Gov 2001: Problem Set 2

Spring 2026

Instructions:

- The Problem set is due on **February 10, 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.

Survey

How many hours (roughly) did you spend on Gov 2001 (other than lectures and sections) in the past week?

Short Questions

1. Let F be the CDF of a continuous r.v., and $f = F'$ be the PDF. Prove that g defined by $g(x) = 2F(x)f(x)$ is also a valid PDF. (Hint: $dF(x) = f(x)dx$)

2. Let X be a random variable with PDF

$$f(x) = \frac{1}{2\sqrt{x}}\mathbf{1}\{x \in (0, 1)\}.$$

Explain how it is possible for f to be a valid PDF even though $\lim_{x\to 0} f(x) = \infty$. Then find the support and CDF of X.

3. Let X be a discrete random variable with support $\{x_1, \ldots, x_n\}$ and let $\mathbb{P}(X = x_k) = p_k$ where $p_k > 0$ and $\sum_{k=1}^n p_k = 1$. The entropy of X is

$$H(X) = \sum_{k=1}^{n} p_k \log_2\left(\frac{1}{p_k}\right).$$

Prove that $H(X) \leq \log_2 n$. (Hint: Let W be an r.v. taking value $\frac{1}{p_k}$ with probability p_k and use Jensen)

4. A random binary sequence of length n is generated, with $n \geq 3$. Each bit (binary digit) is 1 with probability p and 0 with probability $q = 1 - p$, independently. Let X be the number of occurrences of the pattern 110. For example, 11111000011010001101111 has 3 occurrences of 110. Find $\mathbb{E}(X)$.

Long Questions

5. In the United States, approximately 29% of white drivers are stopped by police, compared to approximately 42% of non-white drivers. Among white drivers who are stopped, 25% are found to possess illegal contraband, while 28% of stopped non-white drivers possess illegal contraband.

Let C denote the event that a driver possesses contraband, W the event that the driver is white, and S the event that the driver is stopped by the police. Suppose that the probability of contraband among *non-stopped* drivers is the same across racial groups:

$$P(C \mid S^c, W) = P(C \mid S^c, W^c).$$

(a) What values of the probability of contraband **among non-stopped drivers** would imply the probability of contraband among whites is higher than contraband among non-whites **in general**?

(b) Suppose you are asked to assess whether there is racial bias in police stops. Consider the following measure to quantify racial discrimination:

$$D \;=\; P(S \mid C, W^c) - P(S \mid C, W).$$

The motivation for this measure is that, in the absence of racial bias, $S \perp W \mid C$. This would mean given that if the driver is actually carrying contraband, their race should not update the probability of a police choosing to stop them. Assuming that contraband rates in the population are independent of race, derive the set of feasible values of D using the information provided in this problem.

Hint: Let

$$x = P(C \mid S^c, W) = P(C \mid S^c, W^c),$$

and use Bayes' rule to express D as a function of x.

6. Alice is conducting a survey in a town with population size 1000. She selects a simple random sample of size 100 (i.e., sampling without replacement, such that all samples of size 100 are equally likely). Bob is also conducting a survey in this town. Bob selects a simple random sample of size 20, independent of Alice's sample. Let A be the set of people in Alice's sample and B be the set of people in Bob's sample.

(a) Find the expected number of people in $A \cap B$.

(b) The 1000 people consist of 500 married couples. Find the expected number of couples such that both members of the couple are in Bob's sample.

Hint: Indicator r.v.s

7. It's usually computationally complex and intensive to use LOTUS to calculate moments of distributions. This question studies a simpler way to calculate moments through a specific family of functions: Moment Generating Functions (MGF).

Let X be a discrete random variable with support $\{1, 2\}$ and let $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 2) = q = 1 - p$.

(a) Calculate $\mathbb{E}[X]$, $\mathbb{E}[X^2]$.

(b) Let $M(t) = \mathbb{E}[e^{tX}]$. Calculate $M(t),\ M'(t),\ M''(t)$. Verify that $M'(0) = \mathbb{E}[X],\ M''(0) = \mathbb{E}[X^2]$.

Let c be a constant and Y be a continuous random variable with PDF

$$f(y) = \begin{cases} ce^{-2y}, & y > 0, \\ 0, & \text{otherwise}, \end{cases}$$

(c) Find the constant c that make $f(y)$ a valid PDF.

(d) Use the general result that $M^{(n)}(0) = \mathbb{E}[Y^n]$ to calculate $\mathbb{E}[Y^3]$.