
UM-SJTU JOINT INSTITUTE

Probabilistic Methods in Engineering

(VE401)

Term Project 2

Police Shootings in the United States

Name: Feitong Tang ID: 518370910017 Group 1
Name: Weikai Zhou ID: 518021911039 Group 1

Date: 30 April 2020

1 Abstract

Based on David Spiegelhalter and Arthur Barnett's *London murders: a predictable pattern?*, this project analyzes the pattern of occurrence of fatal police shootings in the United States. The project gets its data from the database of the Washington Post and gives an overview of the data. It uses goodness-of-fit test to check whether the occurrence of fatal police shooting in the United States follows a Poisson distribution in the years 2015-2019. It also tests whether the average number of police shootings depends on weekdays. It gives the confidence interval for the parameter k of a Poisson distribution. Based on the data from Jan. 1 to Apr. 15, 2020, it tests whether it follows a Poisson distribution and calculates \hat{k}_{2020} . It gives the 95% prediction intervals for the number of police shootings in 2020 and analyzes how COVID-19 will influence the data for 2020.

Keywords: Fatal police shooting in the United States, Poisson distribution, goodness-of-fit test, confidence interval, prediction interval.

Contents

2 Introduction

2.1 Background

David Spiegelhalter and Arthur Barnett analyzed the pattern of London murders between April 2004 and March 2007, and predicted the number of murders in London during 2008 [1]. Based on their analysis and the data of fatal police shootings in the USA provided by Washington Post [2], we want to analyze the pattern of fatal police shootings in the USA from January 2015, and predict the number of fatal police shootings in 2020.

2.2 Objectives

- Give an overview of the data of fatal police shootings;
- Figure out the pattern of fatal police shootings and its dependence on weekday;
- Calculate the confidence interval for the parameter of a Poisson distribution;
- Predict the numbers of fatal police shootings in 2020;

3 Data analysis

3.1 Data source

The data we use is based on the database of Washington Post. It records every fatal shooting “in which a police officer, in the line of duty, shoots and kills a civilian” since Jan. 1, 2015. This sentence gives the clear definition of “fatal police shooting”, meaning that the database of Washington Post will not document the deaths of those in custody, fatal shootings by officers who are not on duty and non-shooting deaths. Moreover, the Washington Post gets their information mainly from “news accounts, social media postings and police reports”. It also monitors other database like Killed by Police and Fatal Encounters and updates the database regularly with new fatal police shootings and new facts. Therefore, compared with the database of the FBI and the Centers for Disease Control and Prevention, it is more complete [3].

3.2 Overview of the data

With the help of Mathematica, we can convert the data from Jan. 1, 2015 to Dec. 31, 2019 into a more comprehensible figure, which shows the number of fatal police shootings on each day (Figure 1).

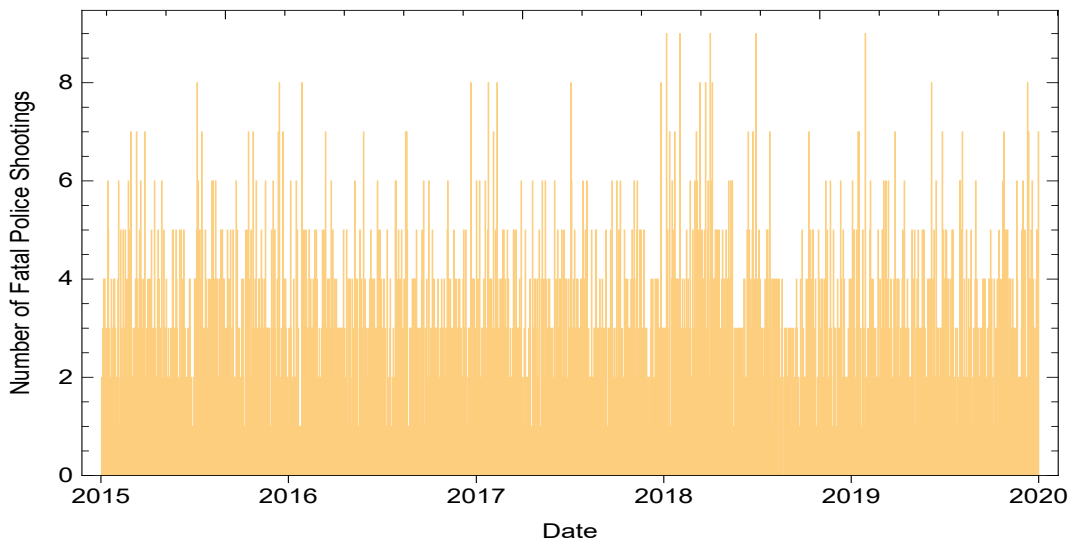


Figure 1: Number of fatal police shootings each day from Jan. 1, 2015 to Dec. 31, 2019.

We notice that there are five days with 9 fatal police shootings recorded and four of them happened in 2018.

4 Goodness-of-fit test for Poisson distribution

In *London murders: a predictable pattern?*, David Spiegelhalter and Arthur Barnett assumed that “if murders happened as random events, the number of murders each day would follow a Poisson distribution” [1]. Similarly, we may assume that the fatal police shootings happened as random events, and the number of fatal police shootings would also follow a Poisson distribution. In order to confirm our assumption, we will test whether the occurrence of fatal police shootings in the USA follows a Poisson distribution or not from 2015 to 2019. From the data, we get the following table (Table 1).

| Number of fatal police shootings in a day | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-----|-----|-----|-----|-----|-----|----|----|----|---|
| Observed days | 139 | 348 | 414 | 382 | 280 | 151 | 66 | 28 | 13 | 5 |

Table 1: Observed days with different numbers of fatal police shootings.

Let X denotes the number of fatal police shootings in a day, then the maximum-likelihood estimator for k is the sample mean [4],

$$\begin{aligned}\hat{k} = \bar{X} &= \frac{1}{5 \cdot 365 + 1} (139 \cdot 0 + 348 \cdot 1 + 414 \cdot 2 + 382 \cdot 3 + \\ &\quad 280 \cdot 4 + 151 \cdot 5 + 66 \cdot 6 + 28 \cdot 7 + 13 \cdot 8 + 5 \cdot 9) \\ &= 2.7043.\end{aligned}$$

Then, in order to use the multinomial distribution, we should calculate

$$\begin{aligned}P[X = 0] &= \frac{e^{-\hat{k}} \hat{k}^0}{0!} = 0.0669; & P[X = 1] &= \frac{e^{-\hat{k}} \hat{k}^1}{1!} = 0.1810; \\ P[X = 2] &= \frac{e^{-\hat{k}} \hat{k}^2}{2!} = 0.2447; & P[X = 3] &= \frac{e^{-\hat{k}} \hat{k}^3}{3!} = 0.2206; \\ P[X = 4] &= \frac{e^{-\hat{k}} \hat{k}^4}{4!} = 0.1491; & P[X = 5] &= \frac{e^{-\hat{k}} \hat{k}^5}{5!} = 0.0807; \\ P[X = 6] &= \frac{e^{-\hat{k}} \hat{k}^6}{6!} = 0.0364; & P[X = 7] &= \frac{e^{-\hat{k}} \hat{k}^7}{7!} = 0.0140; \\ P[X = 8] &= \frac{e^{-\hat{k}} \hat{k}^8}{8!} = 0.0047; \\ P[X \geq 9] &= 1 - P[X = 0] - P[X = 1] - \dots - P[X = 8] = 0.0019.\end{aligned}$$

Therefore, the distribution of X can be expressed as a new distribution with a categorical random variable with parameters $(p_0, p_1, \dots, p_9) = (0.0669, 0.1810, \dots, 0.0019)$.

After that, we need to calculate the expected days with $E_i = np_i$, where i is the category and n is the sample size $n = 5 \cdot 365 + 1 = 1826$.

$$\begin{aligned}E_0 &= 1826 \cdot 0.0669 = 122.19; & E_1 &= 1826 \cdot 0.1810 = 330.45; \\ E_2 &= 1826 \cdot 0.2447 = 446.81; & E_3 &= 1826 \cdot 0.2206 = 402.76; \\ E_4 &= 1826 \cdot 0.1491 = 272.30; & E_5 &= 1826 \cdot 0.0807 = 147.27; \\ E_6 &= 1826 \cdot 0.0364 = 66.38; & E_7 &= 1826 \cdot 0.0140 = 25.64; \\ E_8 &= 1826 \cdot 0.0047 = 8.67; & E_9 &= 1826 \cdot 0.0019 = 3.47.\end{aligned}$$

Besides, we should pay attention to the Cochran’s Rule, which requires that

$$\begin{aligned}E[X_i] &= np_i \geq 1 && \text{for all } i = 1, \dots, k, \\ E[X_i] &= np_i \geq 5 && \text{for 80\% of all } i = 1, \dots, k.\end{aligned}$$

We find that all of the E_i s are greater than 1 and only one out of ten that is smaller than 5, which mean 90% of the E_i s are greater than 5. Therefore, it satisfies the Cochran’s Rule, and we can create the following table (Table 2).

| Number of fatal police shootings in a day | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|--------|--------|--------|--------|--------|--------|-------|-------|------|------|
| Expected days | 122.19 | 330.45 | 446.81 | 402.76 | 272.30 | 147.27 | 66.38 | 25.64 | 8.67 | 3.47 |
| Observed days | 139 | 348 | 414 | 382 | 280 | 151 | 66 | 28 | 13 | 5 |

Table 2: Expected and observed days with different numbers of fatal police shootings.

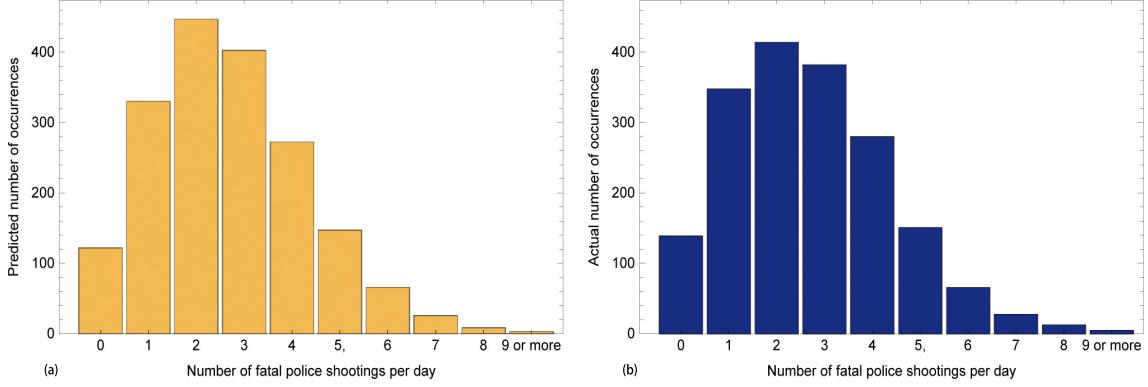


Figure 2: Days with different numbers of fatal police shootings: (a) expected; (b) observed.

Then, the hypothesis “ H_0 : the number of fatal police shootings follows a Poisson distribution with parameter $k = 2.7043$ ” is equivalent to “ H_0 : the number of fatal police shootings follows a multinomial distribution with parameters $(0.0669, 0.1810, \dots, 0.0019)$ ”. Besides,

$$X^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}$$

follows a chi-squared distribution with $N - 1 - m = 10 - 1 - 1 = 8$ degree of freedom, where O_i is the observed value and m is the number of parameters that we estimate. After we plug in the number, we get $X^2 = 10.94$. Let $\alpha = 0.05$, we have $\chi_{0.05, 8}^2 = 15.51 \geq 10.94$. Therefore, we are unable to reject H_0 at the 5% level of significance. We can calculate the P -value as follows

$$P = P[X^2 | H_0] \leq P[\chi_8^2 \geq 10.94] = 1 - P[\chi_8^2 \leq 10.94] = 1 - 0.7949 = 0.2051.$$

The P -value is quite large, therefore, we cannot reject H_0 , and we should consider that the number of fatal police shootings during Jan. 1, 2015 and Dec. 31, 2019 follows a Poisson distribution with parameter $k = 2.7043$ [5].

5 Dependence on weekday

From the data, we can also create the following figures that shows the number of fatal police shootings on each weekday and month (Figure 3).

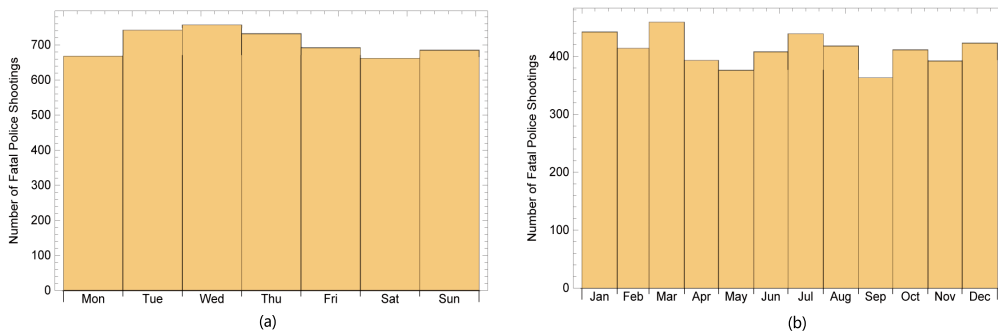


Figure 3: Number of occurrence of police shootings on (a) different weekdays and (b) different months.

We want to test whether there is evidence that the average number of police shootings depends on the weekday, then we should set the null hypothesis as “ H_0 : the number of fatal police shootings follows a multinomial distribution with the same parameter $p = \frac{1}{7}$ ” because there are seven different weekdays. There are 4938 fatal police shootings in total, so we can calculate the expected value on each weekday:

$$E_i = np_i = 4938 \times \frac{1}{7} = 705.43.$$

Then we can get the following table (Table 3).

| Weekdays | Mon. | Tue. | Wed. | Thu. | Fri. | Sat. | Sun. |
|------------------|--------|--------|--------|--------|--------|--------|--------|
| Expected numbers | 705.43 | 705.43 | 705.43 | 705.43 | 705.43 | 705.43 | 705.43 |
| Obseved numbers | 668 | 742 | 757 | 732 | 692 | 662 | 685 |

Table 3: Expected and observed number of fatal police shootings on diferent weekdays.

We notice that it satisfies the Cochran’s Rule because all the E_i s are greater than 5. Therefore,

$$X^2 = \sum_{i=1}^7 \frac{(O_i - E_i)^2}{E_i} = \frac{(668 - 705.43)^2}{705.43} + \dots + \frac{(685 - 705.43)^2}{705.43} = 12.17$$

follows a chi-square distribution with $N - 1 - m = 7 - 1 - 0 = 6$ degree of freedom. Let $\alpha = 0.05$, we have $\chi_{0.05, 6}^2 = 12.59 \geq 12.17$. Therefore, we fail to reject H_0 at the 5% level of significance, which means we can consider that the occurrence of fatal police shootings is independent on weekdays.

6 Confidence interval for the parameter k of the Poisson distribution

We know that for a Poisson distribution, we have $E[X] = k$ and $Var[X] = k$ and we’ve already proved that \bar{X} is an unbiased estimator for the parameter k [4]. In our case we have a large sample size, so we can assume that \bar{X} approximately follows a normal distribution with mean k and variance S/\sqrt{n} . Then,

$$\frac{\hat{k} - k}{S/\sqrt{n}}$$

is apprixomately standard-normally distributed.

It follows that the $100(1 - \alpha)\%$ confidence interval for k is given as follow [6]:

$$\hat{k} \pm z_{\alpha/2} S/\sqrt{n}$$

Furthermore, we note that S^2 is also an unbiased estimator for k [4]. Thus we can substitute S with $\sqrt{\hat{k}}$ and rewrite the $100(1 - \alpha)\%$ confidence interval:

$$\hat{k} \pm z_{\alpha/2} \sqrt{\hat{k}/n}$$

Based on the data from 2015 to 2018, we have 4938 fatal police shooting in the total 1461 days [2]. Thus we have:

$$\hat{k} = \bar{X} = \frac{3934}{1461} = 2.69. \quad (1)$$

We use 95% confidence interval as an example:

$$\hat{k} \pm z_{\alpha/2} \sqrt{\hat{k}/n} = 2.69 \pm 1.96 \times \sqrt{2.69/1461} = 2.69 \pm 0.084$$

7 Goodness-of-fit test for data for 2020

To check whether the data for 2020 so far follows a Poisson distribution, we will apply the goodness-of-fit test. We first use the unbiased estimator \bar{X} to estimate the parameter k . Based on the available data [2], we have:

$$\hat{k}_{2020} = \bar{X} = \frac{303}{106} = 2.86. \quad (2)$$

Then we can calculate the expected probability and the corresponding frequency of the Poisson distribution, which is calculated as follow and recorded in Table 4.

$$\begin{aligned} P[X=0] &= \frac{e^{-\hat{k}} \hat{k}^0}{0!} = 0.057; & P[X=1] &= \frac{e^{-\hat{k}} \hat{k}^1}{1!} = 0.164; \\ P[X=2] &= \frac{e^{-\hat{k}} \hat{k}^2}{2!} = 0.234; & P[X=3] &= \frac{e^{-\hat{k}} \hat{k}^3}{3!} = 0.223; \\ P[X=4] &= \frac{e^{-\hat{k}} \hat{k}^4}{4!} = 0.160; & P[X=5] &= \frac{e^{-\hat{k}} \hat{k}^5}{5!} = 0.091; \\ P[X=6] &= \frac{e^{-\hat{k}} \hat{k}^6}{6!} = 0.044; & P[X=7] &= \frac{e^{-\hat{k}} \hat{k}^7}{7!} = 0.018; \\ P[X \geq 8] &= 1 - P[X=0] - P[X=1] - \dots - P[X=7] = 0.009. \end{aligned}$$

$$\begin{aligned} E_0 &= 106 \cdot 0.057 = 6.1; & E_1 &= 106 \cdot 0.164 = 17.4; \\ E_2 &= 106 \cdot 0.234 = 24.8; & E_3 &= 106 \cdot 0.223 = 23.7; \\ E_4 &= 106 \cdot 0.160 = 16.9; & E_5 &= 106 \cdot 0.091 = 9.7; \\ E_6 &= 106 \cdot 0.044 = 4.6; & E_7 &= 106 \cdot 0.018 = 1.9; \\ E_8 &= 106 \cdot 0.009 = 0.9; \end{aligned}$$

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 or more |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-----------|
| Probability | 0.057 | 0.164 | 0.234 | 0.223 | 0.160 | 0.091 | 0.044 | 0.018 | 0.009 |
| Expected Numbers | 6.1 | 17.4 | 24.8 | 23.7 | 16.9 | 9.7 | 4.6 | 1.9 | 0.9 |

Table 4: Probability and expected frequency of days with different numbers of fatal police shooting in the US.

Besides, we should pay attention to the Cochran's Rule, which requires that

$$\begin{aligned} E[X_i] &= np_i \geq 1 && \text{for all } i = 1, \dots, k, \\ E[X_i] &= np_i \geq 5 && \text{for 80\% of all } i = 1, \dots, k. \end{aligned}$$

Thus we combine the last three categories together and carry out the null hypothesis that

H_0 : The data for 2020 so far follows a Poisson distribution with parameter $k = 2.86$.

We compare the observed data with the expected data to decide whether to reject the null hypothesis or not. The pattern is recorded in Table 5 and Figure 4.

| Numbers of days with the following numbers of fatal police shooting per day: | | | | | | | |
|--|-----|------|------|------|------|-----|-----------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 or more |
| Expected | 6.1 | 17.4 | 24.8 | 23.7 | 16.9 | 9.7 | 7.4 |
| Observed | 7 | 18 | 27 | 18 | 14 | 13 | 9 |

Table 5: Expected and observed number of days with 0, 1, 2 etc. fatal police shooting in the US on 105 days between January 1st 2020 and April 15th 2020.

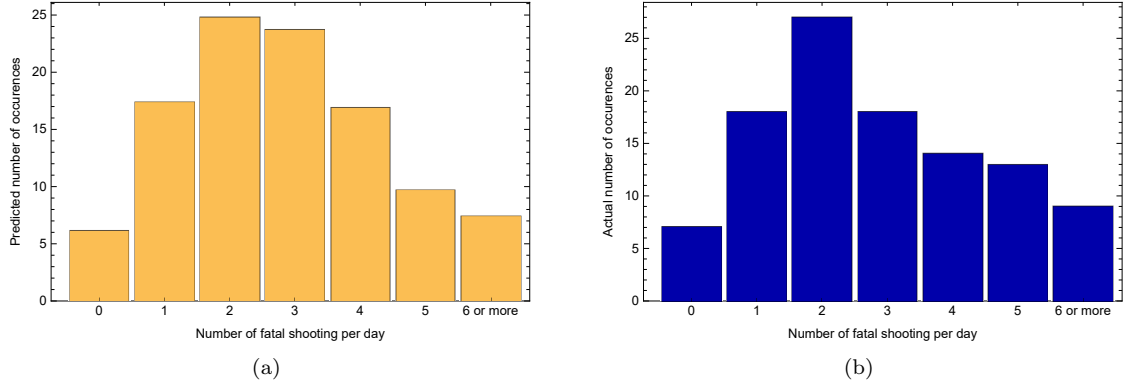


Figure 4: Frequency of occurrence of days with different numbers of fatal police shooting recorded in the US between January 1st 2020 and April 15th 2020: (a) expected; (b) observed.

From the above data and bar charts, we can find that the observed data doesn't fit with the expected one. We will apply the χ^2 test to the data to give a quantitative result.

For $N = 7$ categories, the statistic

$$X^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}$$

then follows a chi-distribution with $N - 1 - m = 7 - 1 - 1 = 5$ degrees of freedom.

$$X^2 = \frac{(7 - 6.1)^2}{6.1} + \frac{(18 - 17.4)^2}{17.4} + \frac{(27 - 24.8)^2}{24.8} + \frac{(18 - 23.7)^2}{23.7} + \frac{(14 - 16.9)^2}{16.9} + \frac{(13 - 9.7)^2}{9.7} + \frac{(9 - 7.4)^2}{7.4} = 3.68$$

On the other hand, the critical value for $\alpha = 0.05$ is $\chi_{0.05,5}^2 = 11.1$. Since $X^2 < \chi_{0.05,5}^2$, we have no evidence to reject the null hypothesis that the data for 2020 so far follows a Poisson distribution with $k = 2.86$.

The result shows that the data for 2020 follows a Poisson distribution, which corresponds to our expectation. However, we also find that the pattern doesn't fit with the expected one very well. From Figure 4 we can find that the number of 3 shooting per day is relatively fewer than expected [5].

8 Prediction interval for 2020

According to the article "Improved closed-form prediction intervals for binomial and Poisson distributions", we can use the Nelson prediction interval to predict the future counts of a Poisson distribution [7].

We will first define some parameters used in calculation: let λ be the mean of the Poisson distribution; in the sample size of n , there are a total of X existing arrivals; in another sample size of m , there will be a total of Y future arrivals [7].

First it is easy to find that the estimator $\hat{\lambda} = \bar{X} = \frac{X}{n}$. At the same time, $\hat{\lambda} = \frac{\hat{Y}}{m}$. According to these two equations, we have:

$$\hat{\lambda} = \frac{X}{n} \hat{Y} = \frac{mX}{n} \quad (3)$$

Also to estimate the variance of $m\hat{\lambda} - Y$, we calculate as follow:

$$\begin{aligned}
\hat{Var}(m\hat{\lambda} - Y) &= \hat{Var}\left(\frac{mX}{n} - Y\right) \\
&= \frac{m^2}{n^2} \hat{Var}(x) + \hat{Var}(Y) \\
&= \frac{m^2}{n^2} n\hat{\lambda} + m\hat{\lambda} \\
&= m^2\lambda\left(\frac{1}{n} + \frac{1}{m}\right)
\end{aligned}$$

Thus we have:

$$\hat{Var}(m\hat{\lambda} - Y) = m^2\hat{\lambda}\left(\frac{1}{n} + \frac{1}{m}\right) \quad (4)$$

Then we need to verify that

$$\frac{m\hat{\lambda} - Y}{\sqrt{\hat{Var}(m\hat{\lambda} - Y)}}$$

is standard-normally distributed, which is equal to verifying that $m\hat{\lambda} - Y$ has mean 0. We've already proved that $\hat{\lambda} = X/n$ and $\hat{Y} = mX/n$. Therefore, the formula above follows a standard normal distribution [7].

According to the article, the Nelson prediction interval is given by $[\lceil L \rceil, \lfloor U \rfloor]$ [7]. Thus instead of conducting the two-sided test, we need to conduct two one-sided respectively. We calculate $\lceil L \rceil$ as an example.

Since $\frac{m\hat{\lambda} - Y}{\sqrt{\hat{Var}(m\hat{\lambda} - Y)}}$ follows a standard normal distribution, we have

$$\begin{aligned}
\alpha &= P\left[\frac{m\hat{\lambda} - Y}{\sqrt{\hat{Var}(m\hat{\lambda} - Y)}} \leq z_{1-\alpha}\right] \\
&= P\left[m\hat{\lambda} - z_{1-\alpha}\sqrt{\hat{Var}(m\hat{\lambda} - Y)} \leq Y\right] \\
&= P\left[\hat{Y} - z_{1-\alpha}\sqrt{m\hat{Y}\left(\frac{1}{n} + \frac{1}{m}\right)}\right]
\end{aligned}$$

Similarly, we can calculate the upper bound and thus we can get

$$[L, U] = \hat{Y} \pm z_{1-\alpha}\sqrt{m\hat{Y}\left(\frac{1}{m} + \frac{1}{n}\right)} \quad (5)$$

From the above formula, we can derive a 95% prediction interval for 2020 based on the data for 2015 to 2019 [2]. In this case, the parameters are listed as follow:

$$\begin{aligned}
X &= 4938 & n &= 1826 \\
\alpha &= 0.05 & z_{1-\alpha} &= 1.645
\end{aligned}$$

Then we can use the data to obtain the prediction interval:

$$\frac{mX}{n} \pm z_{1-\alpha}\sqrt{\frac{m^2X}{n}\left(\frac{1}{m} + \frac{1}{n}\right)} = \frac{4938m}{1826} \pm 1.645 \times \sqrt{\frac{4938m^2}{1826}\left(\frac{1}{m} + \frac{1}{1826}\right)}$$

After simplification, we obtain the 95% prediction interval shown by the following formula:

$$2.704m \pm 1.645\sqrt{2.704m + 0.0015m^2} \quad (6)$$

Figure 5 shows the predicted number of fatal police shooting in the US during 2020 (the black solid line) and the 95% prediction limits (the dash lines) with the data for 2020 so far (the red dots).

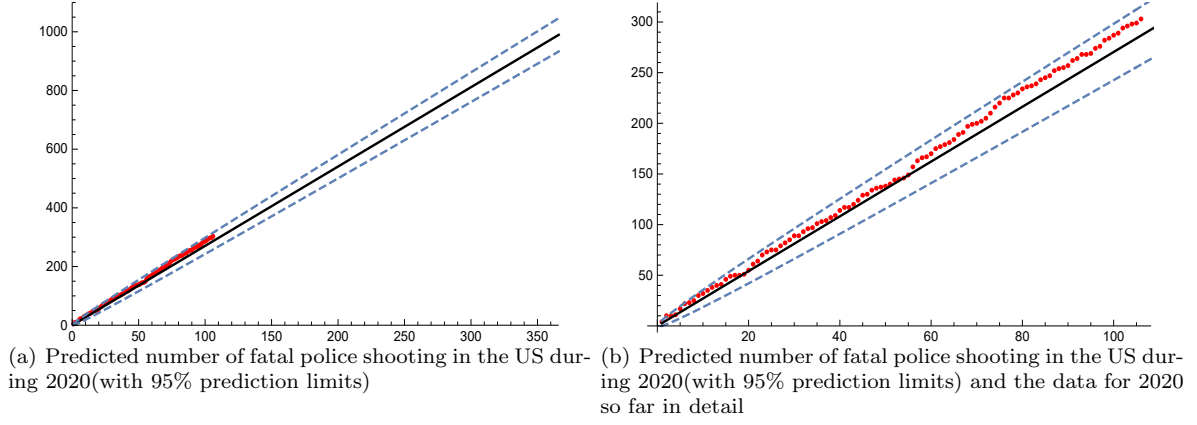


Figure 5: Predicted number of fatal police shooting in the US during 2020 with 95% prediction limits: (a) overview; (b) detailed.

From Figure 5, we can find that the data for 2020 lies in the prediction interval quite well. This indicates that based on the data for 2015-2019, we can roughly predict the future tendency of the fatal police shooting.

9 The influence of the outbreak of Coronavirus

We must note that the Poisson distribution model is based on the assumption that “the level of violence remains the same” [1]. However, when taking the outbreak of the Coronavirus into account, we should be careful whether this will influence our original assumption.

Before focusing on the influence of the outbreak of COVID-19 on fatal police shooting, we may first have a general idea of the situation of COVID-19 in the US. The following figure (Figure 6) is a screenshot of the number of people confirmed COVID-19 in the US provided by Johns Hopkins University. The marked point is the data for March 27th, when the confirmed people first exceeds 100k people and reaches 101.7k people [8].

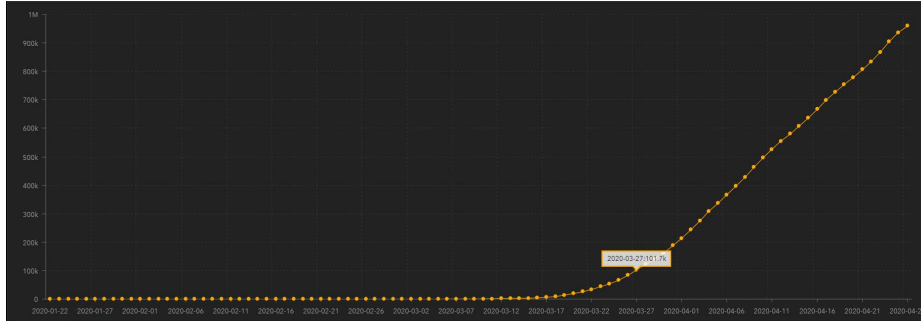


Figure 6: Number of people confirmed COVID-19 in the US [8].

If we look deep into the data, we can find that there’s no strong evidence indicating that the outbreak of COVID-19 has a strong influence on the fatal police shooting generally since the previous sections have reached the conclusion that the data for 2020 still follows a Poisson distribution and the data so far all falls in the prediction interval.

However, we may still get some information from the data. If we turn to Figure 5, we may find that for the first half of the data so far, it is really close to the exact predicted line. For the second half of the data, however, the number of fatal police shooting is rising relatively faster and is approaching the upper bound of the prediction interval. The turning point is in March, which is also the time when COVID-19 broke out in the US.

The reason why the number is rising may be due to the following several reasons:

1. The outbreak of COVID-19 makes the residents panic and thus more people are likely to do extreme things.

2. The US government doesn't respond to the epidemic properly and effectively, leading to some citizens' discontent to the government.
3. Some residents doesn't obey the government's control measures and have conflicts with the police.

10 Conclusion

The project mainly analyze the data of fatal police shooting in the US for 2015-2019 and predict the possible tendency in 2020. We first specify the definition of "fatal" and give a general overview of the data. For the first half of the project, we mainly analyze the data for 2015-2019 by checking its fitness to the Poisson distribution, the confidence interval of the parameter k of the Poisson distribution and the relationship between fatal police shooting and each weekday and month. In the second half of the project, we focus on the data for 2020. We test its fitness to the Poisson distribution based on the existing data for 2020 and also predict its future pattern. At last we take the outbreak of COVID-19 into account and analyze its influence on fatal police shooting.

Through this project, we confirm that the data for fatal police shooting follows a Poisson distribution. Based on the previous data, we can also predict the future possible numbers of fatal police shooting to a certain extent. Moreover, considering the special situation this year, namely the outbreak of COVID-19, it's still not quite clear whether it will significantly influence fatal police shooting.

11 Reference

- [1] D. Spiegelhalter and A. Barnett, "London murders: a predictable pattern?" *Significance*, vol. 6, no. 1, pp. 5-8, 2009. <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2009.00334.x/abstract>.
- [2] The Washington Post, "Data-police-shootings," GitHub. <https://github.com/washingtonpost/data-police-shootings>. Accessed April 28, 2020.
- [3] The Washington Post, "Fatal force." <https://www.washingtonpost.com/graphics/national/police-shootings-2016/>. Accessed April 19, 2020.
- [4] H. Hohberger, "Ve401_video_12," UMJI-SJTU, pp. 18-19, 2020.
- [5] H. Hohberger, "Ve401_video_23," UMJI-SJTU, pp. 9-36, 2020.
- [6] H. Hohberger, "Ve401_video_13," UMJI-SJTU, pp. 8-9, 2020.
- [7] K. Krishnamoorthy and Jie Peng, "Improved closed-form prediction intervals for binomial and Poisson distributions" *Journal of Statistical Planning and Inference*, vol. 141, no.5, pp. 1709-1718, 2011. <http://www.sciencedirect.com/science/article/pii/S0378375810005215>.
- [8] Johns Hopkins University. Coronavirus Resource Center. "COVID-19 United States Cases by County." <https://coronavirus.jhu.edu/us-map>. Accessed April 28, 2020.

12 Appendix

12.1 Code for 3.2 Overview of the data

```
data := Import["shootingdata.csv"][[2 ;; 4939, 3]]
DateHistogram[data, "Day",
  Ticks -> {"2015", "2016", "2017", "2018", "2019", "2020"},
  Automatic}, Frame -> True,
  FrameLabel -> {"Date", "Number of Fatal Police Shootings"}]
```

12.2 Code for 4 Goodness-of-fit test for Poisson distribution

```
BarChart[{122.19, 330.45, 446.81, 402.76, 272.30, 147.27, 66.38,
  25.64, 8.67, 3.47},
  ChartLabels -> {"0", "1", "2", "3", "4", "5", "6", "7", "8",
  "9 or more"}, Frame -> True,
  FrameLabel -> {"Number of fatal police shootings per day",
  "Predicted number of occurrences"}, PlotRange -> {0, 450}]
```

```
BarChart[{139, 348, 414, 382, 280, 151, 66, 28, 13, 5},
  ChartLabels -> {"0", "1", "2", "3", "4", "5", "6", "7", "8",
  "9 or more"}, Frame -> True,
  FrameLabel -> {"Number of fatal police shootings per day",
  "Actual number of occurrences"}, PlotRange -> {0, 450},
  ChartStyle -> Darker[Blue]]
```

12.3 Code for 5 Dependence on weekday

```
data := Import["shootingdata.csv"][[2 ;; 4939, 3]]
DateHistogram[data, "Day", DateReduction -> "Week", Frame -> True,
  FrameLabel -> {None, "Number of Fatal Police Shootings"}]
DateHistogram[data, "Month", DateReduction -> "Year", Frame -> True,
  FrameLabel -> {None, "Number of Fatal Police Shootings"}]
```

12.4 Code for 7 Goodness-of-fit test for data for 2020

```
BarChart[{6.1, 17.4, 24.8, 23.7, 16.9, 9.7, 7.4},
  ChartLabels -> {"0", "1", "2", "3", "4", "5", "6 or more"},
  Frame -> True,
  FrameLabel -> {"Number of fatal shooting per day",
  "Predicted number of occurrences"}]
```

```
BarChart[{7, 18, 27, 18, 14, 13, 9},
  ChartLabels -> {"0", "1", "2", "3", "4", "5", "6 or more"},
  Frame -> True,
  FrameLabel -> {"Number of fatal shooting per day",
  "Actual number of occurrences"}, ChartStyle -> Darker[Blue]]
```

12.5 Code for 8 Confidence interval for the parameter k of the Poisson distribution

```
Show[ ListPlot [
  {4, 10, 10, 11, 17, 22, 23, 25, 30, 32,
  35, 38, 40, 41, 46, 49, 50, 50, 51, 55,
  61, 64, 70, 73, 75, 75, 79, 82, 85, 89, 89,
  93, 96, 97, 101, 103, 104, 107, 109, 114, 117,
  117, 120, 124, 129, 130, 134, 136, 137, 138, 140,
```

```

144, 145, 146, 149, 157, 163, 166, 167, 170,
175, 177, 179, 181, 184, 189, 191, 197, 199, 200,
202, 205, 210, 216, 220, 225, 225, 228, 230, 234,
236, 237, 239, 243, 245, 247, 252, 254, 255, 257, 262,
264, 268, 268, 269, 274, 276, 282, 284, 287, 289,
294, 296, 298, 299, 303}, PlotStyle -> Red,
PlotRange -> {{0, 366}, {0, 1100}}, PlotStyle -> SmallCircle],
Plot[2.704*m, {m, 1, 366}, PlotStyle -> Black, Frame -> True,
PlotRange -> {{0, 366}, {0, 1100}}],
Plot[2.704*m + 1.645*Sqrt[2.704*m + 0.0015*m^2], {m, 1, 366},
PlotStyle -> Dashed, PlotRange -> {{0, 366}, {0, 1100}}],
Plot[2.704*m - 1.645*Sqrt[2.704*m + 0.0015*m^2], {m, 1, 366},
PlotStyle -> Dashed, PlotRange -> {{0, 366}, {0, 1100}}]]

Show[ListPlot[
{4, 10, 10, 11, 17, 22, 23, 25, 30, 32,
35, 38, 40, 41, 46, 49, 50, 50, 51, 55,
61, 64, 70, 73, 75, 75, 79, 82, 85, 89, 89,
93, 96, 97, 101, 103, 104, 107, 109, 114, 117,
117, 120, 124, 129, 130, 134, 136, 137, 138, 140,
144, 145, 146, 149, 157, 163, 166, 167, 170,
175, 177, 179, 181, 184, 189, 191, 197, 199, 200,
202, 205, 210, 216, 220, 225, 225, 228, 230, 234,
236, 237, 239, 243, 245, 247, 252, 254, 255, 257, 262,
264, 268, 268, 269, 274, 276, 282, 284, 287, 289,
294, 296, 298, 299, 303}, PlotStyle -> Red,
PlotStyle -> SmallCircle],
Plot[2.704*m, {m, 1, 366}, PlotStyle -> Black, Frame -> True],
Plot[2.704*m + 1.645*Sqrt[2.704*m + 0.0015*m^2], {m, 1, 366},
PlotStyle -> Dashed],
Plot[2.704*m - 1.645*Sqrt[2.704*m + 0.0015*m^2], {m, 1, 366},
PlotStyle -> Dashed]]

```