

Gov 2002: Problem Set 1

Spring 2023

Problem Set Instructions

This problem set is due on **February 8, 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 (12 points)

- (a) You are trying to schedule qualitative interviews with government officials in some country. 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). The officials can take one of three actions: respond and agree to meet with you with probability $1/3$, respond and reject your request with probability $1/3$, or ignore your email with probability $1/3$. The officials make their decisions independent of each other. Suppose you keep sending out emails until you get one official to agree to meet with you. Find the PMF of the number of emails you send out.
- (b) Now suppose you care about whether an official has responded or not, regardless of the response. What is the PMF of the number of emails you have to send out until there's at least one official who responds to you and one official who doesn't? For example, the first five officials might accept, reject, reject, reject, and ghost you, after which you would stop since you have at least one response and non-response. You might also see the sequence ghost, ghost, ghost, ghost, then reject, after which you stop.

Question 2 (18 points)

Some political scientists believe that voter turnout is low in American elections because most people are not interested in politics. Let C be the event that a person is civically engaged, V_1 be the event that they voted in the previous election and V_2 be the event that they vote in the upcoming election.

For convenience, let $P(C) = c$ and $P(V_1|C) = P(V_2|C) = p_1$ and $P(V_1|C^c) = P(V_2|C^c) = p_2$ where $p_1 > p_2$. Suppose that $V_1 \perp V_2 \mid C$ and $V_1 \perp V_2 \mid C^c$, meaning that given a person's interest (or lack thereof) in politics, their choice to turn out across years is independent.

- a. Do you think V_1 and V_2 would be unconditionally independent? Why or why not?

- b. What is the probability that a citizen is civically engaged if they did not vote in the previous election?
- c. What is the probability that a citizen will vote in next election cycle given that they didn't vote in the last election?

Question 3 (20 points)

In the United States, roughly 29% of white drivers get stopped by police compared to roughly 42% of non-white drivers. Of white drivers who are stopped by police, 25% have illegal contraband, while 28% of stopped non-white drivers have illegal contraband.¹

Let C be the event of a driver possessing contraband, W be the event of the driver being white, and S being the event of the driver getting stopped by the police. Suppose that the probability of contraband found among non-stopped drivers is equal across both racial groups.

- (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 find whether there is (and if there is, how much) racial bias in who is stopped by the police. You use the following measure to quantify racial discrimination: $P(S|C, W^c) - P(S|C, W)$. The reasoning for this measure is as follows: if there is no racial bias in police stops, we might expect that $S \perp W \mid C$. 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. To show that this false – that race does update the probability of a stop – we need to simply show that this measure is equal to zero. In English, it can be interpreted as, “the increase in probability of getting stopped if a contraband carrier is not white”.

Assuming that the rates of contraband in the population are independent of race, plot and interpret possible bounds for this measure using the information provided in this problem. Hint: use Bayes' Rule and compute the bounds using R.² Hint part 2: let $x = P(C|W^c, S^c) = P(C|W, S^c)$.

Question 4

- (a.) Suppose that, in advance of the 2024 presidential election, you know that Pennsylvania and Georgia are pure toss-ups between the Democratic and Republican candidates and are independent of each other. Let X be the number of states the Republican candidate wins of the two. What are the PMF and CDF of X ?
- (b.) Now suppose that you are interested in local elections. Two counties are looking to elect a sheriff, but these elections are not toss-ups - in one, the Republican candidate has a 65% chance of winning, while in the other, the Republican candidate has a 40% chance of winning.

¹These are approximate figures.

²For more on statistical fallacies on estimating racial disparities in policing, see “Administrative Records Mask Racially Biased Policing” (Knox, Dean, Will Lowe, and Jonathan Mummolo).

Again, the two sheriff's races are independent from each other. Let Y be the number of elections the Republican candidate wins of the two. What are the PMF and CDF of Y ?

(c.) Finally suppose that you know the joint PMF of X and Y , $P(X = x, Y = y)$, to be as follows:

| | $Y = 0$ | $Y = 1$ | $Y = 2$ |
|---------|---------|---------|---------|
| $X = 0$ | .0525 | .15 | .065 |
| $X = 1$ | .105 | .265 | .13 |
| $X = 2$ | .0525 | .115 | .065 |

Are X and Y independent?

Question 5 (26 points)

- (a) Let $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$, independent of X . Show that $X - Y$ is *not* binomial. An informal explanation is acceptable here.
- (b) Now let U_1 and U_2 be independent rolls of dice - that is, they follow independent discrete uniform distributions over $\{1, 2, \dots, 6\}$. Show that the random variable $Z = U_1 + U_2$ is *not* uniform. An informal explanation is acceptable here too.