

Homework_04

February 9, 2024

Probability for Data Science

UC Berkeley, Spring 2024

Shobhana Stoyanov

CC BY-NC 4.0

This content is protected and may not be shared, uploaded, or distributed.

```
[ ]: from datascience import *  
import numpy as np  
from scipy import stats
```

1 Homework 4

1.0.1 Instructions

Show your work. Give reasoning. The question isn't always going to ask for it, because we assume that you will provide justification for your answers (you won't get credit without it). Every answer should contain a calculation, reasoning, or diagrams that are clearly labeled to show what's going on.

We encourage you to hand-write your answers. Start each question on a separate page. It is your responsibility to ensure that the homework is submitted completely and properly to Gradescope. **Make sure to assign each page of your pdf to the correct question.** Refer to the bottom of the notebook for submission instructions.

In some of the problems below, you have to do a numerical calculation in a code cell. All the libraries you might need have been imported for you. Run the import cell at the top before you start the homework.

1.1 1. Monkey Dialing 911

Route, a baby Capuchin monkey in Southern California, [managed to call 911](#) on a cell phone. Nobody is quite sure how Route did this, but let's see what the probabilities look like if we make up some assumptions.

- a) Suppose a cell phone screen has Call button as well as a grid of 12 buttons consisting of
- one each for the 10 digits 0 through 9,
 - the * symbol, and

- the # symbol.

Suppose that each time a monkey presses a button on this screen, it has an 80% chance of picking the brightly colored Call button and a 20% chance of picking at a button at random from the other 12 buttons, independent of its choices at all other times.

What is the chance that the first four buttons the monkey presses are 9, 1, 1, Call, in that order? Write the arithmetic expression on paper (not in code) and then provide its numerical value. You can use the code cell below, or a calculator, or any such computational device. We will not grade your code.

```
[ ]: # Calculation for 1a
...

```

1.2 2. Thoughtful Politicians

[This problem setting is due to [Prof. James Norris](#) of Cambridge University.]

In a group of 100 politicians, 40 are in Party I and 60 are in Party II. Members of Party I are incapable of changing their minds about anything. Members of Party II change their minds completely randomly every day, independently of all others and also independently of all evidence.

Tomorrow, the politicians are going to vote on Proposition 88. Here is a summary of their opinions today.

Of the 40 members in Party I, 25 are in favor of Prop 88 and 15 are against. Of the 60 members in Party II, 10 are in favor and 50 are against. Assume that each member in Party II changes their mind based on the toss of a coin: if it lands heads they switch to the opposite opinion and if it lands tails they don't change their mind.

a) What is the distribution of the number of members in Party II who will favor Prop 88 tomorrow? In what way, if any, does it depend on the information given about the opinions of Party II members today?

b) Write a math expression for the chance that a majority (more than 50) of the 100 politicians favor Prop 88 tomorrow. Use the code cell below to compute it and provide the final numerical answer on paper along with your math expression.

```
[ ]: # Calculation for 1b
...

```

1.3 3. Treatment and Control

A randomized controlled experiment has 50 participants, of whom 20 are women. A simple random sample of 25 participants is assigned to the treatment group and the remainder to the control group.

a) Say whether the following statement is true or false. If it is true, provide a math expression for the chance. If it is false, provide math expressions for the two chances.

$P(\text{the treatment group has 16 women}) = P(\text{the control group has 16 women})$

b) Say whether the following statement is true or false, and justify your answer.

The event “the treatment group has 16 women” is independent of the event “the control group has 16 women”.

c) Write a math expression for the chance that there are at least six women in each group.

[*Hint:* This can be done by just thinking about the women in one of the groups. Any other way will prove quite a bit harder.]

1.4 4. Using a CDF

The figure below shows the graph of a function F which is the cumulative distribution function (cdf) of a random variable X .

a) Make a distribution table for X .

b) Find $P(|X - 1| \leq 1)$.

c) Find $P((X - 1)(X - 3) = 0)$.

1.5 5. Can this Problem Die

A die is rolled 24 times. Let X be the number of fives in the first 16 rolls and let Y be the number of fives in all 24 rolls.

- a) What distribution does X follow?
- b) What distribution does Y follow?
- c) For what values of x is $P(X = x, Y = 4)$ nonzero?
- d) For the values of x in the range you found in part (c), find $P(X = x, Y = 4)$.
- e) For the values of x in the range you found in part (c), find $P(X = x|Y = 4)$.

1.6 Submission Instructions

Please follow the directions below to properly submit your work.

1.6.1 Written Work

- Scan all the pages into a PDF. You can use any scanner or a phone using applications such as CamScanner. Please **DO NOT** simply take pictures using your phone.
- Please start a new page for each question. If you have already written multiple questions on the same page, you can crop the image in CamScanner or fold your page over (the old-fashioned way). This helps expedite grading.
- It is your responsibility to check that all the work on all the scanned pages is legible.

1.6.2 Submitting

- Submit the assignment to Homework 4 on Gradescope.
- **Make sure to assign each page of your pdf to the correct question.**
- **It is your responsibility to verify that all of your work shows up in your final PDF submission.**

If you have questions about scanning or uploading your work, please post a follow-up to the Ed thread on [Gradescope Assignment Submission](#).