Lisa Yan and Jerry Cain
CS 109

Quiz #2
October 28-30, 2020

CS109 Quiz #2

Exam # 00000

Take-Home Quiz information

Each quiz will be a 47-hour open-book, open-note exam. We have designed this quiz to approximate about 1-3 hours of active work (*before* typesetting).

- You can submit multiple times; we will only grade the last submission you submit before 1:00pm (Pacific time) on Friday, October 30th. No exam submissions will be accepted late. When uploading, please assign pages to each question. Failure to do so will result in a 2-point deduction. ***Please double-check that you submit the right file.***
- You should upload your submission as a PDF to Gradescope. We provide a LaTeX template if you find it useful, but we will accept any legible submission. You may also find the CS109 Probability LaTeX reference useful: `https://www.overleaf.com/read/wyhtzmdsfwkb`
- Course staff assistance will be limited to clarifying questions of the kind that might be allowed on a traditional, in-person exam. If you have questions during the exam, please ask them as private posts via our discussion forum. We will not have any office hours for answering quiz questions during the quiz.
- **For each problem, briefly explain/justify how you obtained your answer** at a level such that a future CS109 student would be able to understand how to solve the problem. It is fine for your answers to be a well-defined mathematical expression including summations (but not integrals), products, factorials, exponentials, and combinations, unless the question *specifically* asks for a numeric quantity or closed form. Where numeric answers are required, fractions are fine.

Honor Code Guidelines for Take-Home Quizzes

This exam must be completed individually. It is a violation of the Stanford Honor Code to communicate with any other humans about this exam (other than CS109 course staff), to solicit solutions to this exam, or to share your solutions with others.

The take-home exams are open-book: open lecture notes, handouts, textbooks, course lecture videos, and internet searches for conceptual information (e.g., Wikipedia). Consultation of other humans in any form or medium (e.g., communicating with classmates, asking questions on sites like Chegg or Stack Overflow) is prohibited. All work done with the assistance of any external material in any way (other than provided CS109 course materials) must include citation (e.g., "Referred to Wikipedia page on X for Question 2."). Copying solutions is unacceptable, even with citation. If by chance you encounter solutions to the problem, navigate away from that page before you feel tempted to copy.

If you become aware of any Honor Code violations by any student in the class, your commitments under the Stanford Honor Code obligate you to inform course staff. ***Please remember that there is no reason to violate your conscience to complete a take-home exam in CS109.***

I acknowledge and accept the letter and spirit of the Honor Code:

Name (typed or written): ______________________________

1 Undergraduates and Part-Time Jobs [22 points]

According to the California Center for Overextended Students, 74% of Stanford undergraduates and 86% of Berkeley undergraduates maintain part-time jobs in addition to their coursework, all independently of each other.

a. (4 points) You interview five Stanford undergraduates and five Berkeley undergraduates. What is the probability that exactly nine of the ten hold part-time jobs?

In addition to providing justification above,
please compute a numeric answer: ________________________

b. (5 points) Assume you interview Stanford undergraduates, one after another, to survey who holds a part-time job. Let S_k be the number of Stanford undergraduate that must be interviewed until you find the k^{th} student who holds a part-time job. What is $P(S_{14} = 14)$?

In addition to providing justification above,
please compute a numeric answer: ________________________

Problem 1 continued on next page.

c. (7 points) Suppose there are 7000 Stanford undergraduates and 33000 Berkeley undergraduates, and you choose two students at random from the 40000 combined and learn they each have part-time jobs? What is the conditional probability that both students attend Berkeley, given both have part-time jobs?

In addition to providing justification above,
please compute a numeric answer: __________________________

d. (6 points) Suppose you interview 120 Stanford undergraduates individually. What is the approximate probability that strictly more than 90 of them hold part-time jobs?

In addition to providing justification above,
please compute a numeric answer: __________________________

2 Fish Sticks [12 points]

Fish Sticks (a frozen meal company) wants to analyze the success of their new home-delivery website. Suppose users independently visit the homepage at an average rate of 3 users per minute.

a. (2 points) What is the expected number of users who visit the homepage in the next 5 minutes?

In addition to providing justification above,
please compute a numeric answer: ______________________________ users

b. (4 points) What is the probability that there are exactly 20 users who visit the homepage in the next 5 minutes?

In addition to providing justification above,
please compute a numeric answer: ______________________________

Each user that visits the homepage downloads exactly 5 MB of website data. We define **homepage load** in a minute to be the total amount of website data downloaded across all users who visit the homepage in that minute, e.g., 10 homepage visitors in the next minute means the homepage load in that minute is 50 MB. As before, users still independently visit the homepage at an average rate of 3 users per minute.

c. (2 points) What is the expected homepage load in the next minute?

In addition to providing justification above,
please compute a numeric answer: ______________________________ MB

Problem 2 continued on next page.

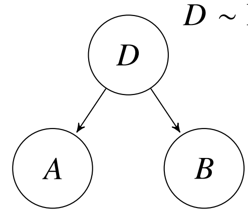
d. (4 points) What is the probability that the homepage load is exactly 20 MB in the next minute?

In addition to providing justification above,
please compute a numeric answer: ____________________

3 Stanford Fish Sticks [30 points]

Fish Sticks (the same frozen meal company) now wants to model their hourly homepage traffic from Stanford. The company decides to model two different behaviors for homepage visits according to the Bayesian Network on the right:

$D \sim \text{Ber}(p = 0.3)$



	A
$D = 0$:	$\text{Poi}(5)$
$D = 1$:	$\text{Poi}(8)$

	B
$D = 0$:	$\text{Poi}(3)$
$D = 1$:	$\text{Poi}(6)$

A and B are the numbers of Stanford students and faculty, respectively, who visit the Fish Sticks homepage in an hour. Since Fish Sticks does not know when Stanford people eat, the company models demand as a "hidden" Bernoulli random variable D, which determines the distribution of A and B. Recall that in a Bayesian Network, random variables are conditionally independent given their parents. For example, given $D = 0$, $A \sim \text{Poi}(5)$ and $B \sim \text{Poi}(3)$, two independent random variables.

a. (6 points) Given that 6 users from group A visit the homepage in the next hour, what is the probability that $D = 0$?

In addition to providing justification above,
please compute a numeric answer: ____________________

b. (10 points) What is the probability that in the next hour, the *total* number of users who visit the homepage from groups A and B is equal to 12, i.e., what is $P(A + B = 12)$?

In addition to providing justification above,
please compute a numeric answer: ____________________

Problem 3 continued on next page.

c. (14 points) Simulate $P(A + B = \texttt{total})$, where `total` = 12, by implementing the `infer_prob_total(total, ntrials)` function below using rejection sampling.

- `total` is the total number of users from groups A and B in the event $A + B = \texttt{total}$.
- `ntrials` is the number of observations to generate for rejection sampling.
- `prob` is the return value to the function, where `prob` $\approx P(A + B = \texttt{total})$.
- The function call is implemented for you at the bottom of the code block.

You can call the following functions from the `scipy` package:

- `stats.bernoulli.rvs(`p`)`, which randomly generates a 0 or 1 with probability p
- `stats.poisson.rvs(`λ`)`, which randomly generates a value according to a Poisson distribution with parameter λ

You are not required to use lists or NumPy arrays in this question (but you can if you want). **Pseudocode is fine** as long as your code accurately conveys your approach. We are not grading on style nor syntax.

```
import numpy as np
from scipy import stats

def infer_prob_total(total, ntrials):
        # YOUR CODE BELOW

        # END YOUR CODE
        return prob


ntrials = 50000
total = 12
print("Simulated P(A + B) =", infer_prob_total(total, ntrials))
```

4 Mutually Recursive Code Analysis [16 points]

After much research, you've finally arrived at function to how many pairs of flip flops you should bring to the beach, since flip flops tend to disappear in the sand and get washed away in the tide, and you need to have at least one pair when you leave.

Consider the following implementation of the mutually recursive flip and flop functions:

```
from scipy import stats
from random import choice

def flip():
    eel = choice([1, 2, 3])
    if eel == 1: return 3
    if eel == 2: return 5 + flip()
    return 11 + flop()

def flop():
    fish = stats.poisson.rvs(10)
    if fish < 5: return 1
    return 2 * flip()
```

a. (6 points) What are the two smallest numbers that can be returned by a call to flop, and what are the probabilities of each being returned?

In addition to providing justification above, please compute numeric answers:

Smallest `flop` return: ________________ probability: ________________

Second smallest `flop` return: ________________ probability: ________________

Problem 4 continued on next page.

(The code block for this problem is included here for your convenience)

```
from scipy import stats
from random import choice

def flip():
    eel = choice([1, 2, 3])
    if eel == 1: return 3
    if eel == 2: return 5 + flip()
    return 11 + flop()

def flop():
    fish = stats.poisson.rvs(10)
    if fish < 5: return 1
    return 2 * flip()
```

b. (10 points) What are the expected return values of each of the two functions, `flip` and `flop`? Please provide an analytical derivation (i.e., do not just run code).

In addition to providing justification above, please compute numeric answers: `flip`: ________ `flop`: ________

That's the end of this quiz!