

# Gov 2001: Problem Set 5

Spring 2026

## Instructions:

- The Problem set is due on **March 3, 11:59 PM Eastern Time.**
- Please upload a PDF of your solutions to Gradescope. Make sure to assign to each question all the pages with your work on that question.
- **Do not use AI assistants (ChatGPT, Claude, Copilot, etc.) on this problem set.** Work with each other instead. The struggle is where learning happens.
- Remember: 70% of your grade comes from in-class exams. Use problem sets to *learn*, not just to get answers.

## Short Questions

1. Let  $X \sim \text{Pois}(\lambda)$ . Find  $\mathbb{E}[X \mid X \geq 1]$  and  $\text{Var}(X \mid X \geq 1)$ .
2. Let  $X_1, X_2$ , and  $Y$  be random variables and  $Y$  has finite variance. Let  $A = \mathbb{E}[Y \mid X_1]$  and  $B = \mathbb{E}[Y \mid X_1, X_2]$ . Show that  $\text{Var}(A) \geq \text{Var}(B)$ . (Hint: Use law of total variance on  $B$  and condition on  $X_2$ )

## Long Questions

3. Let  $X_1, \dots, X_n$  be i.i.d. random variables with mean  $\mu$  and variance  $\sigma^2$ , and  $n \geq 2$ . A bootstrap sample of  $X_1, \dots, X_n$  is a sample of  $n$  random variables  $X_1^*, \dots, X_n^*$  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  $\text{Var}(X_j^*)$  for each  $j$ . (Hint: What is the distribution of  $X_j^*$ ?)
- (b) Find  $\mathbb{E}[\bar{X}^* \mid X_1, \dots, X_n]$  and  $\text{Var}(\bar{X}^* \mid X_1, \dots, X_n)$ . (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  $\text{Var}(\bar{X}^*)$  using the law of total expectation and the law of total variance.
4. Suppose a political scientist wants to use observational data to assess whether a new campaign strategy is more effective than the standard approach. She worries that campaign staff tend to deploy the new strategy to voters that are easier to mobilize (e.g., stronger partisan leanings). If so, a naive comparison of outcomes between those who did and did not receive the campaign would be an “apples-to-oranges” comparison. In order to compare apples to apples, the researcher wants to match every voter who received the campaign to a voter with similar background variables who didn’t. However, since voter characteristics are usually high-dimensional, it’s very difficult to find a match with exact similar backgrounds.

The *propensity score* reduces the possibly high-dimensional vector of background variables down to a single number. Then it is much easier to match someone to a person with a similar propensity score than to match

someone to a person with similar high-dimensional backgrounds. This question explores why this method works.

Let  $X$  denote background characteristics of a voter (possibly a high-dimensional vector). Let  $Z \in \{0, 1\}$  indicate whether the unit received the new strategy. The propensity score of a person with background characteristics  $X$  is defined as

$$S = \mathbb{E}[Z | X] = P(Z = 1 | X).$$

- (a) Prove  $\mathbb{E}[Z|S] = S$ . (Hint: Use LIE with extra conditioning on  $X$ )
- (b) Use the results in (a) to show that conditional on  $S$ ,  $Z$  and  $X$  are independent:  $P(Z = 1|S, X) = P(Z = 1|S)$ . How does this result help the matching process?
- (c) Besides *matching*, another way to use the propensity score in causal inference is *weighting*: we can re-weight observational data to mimic a randomized experiment (this is called inverse probability weighting, IPW).

Let  $m_1(X) = \mathbb{E}[Y | Z = 1, X]$  and  $m_0(X) = \mathbb{E}[Y | Z = 0, X]$  be the CEFs of  $Y$  given  $X$  in the treated ( $Z = 1$ ) and control ( $Z = 0$ ) groups, respectively. We can use the propensity score  $S$  to reweight outcomes:

$$Y_1^* = \frac{ZY}{S}, \quad Y_0^* = \frac{(1-Z)Y}{1-S}.$$

Where  $Y_1^*$  and  $Y_0^*$  are the weighted outcomes in the treatment group (by  $S$ ) and control group (by  $1 - S$ ). Prove that  $\mathbb{E}[Y_1^*] = \mathbb{E}[m_1(X)]$ ,  $\mathbb{E}[Y_0^*] = \mathbb{E}[m_0(X)]$ .

Hint:

$$\mathbb{E}[ZY | X] = \mathbb{E}[ZY | Z = 1, X] \mathbb{P}(Z = 1 | X)$$

and similarly

$$\mathbb{E}[(1-Z)Y | X] = \mathbb{E}[(1-Z)Y | Z = 0, X] \mathbb{P}(Z = 0 | X)$$

5. Suppose  $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bern}(p)$ , where  $p \in (0, 1)$  is the probability of success. The estimand of interest is  $\theta = p$ . Consider the following three estimators of  $\theta$ :

$$\hat{\theta}_1 = X_1, \quad \hat{\theta}_2 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\theta}_3 = c\bar{X}, \quad \text{where } c \in (0, 1) \text{ is a known constant.}$$

- (a) Which one is the plug-in estimator?
- (b) Compute the bias of each estimator. What estimator(s) should we use if we want unbiasedness?
- (c) Compute the variance of each estimator. What estimator(s) should we use if we want its variance becomes smaller when sample size increases?
- (d)  $\hat{\theta}_3$  is usually called a *shrinkage estimator* since it shrinks an unbiased estimator towards zero. Compute the MSE of  $\hat{\theta}_2$  and  $\hat{\theta}_3$  and find the range of  $c$  (as a function of  $c$  and  $n$ ) such that  $\hat{\theta}_3$  has a smaller MSE than  $\hat{\theta}_2$ . Based on these results, when should we consider shrinkage estimators over the unbiased estimator?