

Ve401 Probabilistic Methods in Engineering

Summer 2018 Term Project 2

Date Due: 12:00 PM, Friday, the 3rd of August 2018



General Information

The goal of this term project is to help you apply your new-found knowledge of probability theory and statistics in extended tasks that are beyond the scope of ordinary assignments. **It is strongly recommended that you do not leave the entire project to the last minute** but rather commence work on individual parts as soon as you are able to do so.

Group Work

You will be divided into groups of *4–5 students* each.

Each group member must be familiar with and have contributed to each part of the project report. **You may not divide up the work in such a way that only certain members are involved with certain parts.** In the event of an Honor Code violation (plagiarism or other), all members of the group will be held equally responsible for the violation. Exceptions may only be made, at my discretion, in exceptional situations.

It is therefore all group members' duty to ensure that all collaborators' contributions are plausibly their own and to check on all collaborators' work progress and verify their contributions within reason.

Project Report

The term project will be submitted **electronically only** as a typed report. Handwritten submission will not be accepted! It is recommended that you use a professional type-setting program (such as \LaTeX) for your report. Unless you are able to ensure a unified font size and style for formulas and text in Microsoft Word, use of Word is *not recommended*.

Grading Policy

This term project accounts for 20% of the course grade; it will be scored based on

- **Form (4 points):** Does the report contain essential elements, such as a cover page (with title, date, list of authors), a synopsis (abstract giving the main conclusions of the project), table of contents, clear section headings, introduction, clear division into sections and appendices with informative titles and bibliography (if applicable)? Are the pages numbered? Are the text and formulas composed in a unified font? Are all figures (graphs and images) clearly labeled with identifiable source?
- **Language (4 points):** Is the style of english appropriate for a technical report? Do not treat the project as an assignment and simply number your results like part-exercises. Your text should be a single, coherent whole. The text should be a pleasant read for anyone wanting to find out about the subject matter.
Errors in grammar and orthography (use a spell-checker!) will be penalized. Make sure that the report is interesting to read. Avoid simply repeating sentences by cut-and-paste.
- **Content (12 points):** Are the mathematical and statistical methods and deductions clearly exhibited and easy to follow? Are the conclusions well-supported by the mathematical analysis? It is important to not just copy calculations from elsewhere, but to fully make them your own, adding details and comments where necessary.

All group members will receive the same grade for the term project. (Exceptions are possible in special circumstances.)

On Plagiarism

Study JI's Honor Code carefully. **Any** information from third parties (books, web sites, even conversations) that you use in your project must be accounted for in the bibliography, with a reference in the text. Follow the rules regarding the correct attribution of sources that you have learned in your English course (e.g., Vy100, Vy200). All members of a group are jointly responsible for the correct attribution of all sources in all parts of the project essay, i.e., any plagiarism will be considered a violation of the Honor Code by all group members. Every group member has a duty to confirm the origin of any part of the text.

The following list includes some specific examples of plagiarism:

- Use of any passage of three words or longer from another source without proper attribution. Use of any phrase of three words or more must be enclosed in quotation marks (“example, example, example”). This excludes set phrases (e.g., “and so on”, “it follows that”) and very precise technical terminology (e.g., “without loss of generality”, “reject the null hypothesis”) that cannot be paraphrased,
- Use of material from an uncredited source, making very minor changes (like word order or verb tense) to avoid the three-word rule.
- Inclusion of facts, data, ideas or theories originally thought of by someone else, without giving that person (organization, etc.) credit.
- Paraphrasing of ideas or theories without crediting the original thinker.
- Use of images, computer code and other tools and media without appropriate credit to their creator and in accordance with relevant copyright laws.

Mass Shootings in the United States

This project is based on the article *London murders: a predictable pattern?* by David Spiegelhalter and Arthur Barnett [4]. It discusses the pattern of murders in London between April 2004 and September 2007 based on data of the London Metropolitan Police, obtained from the British Home Office.

Using data obtained from the *Gun Violence Archive* [1] the occurrence of mass shootings in the US can be analyzed in manner analogous to that of Spiegelhalter and Barnett.

- i) Explain what a “mass shooting” and a “mass murder” is. How does the Gun Violence Archive (GVA) define these terms?
- ii) Use Mathematica to re-create a version of Figure 1 in [4] from the Mass Shooting data between January 1st, 2013 and December 31st, 2017 available from the Gun Violence Archive [1].
- iii) From their data, Spiegelhalter and Barnett estimated that the London homicides follow a Poisson distribution with parameter $k = 0.44$. Using the GVA data, test the hypothesis that the occurrence of mass shootings in the US has followed a Poisson distribution in the years 2013-2017. If your null hypothesis is rejected, test the individual years for adherence to a Poisson distribution. Give the P -value of your test. What are your conclusions?
- iv) Does it appear that the occurrence of a mass shooting depends on the weekday or the month? Create charts analogous to those of Figure 3 in [4] using Mathematica.
- v) If you find the the number of shootings depends on the weekday, could that be a reason for the data not to follow a Poisson distribution? Explain! If the data is summarized to become weekly (instead of daily) data, does it then follow a Poisson distribution? Test this! If you reject the null hypothesis, consider further measures, such as considering only individual years.
- vi) Spiegelhalter and Barnett discuss the distribution of the “number of days without any homicides in London”. Unfortunately, there are far too many mass shootings for this to be practical. Instead, find the predicted and actual number of weeks with n shootings, $n = 0, 1, 2, 3, \dots$ and create graphs analogous to Figures 4 a) and b) in [4], the abscissa giving n and the ordinate giving the (expected and observed, respectively) number of weeks with n shootings.

- vii) Confidence intervals for the parameter k of a Poisson distribution have been studied extensively; the survey [3] gives *19 different* expressions! However, in our case we have a very large sample size and it is safe to assume that \bar{X} (the estimator for k) follows a normal distribution. In the spirit of the discussion of confidence intervals for proportions, show that a $(1 - \alpha)100\%$ confidence interval for k is given by

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

and calculate such an interval using the weekly data of the years 2013 to 2017.

- viii) Using the weekly data of January - June 2018, check whether it follows a Poisson distribution and calculate \hat{k}_{2018} . Does this estimate fall into the confidence interval calculated above?
- ix) Prediction intervals for the number of observations are also well-studied (see [2] for example). Derive Nelson's formula [2, (18)], which is also valid under the assumption of large sample sizes and an approximate normal distribution of the estimator for k . Using this formula, obtain 95% prediction intervals for the number of mass shootings in 2018 based on the data for 2013 to 2017. Plot the data 2018 in a single graph along with the prediction intervals, mirroring Figure 5 in [4]. Comment on the result.

References

- [1] Gun Violence Archive. Mass shootings. <http://www.shootingtracker.com>.
- [2] K. Krishnamoorthy and Jie Peng. Improved closed-form prediction intervals for binomial and poisson distributions. *Journal of Statistical Planning and Inference*, 141(5):1709 – 1718, 2011.
- [3] VV Patil and HV Kulkarni. Comparison of confidence intervals for the poisson mean: some new aspects. *REVSTAT-Statistical Journal*, 10(2):211–227, 2012.
- [4] 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].