Gov 2002: Problem Set 5

Problem Set Instructions

This problem set is due on **October 25, 11:59 pm** Eastern time. Please upload a PDF of your solutions to Gradescope. We will accept hand-written solutions but we strongly advise you to typeset your answers in Rmarkdown. Please list the names of other students you worked with on this problem set.

Question 1 (25 points)

Suppose we want to model the relationship between legislation and politician quality. There are two types of politician quality: high and low. When a high quality politician propose a bill, it has a probability p_1 to pass; conversely, when a low quality politician propose a bill, it has a probability p_2 to pass, where $p_1 > p_2$. Unfortunately, we cannot directly observe politicians' quality, but instead rely on our prior that a politician is a high type with probability h and low type with probability $1 - h$, where $h \in (0, 1)$. Let X be the number of passed bills after a randomly picked politician has made n proposals.

(a) Find the marginal distribution of X.

(b) Find the mean and variance of X.

Question 2 (25 points)

We know from the definition of the variance that $\mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$. Prove that this equality still holds when we condition on X, i.e., $\mathbb{E}\left[(Y - \mathbb{E}[Y \mid X])^2 \mid X\right] = \mathbb{E}[Y^2 \mid X] - (\mathbb{E}[Y \mid X])^2$

Question 3 (30 points)

Let $X_1 \dots X_n$ be i.i.d. r.v.s with mean μ and variance σ^2, and $n \geq 2$. A bootstrap sample of $X_1 \dots X_n$ is a sample of n r.v.s $X_1^* \dots X_n^*$ formed from the X_j by sampling with replacement with equal probabilities. Let $\bar{X}^*$ denote the sample mean of the bootstrap sample:

$$\bar{X}^* = \frac{1}{n}(X_1^* + \dots X_n^*)$$

(a) Find $\mathbb{E}[X_j^*]$ and $\mathbb{V}[X_j^*]$ for each j. (Hint: What is the distribution of X_j^*?)

(b) Find $\mathbb{E}\left[\bar{X}^* \mid X_1, \ldots, X_n\right]$ and $\mathbb{V}\left[\bar{X}^* \mid X_1, \ldots, X_n\right]$ (Hint: Conditional on $X_1 \ldots X_n$, the X_j^* are independent, with a PMF that puts probability $1/n$ at each of the points $X_1 \ldots X_n$.)

(c) Find $\mathbb{E}\left[\bar{X}^*\right]$ and $\mathbb{V}\left[\bar{X}^*\right]$ (Hint: Recall that the sample variance $\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$ is an unbiased estimator of the population variance σ^2)

Question 4 (20 points)

Jon commutes on the Boston subway from Park Street Station to Harvard Square. He records in minutes every day how long he waits for the train to arrive. He assumes a statistical model that says his waiting times $Y_1 \ldots Y_n$ are i.i.d. from Unif(0, θ).

(a) Find an unbiased plug-in estimator $\hat{\theta}_{PI}$

(b) Find the variance and mean square error of $\hat{\theta}_{PI}$