

Chris Piech  
CS 109

Problem Set #5  
Nov 7<sup>th</sup>, 2018

## Problem Set #5

### Due: 1pm on Friday, Nov 16th

With problems by Mehran Sahami and Chris Piech

**For each problem, briefly explain/justify how you obtained your answer.** In fact, most of the credit for each problem will be given for the derivation/model used as opposed to the final answer. Make sure to describe the distribution and parameter values you used (e.g.,  $\text{Bin}(10, 0.3)$ ), where appropriate. Provide a numeric answer for all questions when possible.

### Warmup

1. You are developing medicine that sometimes has a desired effect, and sometimes does not. With FDA approval, you are allowed to test your medicine on 9 patients. You observe that 7 have the desired outcome. Your belief as to the probability of the medicine having an effect before running any experiments was  $\text{Beta}(2, 2)$ .
  - a. What is the distribution for your belief of the probability of the medicine being effective after the trial?
  - b. Use your distribution from (a) to calculate your confidence that: the probability of the drug having effect is greater than 0.5. You may use `scipy.stats` or an online calculator.
2. (Coding) Let  $X$  be the sum of 100 independent uniform random variables each of which are identically distributed as  $\text{Uniform}(0, 1)$ . Simulate 100,000 calculations of  $X$ .
  - a. Use the simulations to calculate the probability that  $X$ , rounded to the ones digit, is 30, 31, 32 and so on up until 60. Draw a bar graph of your results.
  - b. Use the Central Limit Theorem to come up with a distribution for  $X$ .
  - c. Use your answer to part (b) to calculate the probability that  $X$  is in the range 47.5 to 48.5. Make sure that your answer aligns with the result you reported in part (a). Round your result to two decimal places.
3. An amateur university band passes around a pot for donations after a concert. There are 50 people in the audience. Each person gives money independently with the same distribution (IID). Each individual has a:
  - 0.10 probability that they give \$0
  - 0.20 probability that they give \$1
  - 0.35 probability that they give \$5
  - 0.30 probability that they give \$10
  - 0.05 probability that they give \$20
  - a. What is the expected amount of money that each person gives?
  - b. What is the variance of the amount of money that each person gives?
  - c. Give a probability distribution that approximates the total amount of money that the band earns.
  - d. What is the approximate probability that the band makes at least \$350?

4. Let  $X$  be the outcome of a dice roll. Let  $Y = X^2$ . What is the covariance between  $X$  and  $Y$ ?
5. Let  $X \sim N(\mu = 1, \sigma^2 = 2)$  and  $Y \sim N(\mu = 1, \sigma^2 = 2)$ . What is the distribution for  $2X + Y$ ?
6. You roll 6 dice. How much more likely is a roll with: [one 1, one 2, one 3, one 4, one 5, one 6] than a roll with six 6s? Think of your dice roll as a multinomial.

### Probabilistic Program Analysis:

7. Consider the following function, which simulates repeatedly rolling a 6-sided die (where each integer value from 1 to 6 is equally likely to be "rolled") until a value  $\geq 3$  is "rolled".

```
int roll() {
    int total = 0;
    while (true) {
        // equally likely to return 1,...,6
        int roll = randomInteger(1, 6);
        total += roll;
        // exit condition
        if (roll >= 3) break;
    }
    return total;
}
```

- a. Let  $X$  be the value returned by the function `roll()`. What is  $E[X]$ ?
  - b. Let  $Y$  be the number of times that the die is "rolled" (i.e., the number of times that `randomInteger(1, 6)` is called) in the function `roll()`. What is  $E[Y]$ ?
8. In class, we considered the following recursive function:

```
int Recurse() {
    // equally likely to return 1, 2, or 3
    int x = randomInteger(1, 3);
    if (x == 1) return 3;
    else if (x == 2) return (5 + Recurse());
    else return (7 + Recurse());
}
```

Let  $Y$  be the value returned by `Recurse()`. We found that  $E[Y] = 15$ . What is  $\text{Var}(Y)$ ?

9. Program A will run 20 algorithms in sequence, with the running time for each algorithm being independent random variables with mean = 50 seconds and variance = 100 seconds<sup>2</sup>. Program B will run 20 algorithms in sequence, with the running time for each algorithm being independent random variables with mean = 52 seconds and variance = 200 seconds<sup>2</sup>.

- a. What is the approximate probability that Program A completes in less than 950 seconds?
- b. What is the approximate probability that Program B completes in less than 950 seconds?
- c. What is the approximate probability that Program A completes in less time than B?

## A/B Testing

In this question you are going to learn how to calculate p-values for experiments that are called "a/b tests". These experiments are ubiquitous. They are a staple of both scientific experiments and user interaction design.

Massive online classes have allowed for distributed experimentation into what practices optimize students learning - and promise to be able to scale more personalized educational experiences. Coursera, a free online education platform that started at Stanford, is testing out a set of ways of teaching a concept in probability. They have two different learning activities activity1 and activity2 and they want to figure out which activity leads to better learning outcomes. After interacting with a learning activity Coursera evaluates a student's learning outcome by asking them to solve a set of questions.

10. Bad testing. For both of the experiments bellow, point out a flaw in the experimental design:
  - a. For the duration of two weeks Coursera tests activity1. Then for subsequent two weeks Coursera tests activity2. Coursera compares student learning outcomes.
  - b. For the duration of two weeks, Coursera gives all users in the western hemisphere activity1 and all the users in the eastern hemisphere activity2. Coursera decides which activity is more useful in general based on the difference in learning outcomes.
11. A/B testing. Over a two-week period, Coursera randomly assigns each student to either be given activity1 (group A), or activity2 (group B). The activity that is shown to each student and the student's measured learning outcomes can be found in the file: learningOutcomes.csv
  - a. What is the difference in sample means of learning outcomes between students who were given activity1 and students who were given activity2?
  - b. Calculate a p-value for the observed difference in means reported in part (a). In other words: assuming the learning outcomes for students who had been given activity1 and activity2 were identically distributed, what is the probability that you could have sampled two groups of students such that you could have observed a difference of means as extreme, or more extreme, than the one calculated from your data? Describe any code you used to calculate your answer.
  - c. File background.csv stores the background of each user. Student backgrounds fall under three categories: more experience, average experience, less experience. For each of the three backgrounds calculate a difference in means in learning outcome between activity1 and activity2, and the p-value of that difference. Describe your methodology.

## Better Peer Grading

12. (Coding) Stanford's HCI class runs a massive online class that was taken by ten thousand students. The class used peer assessment to evaluate student's work. We are going to use their data to learn more about peer graders. In the class, each student has their work evaluated by 5 peers and every student is asked to evaluate 6 assignments: five peers and the control assignment (the graders were un-aware of which assignment was the control). All 10,000 students evaluated the same control assignment and the scores they gave are in the file `peerGrades.csv`. You may use simulations to solve any part of this question.
- What is the sample mean of the 10,000 control assignment grades?
  - Students could be given a final score which is the mean of the 5 grades given by their peers. Imagine the control experiment had only received 5 peer-grades. What is the variance of the mean of those five grades? Describe how you calculated your answer.
  - Students could be given a final score which is the median of the 5 grades given by their peers. Imagine the control experiment had only received 5 peer-grades. What is the variance of the median grade that the control experiment would have been given? Show your work and describe any code you used to calculate your answer.
  - Is the expected median of 5 grades different than the expected mean of 5 grades?
  - Would you use the mean or the median of 5 peer grades to assign scores in the online version of Stanford's HCI class? Hint: it might help to visualize the scores.

## Learning While Helping

13. You are designing a randomized algorithm that delivers one of two new drugs to patients who come to your clinic—each patient can only receive one of the drugs. Initially you know nothing about the effectiveness of the two drugs. You are simultaneously trying to learn which drug is the best and, at the same time, cure the maximum number of people. To do so we will use the Thompson Sampling Algorithm.

**Thompson Sampling Algorithm:** For *each* drug we maintain a Beta distribution to represent the drug’s probability of being successful. Initially we assume that drug  $i$  has a probability of success:  $\theta_i \sim \text{Beta}(1, 1)$ .

When choosing which drug to give to the next patient we **sample** a value from each Beta and select the drug with the largest **sampled** value. We administer the drug, observe if the patient was cured, and update the Beta that represents our belief about the probability of the drug being successful. Repeat for the next patient.

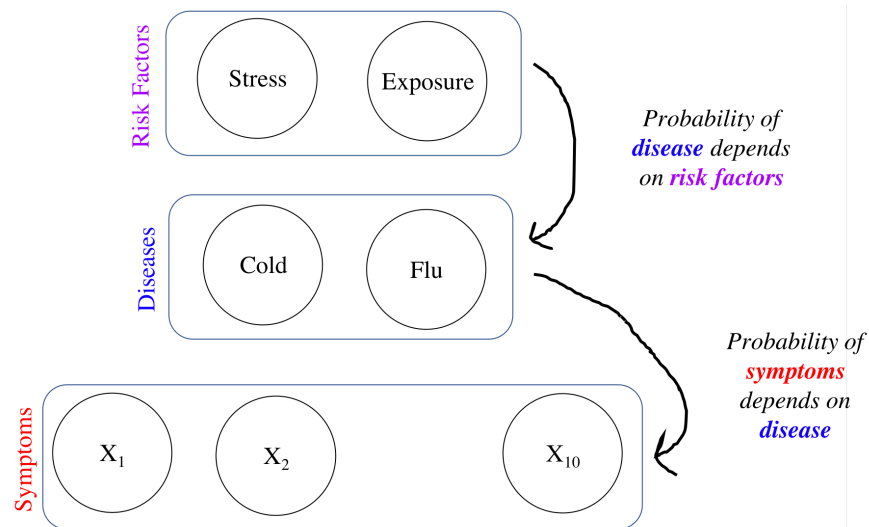
- a. Say you try the first drug on 7 patients. It cures 5 patients and has no effect on 2. What is your belief about the drug’s probability of success,  $\theta_1$ ? Your answer should be a Beta.

Method	Description
$V = \text{sampleBeta}(a, b)$	Returns a real number value in the range $[0, 1]$ with probability defined by a PDF of a Beta with parameters $a$ and $b$ .
$R = \text{giveDrug}(i)$	Gives drug $i$ to the next patient. Returns a True if the drug was successful in curing the patient or False if it was not. Throws an error if $i \notin \{1, 2\}$ .
$I = \text{argmax}(\text{list})$	Returns the index of the largest value in the list.

- b. Write pseudocode to administer either of the two drugs to 100 patients using Thompson’s Sampling Algorithm. Use functions from the table above. Your code should execute `giveDrug` 100 times.
- c. After running Thompsons’ Algorithm 100 times, you end up with the following Beta distributions:
- $\theta_1 \sim \text{Beta}(11, 11)$ ,  
 $\theta_2 \sim \text{Beta}(76, 6)$
- What is the expected probability of success for each drug?

## WebMD

14. We are writing a WebMd program that is slightly larger than the one we worked through in class. In this program we predict whether a user has a flu ( $F = 1$ ) or cold ( $C = 1$ ) based on knowing any subset of 10 potential binary symptoms (eg headache, sniffles, fatigue, cough, etc) and a subset of binary risk factors (exposure, stress).



Write psuedocode that calculates the probability of flu *conditioned on observing* that the patient has had exposure to a sick friend and that they are experiencing symptom 2 (sore throat). In terms of random variables  $P(\text{Flu} = 1 \mid \text{Exposure} = 1 \text{ and } X_2 = 1)$ :

```
def probFlu(): // P(Flu = 1 | Exposure = 1 and X2 = 1)
```

We know the prior probability for Stress is 0.5 and Exposure is 0.1.

You are given functions `probCold(s, e)` and `probFlu(s, e)` which return the probability that a patient has a cold or flu, given the state of the risk factors stress (*s*) and exposure (*e*).

You are given a function `probSymptom(i, f, c)` which returns the probability that the *i*th symptom ( $X_i$ ) takes on value 1, given the state of cold (*c*) and flu (*f*):  $P(X_i = 1 \mid F = f, C = c)$ .

- Write psuedocode that calculates `probFlu()` using **Joint Sampling**.
- (Reach) Write pseudocode that calculates `probFlu()` without using sampling.

Note: Causality implies the following: (1) that risk factors are independent. (2) all diseases are independent of one another conditioned on risk factors. (3) that all symptoms are independent of one another conditioned on knowing the state of diseases:

$$P(S = s, E = e) = P(S = s)P(E = e) \quad (1)$$

$$P(C = c, F = f \mid S = s, E = e) = P(C = c \mid S = s, E = e) \cdot P(F = f \mid S = s, E = e) \quad (2)$$

$$P(\text{symptoms} \mid F = f, C = c) = \prod_j P(X_j = k_j \mid F = f, C = c) \quad (3)$$

"Reach" means I don't *expect* CS109 students to be able to solve the problem. But thinking about it will be useful. Show your work. If you get stuck, explain what is hard.