Homework 3

Due: 10/17/19

1. *Exit polling.* Recall the exit polling example from class: Define μ to be the probability that a voter will vote for Warren in the 2020 primary. Suppose that we poll n voters each at p different polling places. Let

$$z_{i,j} = \begin{cases} 1 & \text{if the } i\text{-th voter at location } j \text{ voted for Warren} \\ 0 & \text{otherwise} \end{cases}$$

Let $\nu_j = \frac{1}{n}\sum_{i=1}^{n} z_{i,j}$, the sample mean at polling location j.
The probability of obtaining k votes for Warren at a given location is given by the binomial distribution:

$$\mathbb{P}[k|n,\mu] = \binom{n}{k}\mu^k(1-\mu)^{n-k}$$

   (a) Assume the sample size $n = 10$ at each polling location. If all the voters have $\mu = 0.05$ compute the probability that at least one polling location will have $\nu_j = 0$ for the case of $p = 1$, $p = 1000$, and $p = 1,000,000$. Repeat for $\mu = 0.8$.

   (b) For the case $n = 6$ and $p = 2$ with $\mu = 0.5$ for both locations, plot the probability

$$\mathbb{P}[\max_j |\nu_j - \mu| > \epsilon]$$

   for $\epsilon \in [0, 1]$ (the max is over polling locations). On the same plot show the bound that would be obtained using the Hoeffding Inequality. Remember that for a single location, the Hoeffding bound is

$$\mathbb{P}[|\nu - \mu| > \epsilon] \leq 2e^{-2n\epsilon^2}$$

   Hint: if events A and B are independent, $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A]\mathbb{P}[B]$

2. *Airbnb price prediction.* Work through the notebook `https://github.com/ORIE4741/homework/blob/master/airbnb.ipynb`.

   (Yes, you can still use the AirBnB dataset for your project if you wrote your proposal on it. This homework by no means answers all the questions you can ask about this dataset!)