

Gov 2002: Problem Set 2

Fall 2023

Problem Set Instructions

This problem set is due on **September 20, 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)

Suppose you are a junior infectious disease physician working in a hospital. The annual flu season has started, and you know that approximately 40% of your patients have seasonal flu. You have a rapid diagnostic test kit with the following diagnostic properties:

- Among patients with seasonal flu, 95% get tested positive with this test kit.
- Among patients without seasonal flu, 45% get tested negative with this test kit.

Using these properties, answer the following questions.

- (a) A patient of yours tested positive with this rapid test kit. What is the probability that the patient actually has the seasonal flu?
- (b) After seeing several patients, you realize that you can do a better job roughly estimating the probability of your patient having seasonal flu before testing. Suppose that, after listening to a patient’s history, you find that one of your patients is extremely unlikely to have the flu: Your rough estimate (the “pre-test probability”) is that the probability of this patient having seasonal flu is 0.1. What is the probability of this patient having seasonal flu if they test positive (the “post-test probability”)?
- (c) Plot the post-test probability as a function of the pre-test probability, assuming that all of your patients receive a positive test result.
- (d) Upon seeing your plot from question (c), your attending physician told you that you don’t have to, and perhaps shouldn’t, perform rapid tests in some cases. Is this true? If so, when should you not perform rapid tests?

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 $x = P(C|W^c, S^c) = P(C|W, S^c)$.

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).

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. 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) You are a player on a basketball team in your town. Your team has 10 players in total. In the coming season, there will be n games in total, and each game is considered as independent. The probability of winning one game is 0.6. If you win a game, your team gains 5,000 dollars as a team. The team distributes the money evenly among 10 players. What is the PMF of the amount of money each *team member* can gain in the coming season?
- (b) Whereas the coming season has n games, the subsequent season will have m games. Again, each game is played independently, and the probability of winning one game is 0.6. What is the PMF of the number of games that the team wins in the coming two seasons combined?
- (c) Let $W_1 \sim \text{Bin}(n, 0.6)$ and $W_2 \sim \text{Bin}(m, 0.6)$, where W_1 is the number of games that your team wins in the first season, and W_2 is the number of games that your team wins in the second season. You can think of W_1 and W_2 as independent of each other. You wonder if the numbers of games your team wins in the two seasons are different and take the difference between the two: $W_1 - W_2$. A friend of yours who is interested in statistics tells you that $W_1 - W_2$ is a Binomial random variable since it is the difference between two Binomial random variables. Is it true? Explain.