Gov 2001: Problem Set 2

Spring 2025

February 19, 2025

Problem Set Instructions:

- This problem set is due on **Feb 26, 11:59 pm** Eastern time. Please upload a PDF of your solutions to **Gradescope**.
- We will accept hand-written solutions but we strongly advise graduate students to typeset your answers in LaTeX.
- Citing your sources is always a good practice in academia. Please list the names of other students / sources / AI you obtained help from on this problem set.

1 Variance

1.1 Alternative expression (10 pts)

Prove that we can write variance as below:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

1.2 Counter example (5 pts)

Show an example where $V[X + Y] = V[X] + V[Y]$ does not hold, and explain why.

2 Ranking Continuous R.V.s

(10 pts) Let $X_1, \ldots, X_n$ be i.i.d. from a continuous distribution. For any permutation $a_1, a_2, \ldots, a_n$ of $1, 2, \ldots, n$, calculate the following probability:

$$\mathbb{P}(X_{a_1} < X_{a_2} < \cdots < X_{a_n})$$

2.1 Discrete R.V. (5 pts)

Would your result hold for discrete R.V.s? Why?

3 Binomial Distribution

A common approximation for a Binomial random variable $X \sim \text{Bin}(n, p)$ is a Poisson random variable $Y \sim \text{Pois}(\lambda)$ where $\lambda = n \cdot p$. One of the reasons for this approximation is that the expectations of Binomial and Poisson random variables match (both np, in our case). We will explore this a bit further.

(a) Holding n fixed, would Y better approximate X if $p = \frac{1}{10}$ or if $p = \frac{1}{2}$? Compare the variances of X and Y by checking the difference and ratio of the two variances. (10 pts)

(b) Holding p fixed, would Y better approximate X if $n = 1{,}000$ or if $n = 100{,}000$? Compare the variances of X and Y by checking the difference and ratio of the two variances. (10 pts)

4 Expected Value

Suppose you're interested in studying the distribution of political ideology in the US, a random variable that we'll call X. Individuals are placed on a continuous one-dimensional ideology scale that varies from -1 to 1, where lower score are more liberal. Instead of having to do sampling to estimate the distribution, the Data Generating God comes to you in a dream and tells you the unnormalized distribution of this random variable, which is as follows:

$$f(x) = \begin{cases} c(\frac{1}{2}x + 1) & -1 \leq x \leq 1, \\ 0 & \text{else} \end{cases}$$

where c is a normalizing constant[1].

(a) Find the value of c that would make $f(x)$ a valid probability distribution function, $f_X(x)$. (20 pts)

(b) Calculate $E[X]$. (10 pts)

5 Count Data

Suppose $X \sim \text{Pois}(\lambda)$, where λ is fixed but unknown.

An estimator is a function of the data and the **bias** of an estimator, $f(X)$, is defined as $E[f(X)] - \theta$, where θ is the **estimand** (an unknown quantity we would like to estimate from the observable data).

For instance our estimand could be λ, and we know by the properties of a Poisson random variable that the bias of the estimator, $f(X) = X$, is $E[X] - \lambda = \lambda - \lambda = 0$. We call an estimator with 0 bias an **unbiased** estimator.

For this question, suppose that our estimand is λ^3 rather than λ.

(a) Show that X^3 is **not** an unbiased estimator of λ^3 and specify the bias as a function of λ.(15 pts)

Hints:

[1]Note that this distribution is completely made up and is not based on actual data.

1. You may use the following result: if $X \sim \text{Pois}(\lambda)$, then $E[X \cdot g(X)] = \lambda E[g(X+1)]$ for any function $g(\cdot)$.

2. You may use the result for $E[X^2]$ derived in lecture and section, (i.e., no need to derive it again).

(b) Suppose $\lambda = 5$. Use $150,000$ simulations to validate your result to part (a). That is, calculate the bias of you estimator from both the simulations and the analytical results. Let's use CGIS Zipcode for seed. (5 pts)

```
set.seed(02138)
```