UM–SJTU Joint Institute

Probability and Statistics in Engineering

(Ve401)

Project Report

Project 2

Police Shootings in the United States

Group 16

Name: Chen Ronghan	ID: 5143709200
Name: Zhang Shengnan	ID: 5143709004
Name: Li Lewei	ID: 515370910054
Name: Sui Zijian	ID: 515370910038
Name: Yu Fan	ID: 515370910237

Date: 2017/4/23

Abstract

aaaaaaaa

Contents

1 Introduction

aaaaaaa

2 Summary of the source

The data are excerpted from the database of every fatal shooting in the United States by a police officer in the line of duty since Jan.1^{st}, 2015 compiled by the Washington Post. The details of killing were tracked by the Post by culling local news reports, law enforcement websites and social media, and by monitoring independent databases. After 2016, the Post gathered additional information about officers by starting filing open-records requests with departments. In addition, the Post collected information from people who knew about the details by email.

The term "fatal police shooting" here only refers to those shootings in which a police officer, in the line of duty, shoots and kills a civilian. In other words, it doesn't include deaths of people in police custody, fatal shootings by off-duty officers or non-shooting deaths.

3 Police Shooting data

By Mathematica, we obtain the histogram of Police Shooting data between January 1^{st}, 2015 and December 31^{st}, 2016 available from the database [2].

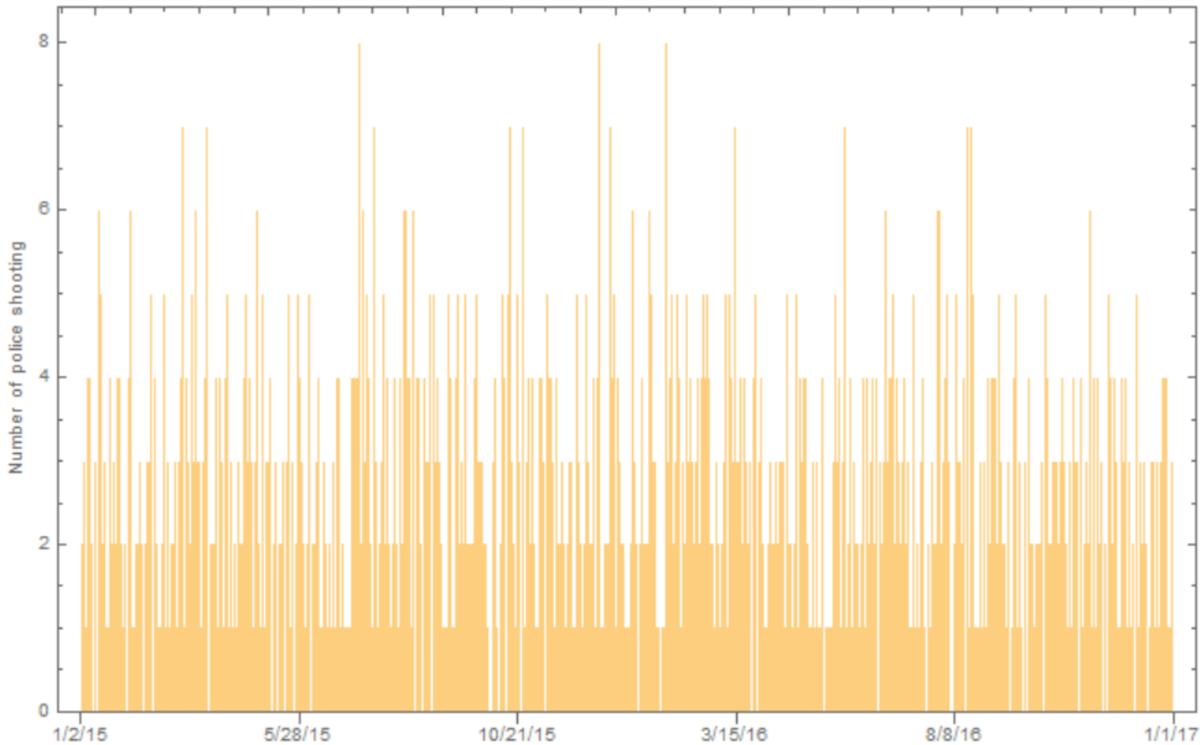

Figure 1: Police Shooting data between January 1^{st}, 2015 and December 31^{st}, 2016.

4 Test for Poisson distribution

Spiegelhalter and Barnett gave out an estimation that the data follows a Poisson distribution with parameter $\hat{k} = 0.44$. Now for the police shooting data in Year 2015-2016, the estimator for k is

$$\hat{k} = \bar{X} = 1954/731 = 2.67.$$

In order to apply the Goodness-of-fit Test, we first calculate

$$P[X=0] = \frac{e^{-\hat{k}}\hat{k}^0}{0!} = 0.069$$

$$P[X=1] = \frac{e^{-\hat{k}}\hat{k}^1}{1!} = 0.185$$

$$P[X=2] = \frac{e^{-\hat{k}}\hat{k}^2}{2!} = 0.247$$

$$P[X=3] = \frac{e^{-\hat{k}}\hat{k}^3}{3!} = 0.220$$

$$P[X=4] = \frac{e^{-\hat{k}}\hat{k}^4}{4!} = 0.147$$

$$P[X=5] = \frac{e^{-\hat{k}}\hat{k}^5}{5!} = 0.078$$

$$P[X=6] = \frac{e^{-\hat{k}}\hat{k}^6}{6!} = 0.034$$

$$P[X=7] = \frac{e^{-\hat{k}}\hat{k}^7}{7!} = 0.013$$

$$P[X >= 8] = 1 - P[X=0] - P[X=1] - ...... - P[X=7] = 0.007$$

Table 1: Observed and expected number of days in TWO years(731 days)

	Numbers of days:							
	0	1	2	3	4	5	6	>= 7
Expected	50.44	135.24	180.56	160.82	107.46	57.02	24.85	14.62
Observed	50	149	163	155	115	60	23	16

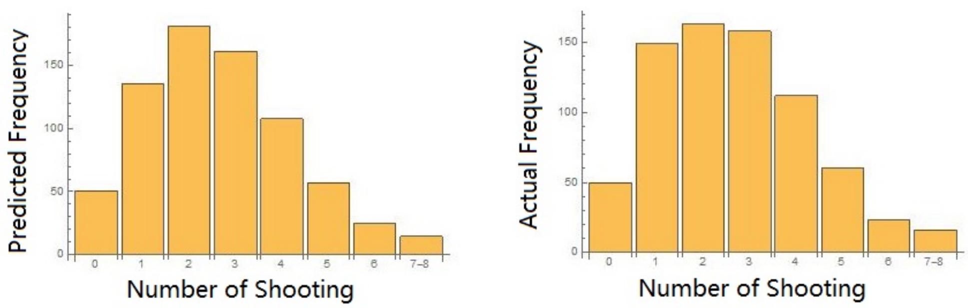

Figure 2: Frequency of Occurrence in two years

Then the test

$$H_0 : \text{the data follows a Poisson distribution with } k = 2.67$$

For $N = 8$ categories, the statistic

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

follows a chi-squared distribution with N-1-m=8-1-1=6 degree of freedom. Now

$$X^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 4.28 < \chi^2_{0.05,6} = 15.5,$$

so we cannot reject H_0 at 5% level of significance.

Next, we will test the null hypothesis in case of individual year. First, for Year 2015(365 days), the observed and expected frequency are shown below.

Table 2: Observed and expected number of days in 2015

	Numbers of days:							
	0	1	2	3	4	5	6	>= 7
Expected	25.19	67.53	90.16	80.30	53.66	28.47	12.41	7.30
Observed	24	73	88	69	55	33	13	10

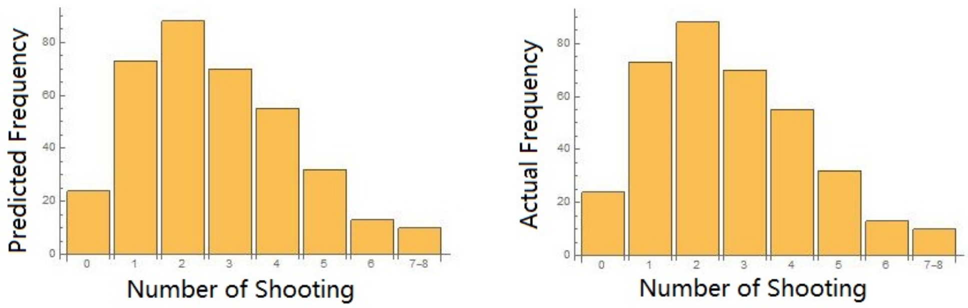

Figure 3: Frequency of Occurrence in 2015

Now N-1-m=8-1-1=6.

$$X^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 3.92 < \chi^2_{0.05,6} = 15.5,$$

so we cannot reject H_0 at 5% level of significance.

Then for Year 2016(366 days),

Table 3: Observed and expected number of days in 2016

	Numbers of days:							
	0	1	2	3	4	5	6	>= 7
Expected	25.25	67.71	90.40	80.52	53.80	28.55	12.44	7.32
Observed	26	76	75	86	60	27	10	6

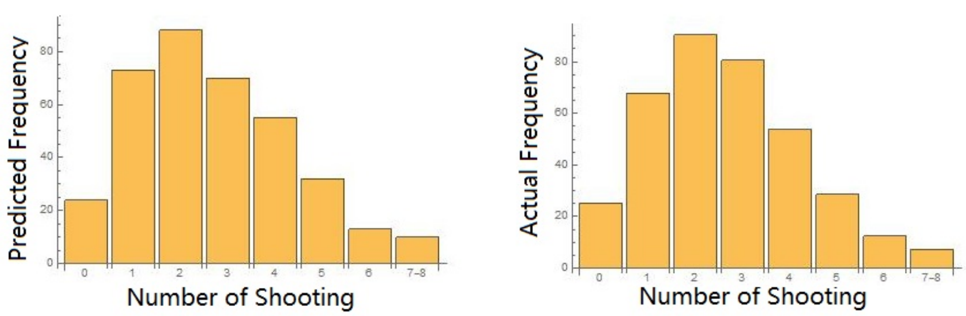

Figure 4: Frequency of Occurrence in 2016

$$X^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 5.55 < \chi^2_{0.05,6} = 15.5,$$

so we still cannot reject H_0 at 5% level of significance.

The P-value we are using for the test is 0.05.

Therefore, using the police shooting data, there is strong evidence that the occurrence of shootings follows a Poisson distribution. The data for both individual years are almost the same, and data for Year 2015 fits better to the model than that for Year 2016 does. The total goodness-of-fit is quite well.

5 Weekday Dependence

Given the data of fatal police shootings on each day, we are interested in whether the number of shootings is dependent on weekdays. Counted by Excel, the number of shootings for each day of the week is shown in the table below.

	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun.
Number of shootings	249	294	307	285	273	269	277

Table 4: Number of shootings for each day of the week.

Then by Mathematica, we obtain the distribution of the number of shootings for each day of the week, which is shown below.

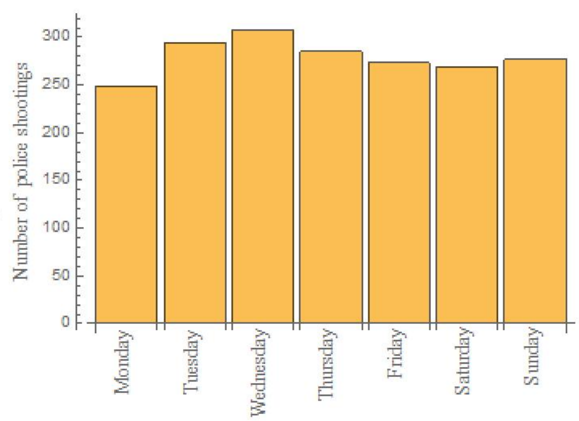

Figure 5: Distribution of the number of shootings for each day of the week.

We want to test whether these data conform to a discrete uniform distribution on $\Omega = 1, 2, ..., 7$ at a level of significance $\alpha = 0.05$.

We test

H_0: The data follow a multinomial distribution with parameters $(p_1, ..., p_7) = (\frac{1}{7}, ..., \frac{1}{7})$.

And we have $E_i = \frac{1}{7} \cdot 1954 = 279.1$ for $i = 1, ..., 7$.

The observed test statistic is

$$\sum_{i=1}^{7} \frac{(O_i-E_i)^2}{E_i} = \frac{(249-279.1)^2}{279.1} + \cdots + \frac{(277-279.1)^2}{279.1} = 7.47.$$

The statistic follows a chi-squared distribution with $7 - 1 = 6$ degrees of freedom. Since $\chi^2_{0.05,6} = 12.592$, the P-value of the test is greater than 5%. There is not enough evidence to reject H_0.

We conclude that there is no evidence that the number of police shootings is dependent on weekdays.

6 Gaps between shooting

(a)If time is considered as a continuum, then the distribution formula is

$$f(t) = 2.673e^{-2.673t}$$

where t refers to consecutive days free of police shooting. Then we define the number of consecutive days respectively as $N(x)$,

$$N(0) = 731 \times \int_0^1 2.673e^{-2.673t}\,dt = 680.53$$

$$N(1) = 731 \times \int_1^2 2.673e^{-2.673t}\,dt = 46.98$$

$$N(\geq 2) = 731 \times \int_2^3 2.673e^{-2.673t}\,dt = 3.48$$

(b)If discrete days are considered, then the distribution formula is

$$f(n) = \frac{2.673^n e^{-2.673}}{n!}$$

where n refers to numbers of police shooting for certain day. Then we have

$$N(0) = 731\times(1 - f(0)) = 731 \times \left(1 - \frac{2.673^0 e^{-2.673}}{0!}\right) = 680.53$$

$$N(1) = 731\times f(0)\times(1 - f(0)) = 46.98$$

$$N(\geq 2) = 731\times f^2(0)\times(1 - f(0)) = 3.48$$

The actual values over 2 years are

$$N(0) = 681$$

$$N(1) = 48$$

$$N(\geq 2) = 1$$

We can plot following figures.

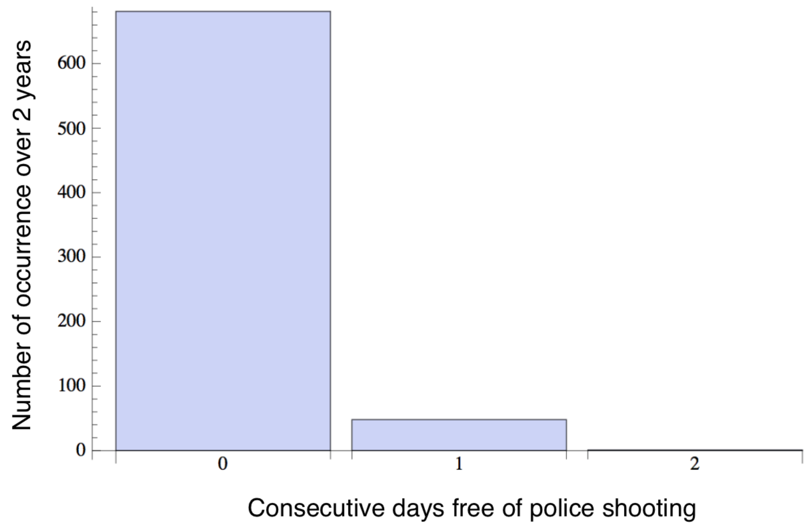

Figure 6: Prediction via method (a)

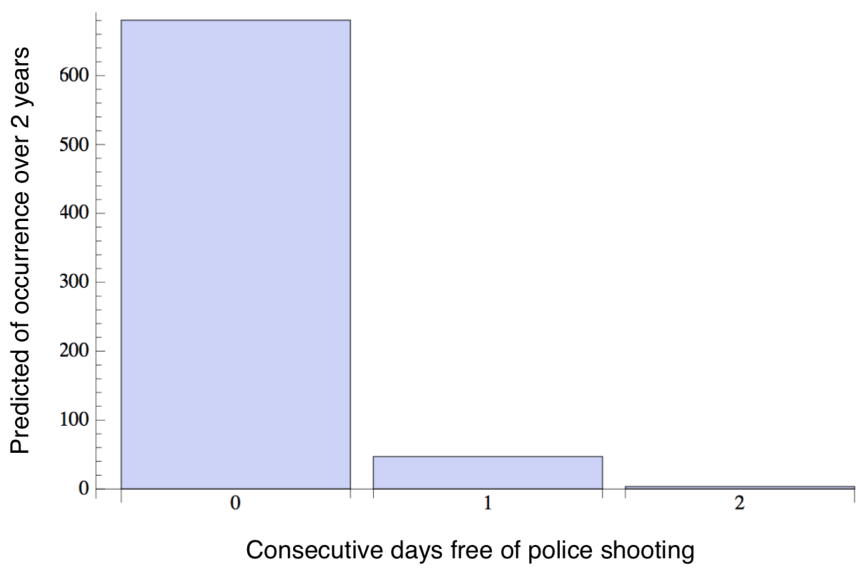

Figure 7: Prediction via method (b)

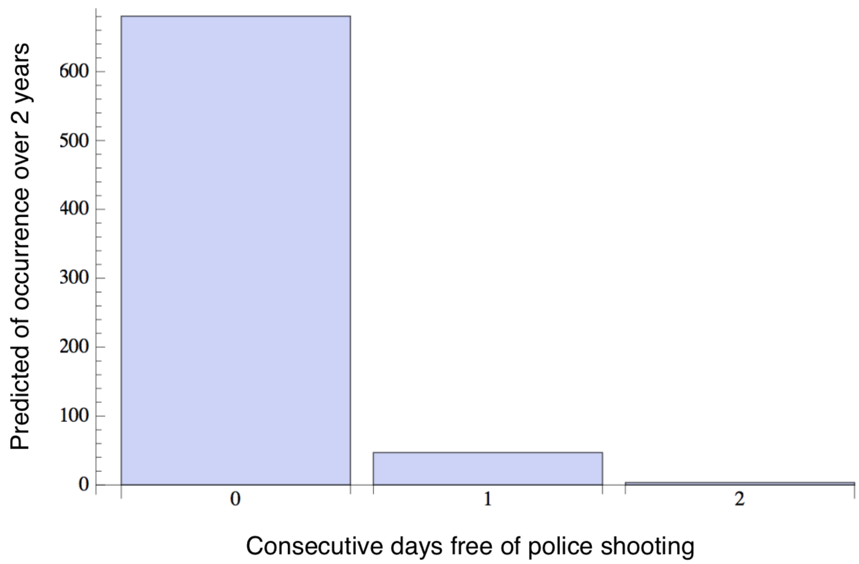

Figure 8: Actual values

7 Cumulative number of shooting

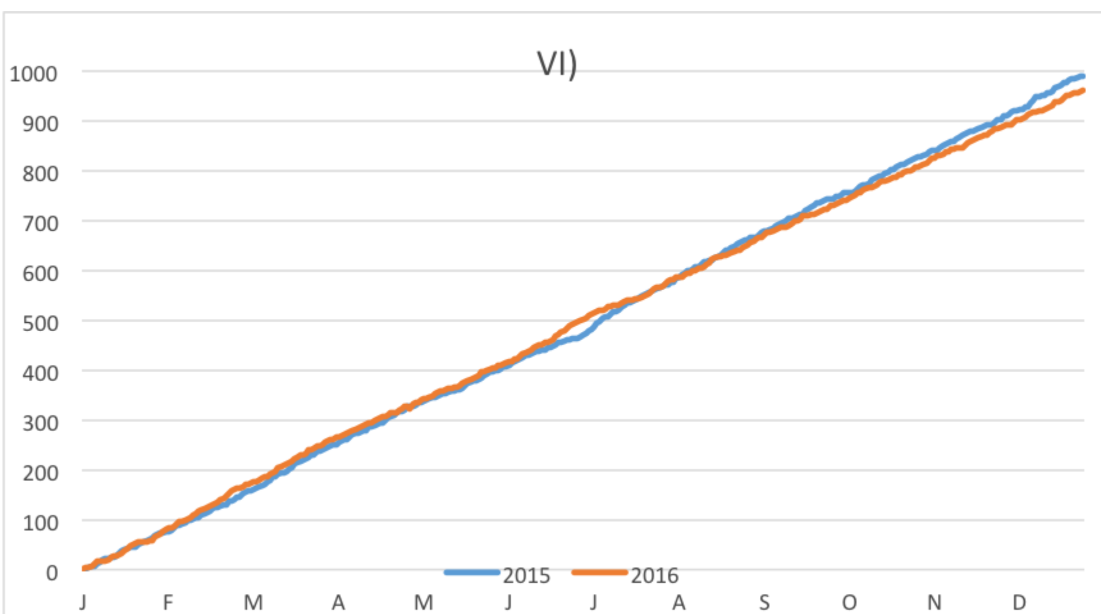

Figure 9: the cumulative number of mass shootings for 2015 and 2016

From the figure above we can see the rate of mass shooting remains almost the same over the years.

References

[1] D. Spiegelhalter and A. Barnett. London murders: a predictable pattern? Significance, 6(1):5?8, 2009. http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2009.00334.x/abstract [Online; accessed 5-July-2015].

[2] The Washington Post. Fatal force. https://www.washingtonpost.com/graphics/national/police-shootings-2016/. Web. Accessed February 16th, 2017.