Ve401 Probabilistic Methods in Engineering

Spring 2020 Term Project 2

Date Due: 12:00 PM (noon), Wednesday, the 30th of April 2020

Group Work

You will be divided into groups of *2–3 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 of **no more than 10 pages** in length. 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*.

Your report should have the appearance, style and contents of a professional report. It should be comprehensible without reference to this document and should be comprehensible by any other student in this course. It is strongly suggested that all members of the project team proof-read the report before submission. **The report should not look like the solution to an assignment.** Do not structure the section titles as "Answer to Question i)" or similarly.

Grading Policy

This term project accounts for 15% of the course grade; it will be scored based on

- **Form (3 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 (3 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 (9 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 generally receive the same grade for the term project. Exceptions are possible in certain circumstances, such as a group member not contributing to the project.

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.

Police Shootings in the United States

This part of the project is based on the article *London murders: a predictable pattern?* by David Spiegelhalter and Arthur Barnett [4]. The article is available for download from the *Resources* section of the Canvas site. It analyzes 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 *Database of Fatal Police Shootings* of the Washington Post [3] the occurence of fatal police shootings in the US can be analyzed in manner analogous to that of Spiegelhalter and Barnett.

i) Summarize the source of the data and characterize how the term "fatal police shooting" is used here.

ii) Use Mathematica to re-create a version of Figure 1 in [4] from the Police Shooting data between January 1st, 2015 and December 31st, 2019 available from https://github.com/washingtonpost/data-police-shootings.

iii) From their data, Spiegelhalter and Barnett estimated that the London homicides follow a Poisson distribution with parameter $k = 0.44$. Using the police shooting data, test the hypothesis that the occurrence of police shootings in the US has followed a Poisson distribution in the years 2015-2019. 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? Recreate tables analogous to Table 1 and Figure 2 in [4] using your data.

iv) Create a table analogous to Figure 3 in [4] using Mathematica. Test whether there is evidence that the average number of police shootings depends on the weekday.

v) Confidence intervals for the parameter k of a Poisson distribution have been studied extensively; the survey [2] gives *19 different* expressions! However, in our case we have a very large sample size and it is safe to assume that $\overline{X}$ (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 data of the years 2015 to 2018.

vi) Using the data of January - April 15th 2020, check whether it follows a Poisson distribution and calculate $\hat{k}_{2020}$. Comment on the result.

vii) Prediction intervals for the number of observations are also well-studied (see [1] for example). Derive Nelson's formula [1, (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 police shootings in 2020 based on the data for 2015 to 2019. Plot the data for 2020 so far in a single graph along with the prediction intervals, mirroring Figure 5 in [4]. Comment on the result.

viii) How has the outbreak of the Coronavirus influenced the data for 2020 so far?

References

[1] K. Krishnamoorthy and J. Peng. Improved closed-form prediction intervals for binomial and Poisson distributions. *Journal of Statistical Planning and Inference*, 141(5):1709 – 1718, 2011. `http://www.sciencedirect.com/science/article/pii/S0378375810005215`.

[2] V. Patil and H. Kulkarni. Comparison of confidence intervals for the Poisson mean: some new aspects. *REVSTAT–Statistical Journal*, 10(2):211–227, 2012.

[3] The Washington Post. Fatal force. `https://www.washingtonpost.com/graphics/national/police-shootings-2016/`. Web. Accessed April 10th, 2019.

[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`.