
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

Contents

1	Abstract	2
2	Introduction	4
2.1	Background	4
2.2	Objectives	4
3	Data analysis	4
3.1	Data source	4
3.2	Overview of the data	4
4	Goodness-of-fit test for Poisson distribution	5
5	Reference	6
6	Appendix	7

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 or independence 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).

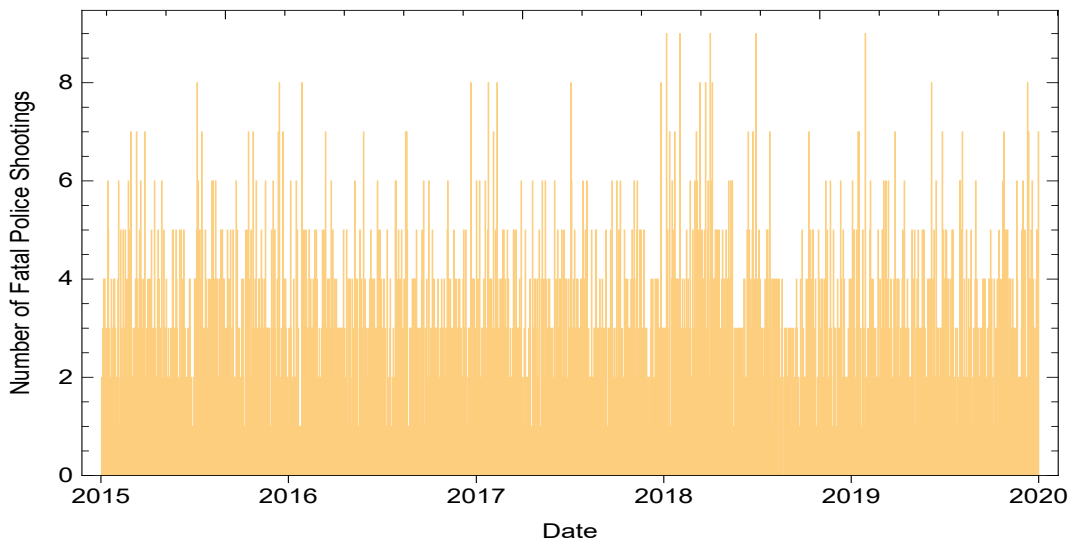


Figure 1: Number of fatal police shootings each day from Jan. 1, 2015 to Dec. 31, 2019.

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.

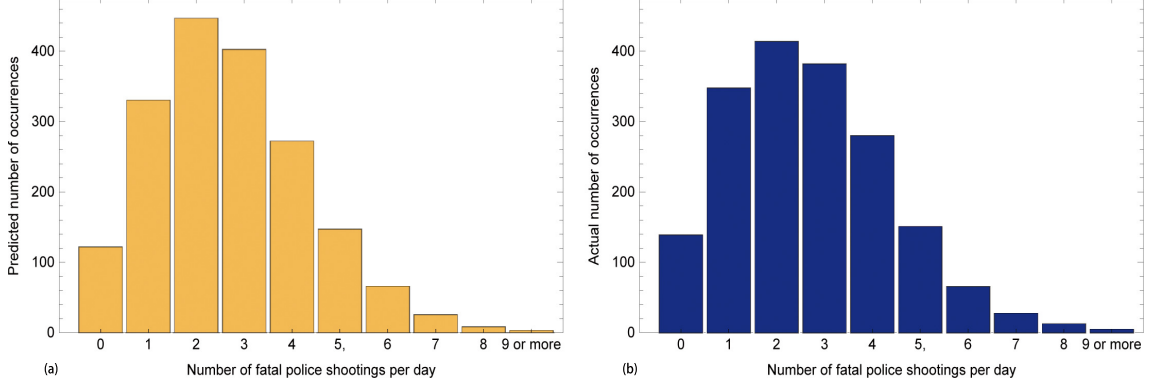


Figure 2: Days with different numbers of fatal police shootings: (a) expected; (b) observed.

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

6 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 19, 2020.
- [3] The Washington Post, “*Fata 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.

7 Appendix

$$f[a_-, b_-] := a^2 + b^2 + 4b^3 + 25a*b*c \quad ;$$