STAT/Q SCI 403: Introduction to Resampling Method
Spring 2019
Homework 04

Instructions:

- You have to submit all your answers in a single PDF file generated by either LaTeX or *Rmarkdown*.
- You may use the LaTeXtemplate `HW_template.tex` to submit your answer.
- For questions using R, you have to attach your code in the PDF file. If the question ask you to plot something, you need to attach the plot in the PDF as well.
- If the question asks you to show a figure, the clarity of the figure will also be graded.
- The total score of this homework is 8 points.
- Questions with ♠ will be difficult questions.

Questions:

1. In this question, we will study the Poisson distribution, a distribution of a discrete random variable that is determined by a rate parameter λ. If an random variable X follows the Poisson distribution with rate λ, denoted as $\mathsf{Pois}(\lambda)$, then

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \cdots .$$

   It is known that for such a distribution, $\mathbb{E}(X) = \lambda$ and $\mathsf{Var}(X) = \lambda$. Given $X_1, \cdots, X_n \sim \mathsf{Pois}(\lambda)$, the MLE of the rate parameter $\widehat{\lambda}_{\mathsf{MLE}} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$ is the same as the sample mean. Assume we observe 100 random variables $X_1, \cdots, X_{100} \sim \mathsf{Pois}(4)$.

   (a) **(1 pt)** What are the bias and variance of the MLE of the rate parameter λ using these 100 observations.?

   (b) **(0.5 pt; ♠)** Write down a 90% confidence interval of λ using these 100 observations. i.e., the confidence interval will be a function of $X_1, \cdots, X_{100}$.

   (c) **(1 pt)** Use R to generate 100 IID random points from $\mathsf{Pois}(4)$, show the histogram. What is the value of MLE using these 100 points?

(d) **(1 pt)** Use R to run $N = 10,000$ Monte Carlo Simulations of obtain the MLE of a size 100 random sample. Namely, you will obtain $10,000$ realizations of the MLE. These realizations provide you an idea of the distribution of the MLE of a size 100 random sample. Plot the histogram of these $10,000$ realizations. Does the fitted value looks like a normal distribution.

(e) **(1 pt; ♠)** (continue from the previous question) What fraction of the realizations of the MLE are within the interval $[3.5, 4.5]$? Can you come up with an explanation of this?

2. Go to the website `http://www.stat.cmu.edu/~larry/all-of-nonpar/data.html` and download the dataset *Earthquake in Fiji*.

(a) **(0.5 pt)** Show the histogram of the variable `stations`.

(b) **(1 pt)** We now fit an exponential distribution to the variable `stations`. What is the fitted value of the rate parameter? Compare the fitted density curve to the histogram.

(c) **(1 pt)** Fit a linear model with the response variable `Y` being `mag` and the covariate `X` being `stations`. What are the fitted slope? Show the scatter plot and attach the fitted regression line.

(d) **(1 pt)** (continue from the previous question) What is a 95% confidence interval of the fitted slope?

3. **(Bonus question–you don't have to do it; ♠♠♠)** Let $X_1, \cdots, X_n \sim \mathsf{Uni}[0, \theta]$, here θ is an unknown quantity. Find a 90% confidence interval of θ. *(You may earn 1 extra grade point if you successfully solve it. Please send you solution of this question directly to me through an email.)*