

Gov 2002: Problem Set 5

Problem Set Instructions

This problem set is due on **March 22, 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 (20 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 (15 points)

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

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 $\mathbb{E}[\bar{X}^* | X_1, \dots, X_n]$ and $\mathbb{V}[\bar{X}^* | X_1, \dots, X_n]$ (Hint: Conditional on $X_1 \dots X_n$, the X_j^* are independent, with a PMF that puts probability $1/n$ at each of the points $X_1 \dots X_n$.)
- (c) Find $\mathbb{E}[\bar{X}^*]$ and $\mathbb{V}[\bar{X}^*]$ (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 $Y_1 \dots Y_n$ are i.i.d. from $\text{Unif}(0, \theta)$.

- (a) Find an unbiased plug-in estimator $\hat{\theta}_{PI}$ for the maximum θ .
- (b) Find the variance and mean square error of $\hat{\theta}_{PI}$.

Question 5 (15 points)

The circus owner is planning to ship his 50 elephants and so he needs a rough estimate of the total weight of the elephants. As weighing an elephant is a cumbersome process, the owner wants to estimate the total weight by weighing just one elephant. Which elephant should he weigh? So the owner looks back on his records and discovers a list of the elephants' weights taken 3 years ago. He finds that 3 years ago, Sambo, the middle-sized elephant was the average (in weight) elephant in the herd. he checks with the elephant trainer who reassures him (the owner) that Sambo may still be considered to be the average elephant in the herd. Therefore, the owner plans to weigh Sambo and take $50y$ (where y is the present weight of Sambo) as an estimate of the total weight $Y = Y_1 + Y_2 + \dots + Y_{50}$ of the 50 elephants. But the circus statistician is horrified when he learns of the owner's purposive sampling plan. 'How can you get an unbiased estimate of Y this way?' protests the statistician. So, together they work out a compromise sampling plan. With the help of a table of random numbers they devise a plan that allots a selection probability of $99/100$ to Sambo, and equal selection probabilities to each of the other 49 elephants. Naturally, Sambo is selected and the owner is happy. 'How are you going to estimate Y ?' asks the statistician. 'Why? The estimate ought to be $50y$ of course,' says the owner. 'Oh! No! That cannot possibly be right,' says the statistician, 'I recently read an article in the *Annals of Mathematical Statistics* where it is proved that the Horvitz-Thompson estimator is the unique hyperadmissible estimator in the class of all generalized polynomial unbiased estimators.' 'What is the Horvitz-Thompson estimate in this case?' asks the owner, duly impressed. 'Since the selection probability of Sambo in our plan was $99/100$,' says the statistician, 'the proper estimate of Y is (i)_____ and not $50y$.' 'And, how would you have estimated Y ,' inquires the incredulous owner, 'if our sampling plan made us select, say the big elephant Jumbo?' 'According to what I understand of the Horvitz-Thompson estimation method,' say the unhappy statistician, 'the proper estimate of Y would then have been (ii)_____, where y is Jumbo's weight.' That is how the statistician lost his circus job (and perhaps became a teacher of statistics!)."

- (a) Show that the Horvitz-Thompson estimator for a finite population is unbiased: $\hat{Y}_{HT} = \sum_{i=1}^N \frac{I_i y_i}{\pi_i} = Y$
- (b) Fill in the blanks in (i) and (ii). Explain why or why not the answers you provided make sense.
- (c) Based on your answers to part (a) and (b), discuss when should (not) one adopt the Horvitz-Thompson estimator.