STATS 551 Homework 0

Prerequisites

Due date: 6:00 pm (EST) Jan. 16, 2018

Review of Probability and Inference (6×10 points).

1. There are 27 lectures in total for STATS 551 (2018, Winter). The instructor flips a fair coin (2 sides, HEAD and TAIL) at the beginning of each lecture. If a 'HEAD' shows up, a quiz is conducted; otherwise not.
   (a) What is the probability of getting a quiz for each lecture?
   (b) What is the distribution of the total number of quizzes during the semester?
   (c) What is the expectation and variance of total number of quizzes during the semester? Can you give a 95% confidence interval?
2. Assume that the instructor flips a biased coin ('HEAD' with probability 0.6) each time, answer the questions above.
3. Assume that the instructor has six coins, with 'HEAD' probabilities equal to $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ respectively. Each time she draws one coin at random and then flips that coin to determine whether a quiz is conducted (HEAD) or not (TAIL). What is the expected number of quizzes during the semester? What is the corresponding variance?
4. Consider the case when the probability of getting a 'HEAD' for each coin flip is sampled independently from a uniform distribution on $[0, 1]$, what is the expectation and variance of the number of quizzes?
5. Can you verify the results in 4 via simulation experiments?
6. Assume that the instructor flips the same coin throughout the semester, but with an unknown probability of 'HEAD', denoted by θ. Out of 27 lectures, 16 of them have quizzes. Derive the maximum likelihood estimator and 95% confidence interval for θ and test the hypothesis "The coin is a fair coin."

Guideline for Submission: submit a hard copy (handwritten or printed).

Exploratory Data Analysis (40 points). Find a data set, preferably the same as the one you use for your final project, and perform any exploratory data analysis you have on mind. Summarize the features of the data set with numbers/tables/figures.

Think about the following questions (without explicitly writing down in your homework submissions). What is interesting or unique about the data? What might be a potential difficult feature of the data? Are you going to use this data set for your final project? Why or why not?

Guideline for Submission: submit R markdown (or jupyter notebook) with annotated code followed by results. Discussions about the results should follow the results.

Prerequisites Self-evaluation. You are allowed to refer to previous textbooks, lecture notes and Internet resources for the homework assignments.

- If you get less than 30 out of 60 points in the 'Review of Probability and Inference' part, you might struggle this course.
- If you get more than 40 points in the 'Review of Probability and Inference' part and more than 20 points in the 'Exploratory Data Analysis' part, you are well prepared for this course.

Optional Reading. Read one of the following papers and post your summary and thoughts on Canvas. Bonus points up to 5 will be rewarded.

1. Induction and deduction in Bayesian data analysis. Andrew Gelman, 2011.
2. Philosophy and the practice of Bayesian statistics (with discussion), A. Gelman and C. Shalizi, British Journal of Mathematical and Statistical Psychology, 66, 8-80.