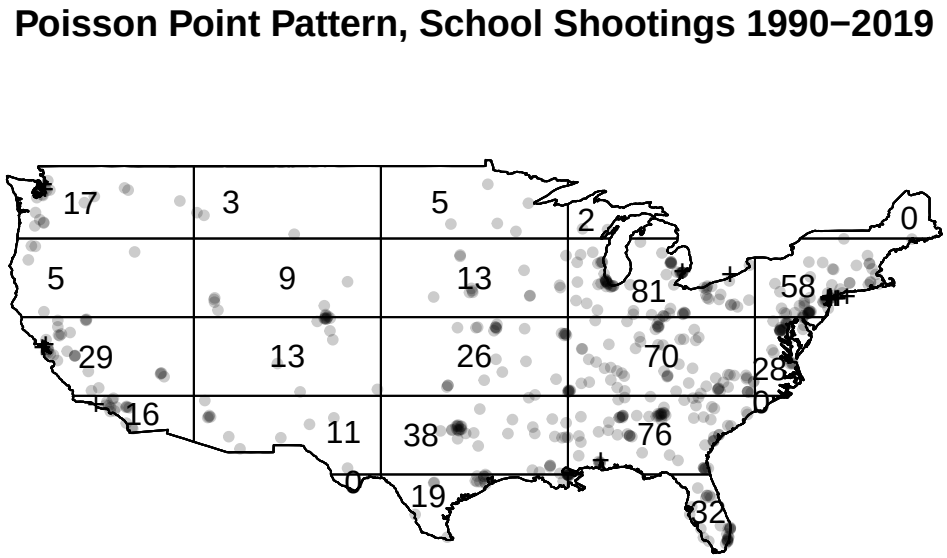
The intensity function $\lambda(s)$, also known as the first-order property of the spatial point process, is defined as

$$\lambda(s) = \lim_{|\Delta s| \to 0} \frac{E[N(\Delta s)]}{|\Delta s|}$$

Firstly we want to ascertain whether the incidences of school shootings are indeed a Poisson process, and if so, determine whether or not the process is *homogeneous* (where the intensity function $\lambda(s)$ assumes a constant λ) or *inhomogeneous*. Contextually, this means we are interested in ascertaining whether the spatial pattern of these shootings is random or not, i.e. are they more likely to take place in certain places, or around each other? The intensity function is the expected number of events per unit area.

Quadrats

While we could estimate the intensity function across the entire area, what we are interested in this particular analysis is to how the intensity varies across different regions contained therein. We do this by splitting up the area into what are referred to as *quadrats*.



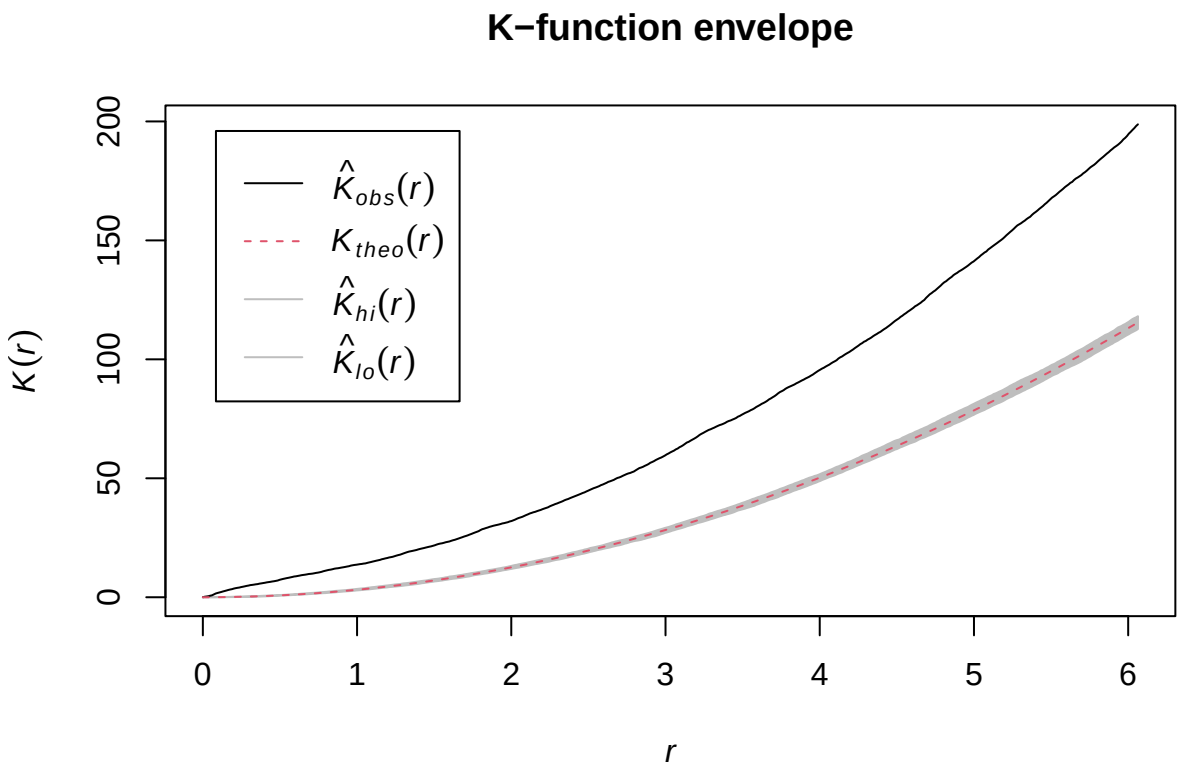

```
##
##  Conditional Monte Carlo test of CSR using quadrat counts
##  Test statistic: Pearson X2 statistic
##
## data:  ppp_9019
## X2 = 582.1, p-value = 2e-04
## alternative hypothesis: clustered
##
## Quadrats: 23 tiles (irregular windows)
```

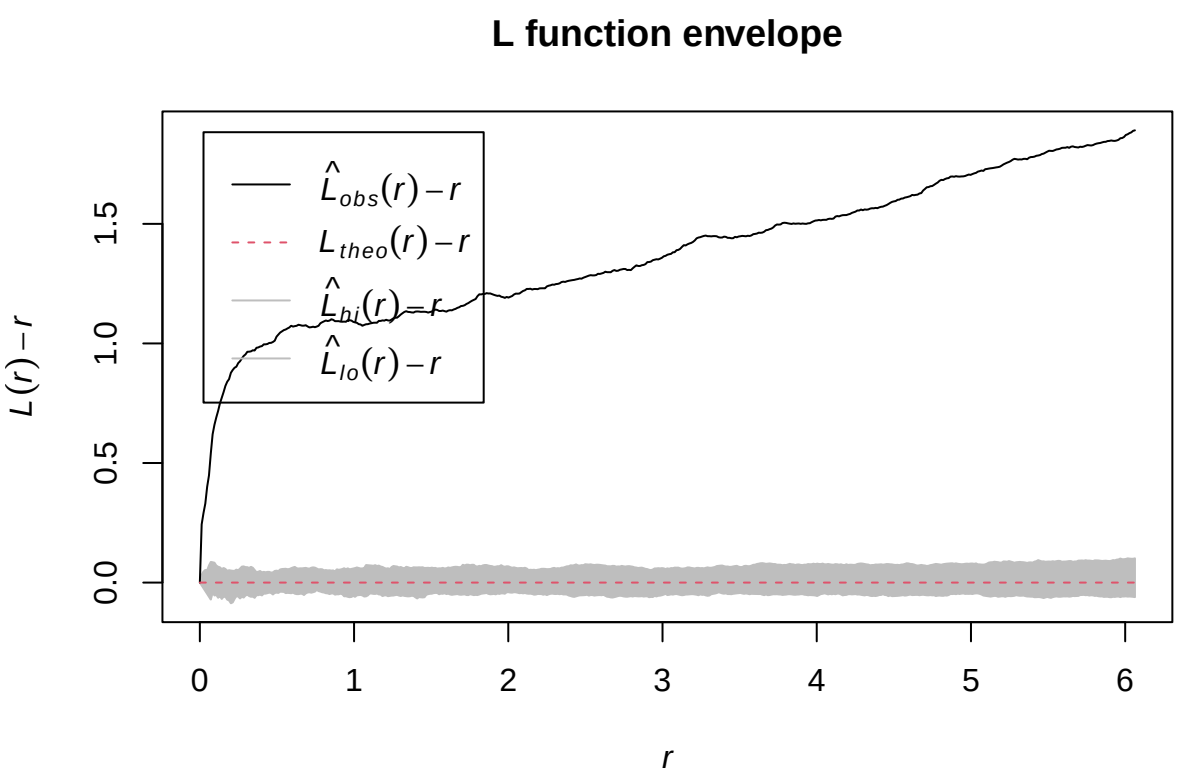
Discussion

References

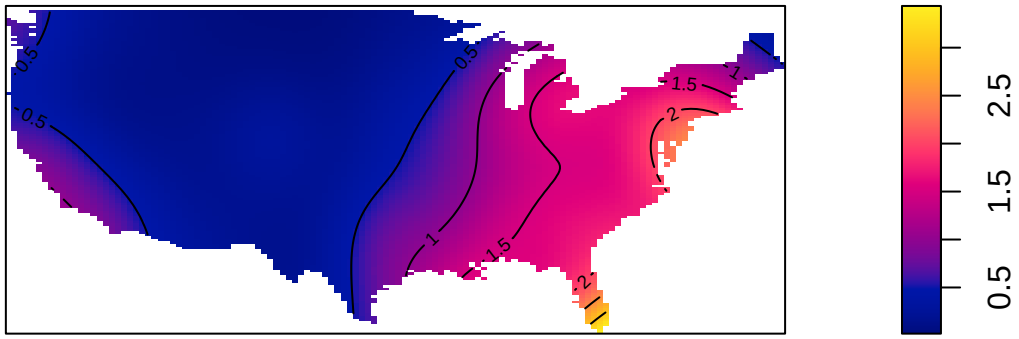
```
## Generating 99 simulations of CSR  ...
## 1, 2,  [etd 9:12] 3,  [etd 9:10] 4,
##  [etd 9:14] 5,  [etd 8:55] 6,  [etd 8:34] 7,  [etd 8:30] 8,
##  [etd 8:20] 9,  [etd 8:14] 10,  [etd 8:09] 11,  [etd 8:05] 12,
```

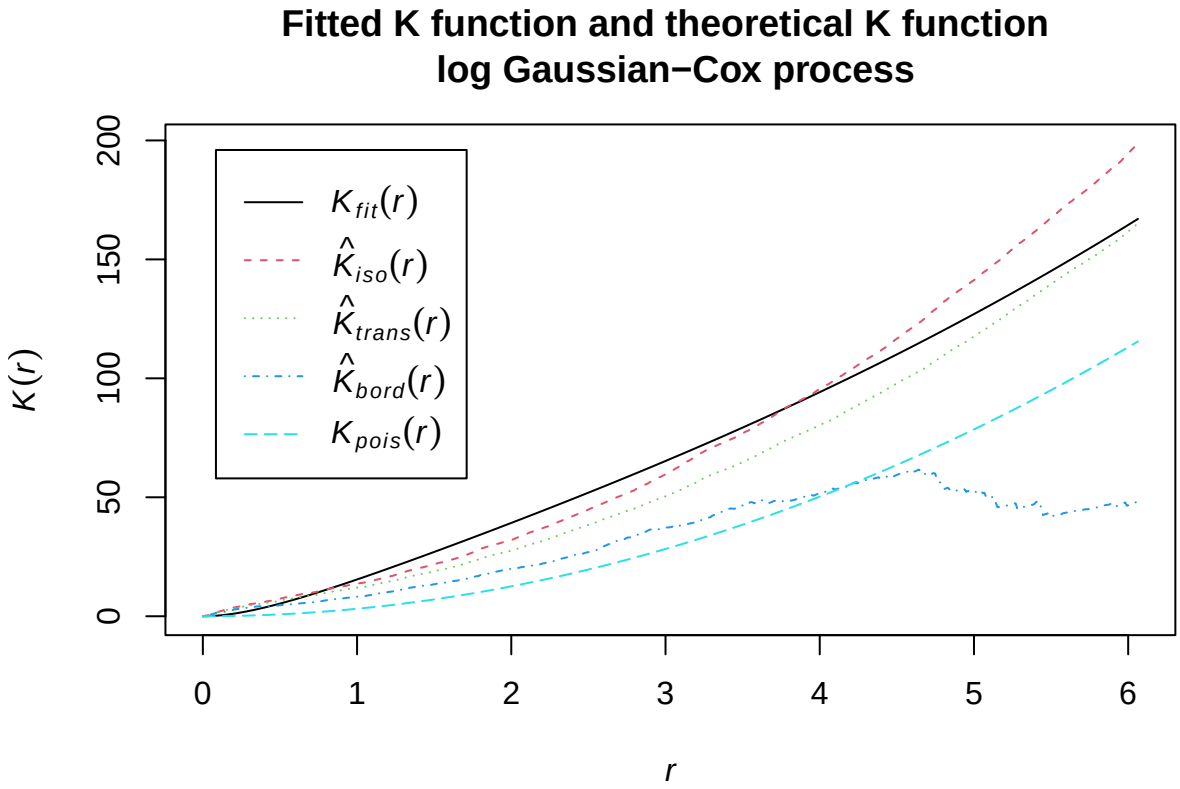
```
##  [etd 7:52] 13,  [etd 7:44] 14,  [etd 7:44] 15,  [etd 7:35] 16,
##  [etd 7:31] 17,  [etd 7:34] 18,  [etd 7:31] 19,  [etd 7:28] 20,
##  [etd 7:21] 21,  [etd 7:16] 22,  [etd 7:17] 23,  [etd 7:12] 24,
##  [etd 7:04] 25,  [etd 7:01] 26,  [etd 6:54] 27,  [etd 6:50] 28,
##  [etd 6:45] 29,  [etd 6:36] 30,  [etd 6:30] 31,  [etd 6:26] 32,
##  [etd 6:22] 33,  [etd 6:16] 34,  [etd 6:12] 35,  [etd 6:06] 36,
##  [etd 6:02] 37,  [etd 5:59] 38,  [etd 5:55] 39,  [etd 5:51] 40,
##  [etd 5:46] 41,  [etd 5:41] 42,  [etd 5:35] 43,  [etd 5:31] 44,
##  [etd 5:26] 45,  [etd 5:21] 46,  [etd 5:16] 47,  [etd 5:09] 48,
##  [etd 5:03] 49,  [etd 4:58] 50,  [etd 4:52] 51,  [etd 4:45] 52,
##  [etd 4:39] 53,  [etd 4:33] 54,  [etd 4:28] 55,  [etd 4:22] 56,
##  [etd 4:16] 57,  [etd 4:10] 58,  [etd 4:04] 59,  [etd 3:59] 60,
##  [etd 3:54] 61,  [etd 3:47] 62,  [etd 3:42] 63,  [etd 3:36] 64,
##  [etd 3:30] 65,  [etd 3:23] 66,  [etd 3:17] 67,  [etd 3:11] 68,
##  [etd 3:05] 69,  [etd 2:59] 70,  [etd 2:53] 71,  [etd 2:47] 72,
##  [etd 2:41] 73,  [etd 2:35] 74,  [etd 2:29] 75,  [etd 2:23] 76,
##  [etd 2:18] 77,  [etd 2:12] 78,  [etd 2:06] 79,  [etd 2:00] 80,
##  [etd 1:54] 81,  [etd 1:48] 82,  [etd 1:42] 83,  [etd 1:36] 84,
##  [etd 1:30] 85,  [etd 1:24] 86,  [etd 1:18] 87,  [etd 1:12] 88,
##  [etd 1:06] 89,  [etd 1:00] 90,  [etd 54 sec] 91,  [etd 48 sec] 92,
##  [etd 42 sec] 93,  [etd 36 sec] 94,  [etd 30 sec] 95,  [etd 24 sec] 96,
##  [etd 18 sec] 97,  [etd 12 sec] 98,  [etd 6 sec]  99.
##
## Done.

## Generating 99 simulations of CSR  ...
## 1, 2,  [etd 10:42] 3,  [etd 9:21] 4,
##  [etd 9:27] 5,  [etd 9:02] 6,  [etd 9:18] 7,  [etd 9:34] 8,
##  [etd 9:26] 9,  [etd 9:12] 10,  [etd 9:12] 11,  [etd 9:14] 12,
##  [etd 9:07] 13,  [etd 8:55] 14,  [etd 8:49] 15,  [etd 8:43] 16,
##  [etd 8:35] 17,  [etd 8:31] 18,  [etd 8:23] 19,  [etd 8:15] 20,
##  [etd 8:08] 21,  [etd 8:02] 22,  [etd 7:57] 23,  [etd 7:50] 24,
##  [etd 7:48] 25,  [etd 7:40] 26,  [etd 7:37] 27,  [etd 7:29] 28,
##  [etd 7:20] 29,  [etd 7:12] 30,  [etd 7:07] 31,  [etd 7:00] 32,
##  [etd 6:54] 33,  [etd 6:47] 34,  [etd 6:41] 35,  [etd 6:36] 36,
##  [etd 6:29] 37,  [etd 6:22] 38,  [etd 6:15] 39,  [etd 6:08] 40,
##  [etd 6:03] 41,  [etd 5:58] 42,  [etd 5:55] 43,  [etd 5:48] 44,
##  [etd 5:43] 45,  [etd 5:36] 46,  [etd 5:31] 47,  [etd 5:24] 48,
##  [etd 5:17] 49,  [etd 5:11] 50,  [etd 5:05] 51,  [etd 4:58] 52,
##  [etd 4:52] 53,  [etd 4:46] 54,  [etd 4:39] 55,  [etd 4:32] 56,
##  [etd 4:27] 57,  [etd 4:20] 58,  [etd 4:14] 59,  [etd 4:08] 60,
##  [etd 4:01] 61,  [etd 3:55] 62,  [etd 3:49] 63,  [etd 3:42] 64,
##  [etd 3:36] 65,  [etd 3:31] 66,  [etd 3:24] 67,  [etd 3:18] 68,
##  [etd 3:12] 69,  [etd 3:05] 70,  [etd 2:59] 71,  [etd 2:53] 72,
##  [etd 2:47] 73,  [etd 2:41] 74,  [etd 2:35] 75,  [etd 2:28] 76,
##  [etd 2:22] 77,  [etd 2:16] 78,  [etd 2:10] 79,  [etd 2:04] 80,
##  [etd 1:58] 81,  [etd 1:52] 82,  [etd 1:46] 83,  [etd 1:39] 84,
##  [etd 1:33] 85,  [etd 1:27] 86,  [etd 1:20] 87,  [etd 1:14] 88,
##  [etd 1:08] 89,  [etd 1:02] 90,  [etd 56 sec] 91,  [etd 50 sec] 92,
##  [etd 43 sec] 93,  [etd 37 sec] 94,  [etd 31 sec] 95,  [etd 25 sec] 96,
##  [etd 19 sec] 97,  [etd 12 sec] 98,  [etd 6 sec]  99.
##
## Done.
```


K−function envelope
$\hat{K}_{obs}(r)$
$K_{theo}(r)$
$\hat{K}_{hi}(r)$
$\hat{K}_{lo}(r)$
K(r)
0
50
100
150
200
0
1
2
3
4
5
6
r


L function envelope
$\hat{L}_{obs}(r) - r$
$L_{theo}(r) - r$
$\hat{L}_{hi}(r) - r$
$\hat{L}_{lo}(r) - r$
$L(r) - r$
0.0
0.5
1.0
1.5
0
1
2
3
4
5
6
r

Density of School Shootings, 1990−2019




Fitted K function and theoretical K function
log Gaussian−Cox process
K(r)
0
50
100
150
200
0
1
2
3
4
5
6
r
$K_{fit}(r)$
$\hat{K}_{iso}(r)$
$\hat{K}_{trans}(r)$
$\hat{K}_{bord}(r)$
$K_{pois}(r)$

References

Appendix