STAT/Q SCI 403: Introduction to Resampling Method
Spring 2019
Homework 02

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. Let $X_1, \cdots, X_n$ be IID random points from the Beta distribution $\mathsf{Beta}(\alpha = 2, \beta = 2)$. The PDF of $\mathsf{Beta}(\alpha = 2, \beta = 2)$ is

$$p(x) = 6 \cdot x \cdot (1 - x)$$

for $x \in [0, 1]$ and $p(x) = 0$ outside $[0, 1]$. Let $F(x)$ be the CDF of $\mathsf{Beta}(\alpha = 2, \beta = 2)$. Let $\widehat{F}_n(x)$ be the EDF using $X_1, \cdots, X_n$.

   (a) **(1 pt)** What is the CDF of $\mathsf{Beta}(\alpha = 2, \beta = 2)$? i.e., what is the function $F(x)$?

   (b) **(1 pt)** What is the mean and variance of the EDF $\widehat{F}_n(x)$ for a given $x \in [0, 1]$?

   (c) **(1 pt)** Plot the CDF of $\mathsf{Beta}(\alpha = 2, \beta = 2)$ using R within the range $[0, 1]$.

2. Let U be a uniform random variable over $[0, 1]$. We define another random variable $W = -2 \log U$.

   (a) **(1 pt)** Show that W has the same distribution as $\mathsf{Exp}(0.5)$ by mathematical derivation.

   (b) **(1 pt)** Show that W has the same distribution as $\mathsf{Exp}(0.5)$ by simulating realizations of both random variables for at least 1000 times (i.e., sample size $n = 1000$) and show the two EDFs in the same plot.

3. Use R to generate 5000 data points from $N(2, 2^2)$. Note that $N(2, 2^2)$ is the normal distribution with mean 2 and variance $2^2 = 4$.

   (a) **(0.5 pt)** Plot the EDF curve within $[-1, 5]$.

   (b) **(1.5 pt)** Repeat the above procedure 10 times to generate another 10 EDF curves from the same distribution and same sample size. Plot the new 10 EDF curves within $[-1, 5]$ and attach the actual CDF curve. Does the EDF look like a good estimator of the CDF?

4. **(1 pt)** ♠ Let $X_1, \cdots, X_n$ be an IID random sample from an unknown CDF $F(x)$. Let $\widehat{F}_n(x)$ be the EDF. For a fixed point x_0, explain why the following can be used as a $1 - \alpha$ confidence interval of $F(x_0)$:

$$\widehat{F}_n(x_0) - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\widehat{F}_n(x_0)(1 - \widehat{F}_n(x_0))}{n}}, \quad \widehat{F}_n(x_0) + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\widehat{F}_n(x_0)(1 - \widehat{F}_n(x_0))}{n}},$$

   where z_γ is the γ quantile of $N(0, 1)$. In other words, explain why

$$\widehat{F}_n(x_0) \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\widehat{F}_n(x_0)(1 - \widehat{F}_n(x_0))}{n}}$$

   is a (asympotitc) $1 - \alpha$ confidence interval of $F(x_0)$.