Gov 2001: Problem Set 1

Spring 2025

February 3rd, 2025

Problem Set Instructions:

- This problem set is due on **Feb 12, 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 Birthday Problem (20 points)

Let's re-consider the birthday problem we discussed during lecture.

Birthday Problem: There are 20 people in a room. Assume each person's birthday is equally likely to be any of the 365 days of the year (no leap babies) and birthdays are independent. What is the probability that at least one pair of people have the same birthday?

1.1 Approach 1 (10 points)

Repeat what we did in class by taking the complement of the event we are interested in. Then, calculate the probability in the birthday problem.

1.2 Approach 2 (10 points)

Instead of the smarter approach we did in class, can you solve the problem other way around:

Due to the complexity of the problem, can you list 5 examples of these smaller events and write down their probabilities?

Why can we naively add these probabilities up?

2 Union of Two Events (10 points)

Recall our discussion during class: Even if B and C are disjoint,

$$\mathbb{P}(A \mid B \cup C) \neq \mathbb{P}(A \mid B) + \mathbb{P}(A \mid C).$$

Can you give us an example to illustrate this idea? You can also use visualization to help with your point. For those who are typesetting the pset, `tikz` package is your friend!

3 Peace Between US and UK (10 points)

You and your coauthor are trying to model the probability over how long the peace between the United States and United Kingdom will last, where the sample space is the set of all positive years ($\Omega = \{1, 2, 3, \ldots\}$). You and your coauthor agree that this sample space is countably infinite, but your coauthor insists that you should just model every year as equally likely. Using the axioms of probability, show your coauthor that each year in the sample space cannot be equally likely.

4 Survey Study (20 points)

Suppose you are conducting a panel study over two waves, three months apart. In the first wave, you sample without replacement n respondents from your pool of N potential panelists. In the second wave, you take a sample of size m without replacement from the same pool. Our goal will be to obtain the probability that exactly k of the m respondents in the second wave were also in the first wave. We'll do this in a few steps.

You should assume that being a respondent in one wave has no effect on being selected in another wave and that (for now) all those selected participate.

1. (5pts) What are the total number of ways to select m respondents from the pool in the second wave?

2. (10pts) How many different ways are there to select m respondents in the second wave such that exactly k are from the n selected in the first wave?

3. (5pts) Assuming everyone is equally likely to be selected, what is the probability that exactly k of the m respondents in the second wave were also selected in the first wave?

5 Independence of Data (20 points)

1. (5pts) 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 anf CDF of X?

2. (5pts) 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 by the number of elections the Republican candidate wins of the two. What are the PMF and CDF of Y?

3. (10pts) 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? Why?

6 Stopping a Driver (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.

1. (5 pts) 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*?

2. (15 pts) 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 <u>does</u> update the probability of a stop – we need to simply show that this measure is not equal to zero.

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

[1] These are approximate figures.

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

7 Challenge Question (Optional, 10 points)

Prove the following property:

$$\sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k} = \binom{m+n}{r}$$