

Gov 2002: Problem Set 4

Spring 2021

Problem Set Instructions

This problem set is due on February 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 (20 points)

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 .

Hints:

1. In section 3, we proved the identity that, if $X \sim \text{Pois}(\lambda)$, then $E[X \cdot g(X)] = \lambda E[g(X+1)]$ for any function $g(\cdot)$. You can make use of this result here.
 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. Print your results in the format below:

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

```
# Simulated result
```

Question 2 (20 points)

Suppose you are studying political ads with a co-author and you have a series of ads in video form. Let the timestamps of the videos be normalized to $(0,1)$ so that 0 represents the start of the video and 1 represents the end. In a misguided attempt to save hard drive space, your co-author has decided to break up all the videos into two clips by splitting them at a proportion of the video given by a uniform random variable $U \sim Unif(0,1)$. Your task is to watch every clip that contains the timestamp p .

- What is the probability that the first clip contains p ?
- What is the expected length of the clip that does contain p ? HINT: you do not need to derive the p.d.f. or c.d.f. of this random variable to calculate this expectation.

Question 3 (30 points)

Consider the following scenario similar to one from a few weeks ago:

You are trying to schedule qualitative interviews with government officials in some country. This time, the probability of any official responding to you is p and, again, you don't have a list with a fixed length of officials ahead of time (i.e. you're just finding emails in real-time). Let $q = 1 - p$. Assume that the officials' emails to you are independent and that $0 < p < 1$. Let X be the number of emails you send until you get one response email back (where the count does not include the email that received a response).

- What is the probability that $P(X = k)$ for $k \geq 0$?
- Let $g(X) = (1 - p)^X$. Find $E[g(X)]$. Express your final answer only in terms of p . Hint: the sum of a **geometric series** of the form $\{a, ar, ar^2, ar^3, ar^4, \dots\}$ for two constants a and r is $\frac{a}{1-r}$.
- Find the CDF F of X , being sure to specify $F(x)$ for all real x .

Hint: Start by considering the case where x is a nonnegative integer. In extending to the case that x is a nonnegative real number, you can express your answer in terms of the floor function: let $\lfloor x \rfloor$ denote the greatest integer less than or equal to x .

Question 4: Fisher's method in forecasting (30 points)

In this problem, we're going to explore a real-world example of Fisher's "lady tasting tea" experiment from lecture: election forecasters – who have, for better or worse, become a big part of politics in the United States and elsewhere. Before running your first R simulation, set your seed: `set.seed(02138)`.

- Suppose that Bob has correctly predicted six of the last eight election outcomes. What is the probability that someone randomly flipping a coin each of the same elections

would have experienced the same success as Bob? Use R to compute your answer.

Forecasting has become so popular that riffraff are flooding the market. These “uniform amateurs” predict the vote share for each state in the U.S. presidential election by drawing a uniform random variable between 0 and 1, independently across states. You are deciding whether or not to hire a forecaster, Nate, to forecast each of the 50 state election winners in the 2024 presidential general election based on the performance of his 2020 election forecast, but you are worried that Nate might be one of these amateurs. When you ask him to justify his 2020 forecasts, he says “my highest predicted vote share was 0.8 which is very unlikely if I were a uniform amateur.” Let’s evaluate his claim.

- (b) Suppose Nate is a uniform amateur and let X be the maximum of the 51 uniform vote share draws (include D.C.). Derive the CDF and PDF of X . Use these to calculate the probability of Nate’s highest vote share being 0.8 or less if he were a uniform amateur.
- (c) To be on the lookout for more uniform amateurs, it’s helpful to know what highest vote share we should expect. To that end, calculate $E[X]$.