Homework 1
Statistics 201B
Due Oct. 25, 2018

1. Let $X_1, X_2, \ldots \overset{iid}{\sim} Unif(0, \theta)$. Consider the following two estimators of θ:

$$\begin{aligned}\hat{\theta}_n &= \max\{X_1, \ldots, X_n\} \\ \tilde{\theta}_n &= 2\bar{X}_n\end{aligned}$$

   (a) Find the PDF of $\hat{\theta}_n$.

   (b) Find the bias, standard error, and MSE of $\hat{\theta}_n$.

   (c) Find the bias, standard error, and MSE of $\tilde{\theta}_n$.

   (d) Fix $\theta = 1$ and use `R` to make a plot of both MSEs as a function of n. (That is, put two lines on the same plot.) What does the plot tell you about the conditions under which we might prefer $\hat{\theta}_n$ or $\tilde{\theta}_n$?

2. Again let $X_1, X_2, \ldots \overset{iid}{\sim} Unif(0, \theta)$. Consider a confidence interval for θ constructed as $[a\hat{\theta}_n, b\hat{\theta}_n]$, where $\hat{\theta}_n = \max\{X_1, \ldots, X_n\}$. Calculate the coverage of this interval and show that it depends only on a and b. If $a = 1$, what should b be to obtain a coverage of 95%?

3. Let $X_1, \ldots, X_n \sim Bernoulli(p)$ and let $Y_1, \ldots, Y_n \sim Bernoulli(q)$. Find the plug-in estimator and estimated standard error for p. Find an approximate 90% confidence interval for p. Find the plug-in estimator and estimated standard error for $p - q$. Find an approximate 90% confidence interval for $p - q$.

4. Let $X_1, \ldots, X_n \sim F$ and let $\hat{F}_n$ be the empirical distribution function. Let $a < b$ be fixed numbers and define $\theta = T(F) = F(b) - F(a)$. Let $\hat{\theta} = T(\hat{F}_n) = \hat{F}_n(b) - \hat{F}_n(a)$. Find the estimated standard error of $\hat{\theta}$. Find an expression for an approximate $1 - \alpha$ confidence interval for θ.

5. Data on the magnitudes of earthquakes near Fiji are available at `http://www.stat.cmu.edu/~larry/all-of-statistics/=data/fijiquakes.dat`. Download this data and load it into `R` using

```
quakes <- read.table(file = "fijiquakes.dat", header = TRUE)
```

(You may need to change the file argument depending on where on your computer you saved the file.) Type

```
head(quakes)
```

to see the first several lines. This is a special type of object in `R` called a dataframe. You can extract elements from the dataframe using the dollar sign; for example

```
hist(quakes$mag)
```

makes a histogram of the magnitudes. Estimate the CDF $F(x)$ for the magnitudes and plot it. Compute and lines to show a 95% confidence envelope for F using the Dvoretzky-Kiefer-Wolfowitz inequality. Turn in your code and your plot.

6. In 1975, an experiment was conducted to see if cloud seeding produced rainfall. Twenty-six clouds were seeded with silver nitrate and 26 were not. The decision to seed or not was made at random. Download the file `clouds.dat` from bCourse. Let θ be the difference in the median precipitation from the two groups (i.e., median of seeding minus median of non-seeding). Find the plug-in estimate of θ. Using the bootstrap, estimate the standard error of the plug-in estimate (by R) and produce an approximate 95% Normal confidence interval for θ. Turn in your code.

7. Consider approximating $E[\theta]$ when θ is distributed as f by monte carlo integration $1/B\sum_{i=1}^{B}\theta_i$, where $\theta_1, ..., \theta_B$ are iid distributed as f: $P_f(\theta = 1) = 1 - 10^{-4}$ and $P_f(\theta = 10^5) = 10^{-4}$. What is the mean and variance of this estimator?

   Now we consider approximating $E[\theta]$ when θ is distributed as f by importance sampling based monte carlo integration $1/B\sum_{i=1}^{B}\theta_i P_f(\theta_i)/P_g(\theta_i)$, where $\theta_1, ..., \theta_B$ are iid distributed as g: $P_g(\theta = 1) = 1/2$ and $P_g(\theta = 10^5) = 1/2$. What is the mean and variance of this estimator?

   In the context of this example, discuss briefly about the advantages of importance sampling.