

Gov 2002: Problem Set 2

Spring 2021

Problem Set Instructions

This problem set is due on February 10, 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)

- You are trying to schedule qualitative interviews with government officials in some country. The probability of any official responding to you is essentially a coin flip and that 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). Suppose you keep sending out emails until you get one response email back. Find the PMF of the number of emails you send out.
- 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?

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 she voted in the previous election and V_2 be the event that she votes 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.

- Do you think V_1 and V_2 would be unconditionally independent? Why or why not?
- What is the probability that a citizen is civically engaged if they did not vote in the previous election?
- 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.) Is the probability of contraband amongst whites greater than amongst non-whites? If not, explain intuitively under which conditions it would be.

(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 brainstorm the four following measures:

- $P(S|W^c) - P(S|W)$
- $P(W^c|S) - P(W|S)$
- $P(S|C, W^c) - P(S|C, W)$
- $P(S|C^c, W^c) - P(S|C^c, W)$

Explain intuitively which one of these might be the best racial bias measure. Calculate and interpret the upper and lower bounds for your chosen measure given the information provided in this problem.

Hint: use Baye's Rule and compute the bounds using R.²

Question 4: Simpson's Paradox (24 points)

- (a) Is it possible to have events A, B, E such that $P(A|E) < P(B|E)$ and $P(A|E^c) < P(B|E^c)$, yet $P(A) > P(B)$? That is, A is less likely under B given that E is true, and also given that E is false, yet A is more likely than B if given no information about E . Show this is impossible (with a short proof) or find a counterexample (with a "story" interpreting A, B, E). Hint: Use the LTP.
- (b) Simpson's paradox is a phenomenon where, for events A, B, E , $P(A|B, E) < P(A|B^c, E)$ and $P(A|B, E^c) < P(A|B^c, E^c)$, yet $P(A|B) > P(A|B^c)$.

For instance:

- Let A be the event that a senator voted to pass the Civil Rights Act of 1964
- Let B be the event that the senator is Republican
- Let E the event that the senator represents a Southern state.

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).

The proportions voting in favour of Act were as follows:

Southern Republicans less likely to support Act than Southern Democrats:

$$0 = P(A|B, E) < P(A|B^c, E) = 0.07$$

Northern Republicans less likely to support Act than Northern Democrats:

$$0.85 = P(A|B, E^c) < P(A|B^c, E^c) = 0.94$$

But Republicans more likely to support Act than Democrats:

$$0.80 = P(A|B) > P(A|B^c) = 0.61$$

Show that this paradox is *not* possible if E is independent of B and B^c . In this example, this would imply that the probability of being a Southern senator is independent of the Senators' party. Hint: apply the law of total probability.

Question 5 (26 points)

Let $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$, independent of X .

- Show that $X + Y \sim \text{Bin}(n + m, p)$. Intuitive answers are acceptable.
- Show that $X - Y$ is *not* binomial. An informal explanation is acceptable here too.
- Why is $2X \sim \text{Bin}(2n, p)$ not correct?