Will Monroe
CS 109

Problem Set #1
June 28, 2017

Problem Set #1
Due: 12:30pm on Wednesday, July 5th

With problems by Mehran Sahami and Chris Piech

For each problem, briefly explain/justify how you obtained your answer. Brief explanations of your answer are necessary to get full credit for a problem even if you have the correct numerical answer. The explanations help us determine your understanding of the problem whether or not you got the correct answer. Moreover, in the event of an incorrect answer, we can still try to give you partial credit based on the explanation you provide. It is fine for your answers to include summations, products, factorials, exponentials, or combinations; you don't need to calculate those all out to get a single numeric answer.

Note: all assignment submissions will be made online through Gradescope. You can find information on signing up to submit assignments though Gradescope on the class webpage. If you handwrite your solutions, you are responsible for making sure that you can produce **clearly legible** scans of them for submission. You may use any word processing software you like for writing up your solutions. On the CS109 webpage we provide a template file and tutorial for the LaTeX system, if you'd like to use it.

This problem set includes one question where we ask you to write some code. You'll need to include a printout of your code in PDF or image form in your Gradescope submission. Double-check that indentation is preserved and the code isn't cut off (at the end of the line or at the end of the page). For LaTeX, we recommend the `minted` package (`https://www.sharelatex.com/learn/Code_Highlighting_with_minted`) with the `breaklines` option.

1. Introduce yourself! Fill out this Google form to tell me a bit about you:

   `https://goo.gl/forms/DuJ8v0UMpsTKDDlB2`

   (No need to copy the answers into your Gradescope submission; you can select an arbitrary page or write "done" so there is something to select.)

2. 10 computers are brought in for servicing (and machines are serviced one at a time). Of the 10 computers, 3 are PCs, 4 are Macs, 2 are Linux machines, and 1 is an Amiga. Assume that all computers of the same type are indistinguishable (i.e., all the PCs are indistinguishable, all the Macs are the indistinguishable, etc.).

   a. In how many distinguishable ways can the computers be ordered for servicing?
   b. In how many distinguishable ways can the computers be ordered if the first 5 machines serviced must include all 4 Macs?
   c. In how many distinguishable ways can the computers be ordered if 1 PC must be in the first three and 2 PCs must be in the last three computers serviced?

3. You are planning out what courses you want to take for the next two years. You have 22 courses to schedule over 6 quarters. All the classes are distinct, and order of classes within a quarter doesn't matter. How many different course plans are possible:

   a. if there are no restrictions? (For example, you could put all 22 in one quarter if you wanted. This is not a recommended course planning strategy.)
   b. if you can only take at most 4 courses in any quarter?

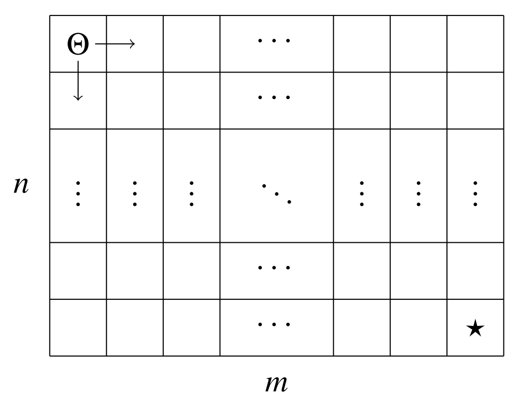
4. A substitution cipher is derived from orderings of the alphabet. How many ways can the 26 letters of the alphabet (21 consonants and 5 vowels) be ordered if each letter appears exactly once and:

   a. there are no other restrictions?
   b. all five vowels must be next to each other?
   c. no two vowels can be next to each other?

5. You are counting cards in a card game that uses two standard decks of cards. There are 104 cards total. Each deck has 52 cards (13 values each with 4 suits). Cards are only distinguishable based on their suit and value, not which deck they came from.

   a. In how many distinct ways can the cards be ordered?
   b. You are dealt two cards. How many distinct pairs of cards can you be dealt? Note: the order of the two cards you are dealt does not matter.
   c. You are dealt two cards. Cards with values 10, Jack, Queen, King and Ace are considered "good" cards. How many ways can you get two "good" cards? Order does not matter.

6. Imagine you have a robot (Θ) that lives on an $n \times m$ grid (it has n rows and m columns):



   The robot starts in cell $(1, 1)$ and can take steps either to the right or down (**no left or up steps**). How many distinct paths can the robot take to the destination ($\star$) in cell (n, m):

   a. if there are no additional constraints?
   b. if the robot must start by moving to the right?
   c. if the robot changes direction exactly 3 times? Moving down two times in a row is not changing directions, but switching from moving down to moving right is. For example, moving [down, right, right, down] would count as having two direction changes.

7. Consider an array x of integers with k elements (e.g., `int x[k]`), where each entry in the array has a distinct integer value between 1 and n, inclusive, and the array is sorted in increasing order. In other words, $1 \leq$ `x[i]` $\leq n$, for all $i = 0, 1, 2, \ldots, k-1$, and the array is sorted, so `x[0]` < `x[1]` < ... < `x[k-1]`. How many such sorted arrays are possible?

8. Given all the start-up activity going on in high-tech, you realize that applying combinatorics to investment strategies might be an interesting idea to pursue. Say you have $20 million that must be invested among 4 possible companies. Each investment must be in integral units of $1 million, and there are minimal investments that need to be made if one is to invest in these companies. The minimal investments are $1, $2, $3, and $4 million dollars, respectively for company 1, 2, 3, and 4. How many different investment strategies are available if
   a. an investment must be made in each company?
   b. investments must be made in at least 3 of the 4 companies?

9. Say we send out a total of 26 distinguishable emails to 10 distinct users, where each email we send is equally likely to go to any of the 10 users (note that it is possible that some users may not actually receive any email from us). What is the probability that the 26 emails are distributed such that there are 4 users who receive exactly 2 emails each from us and 3 users who receive exactly 6 emails each from us?

10. To get good performance when working binary search trees (BST), we must consider the probability of producing completely degenerate BSTs (where each node in the BST has at most one child). See Lecture Notes #2, Example 2 for more details on binary search trees.
    a. If the integers 1 through n are inserted in arbitrary order into a BST (where each possible order is equally likely), what is the probability (as an expression in terms of n) that the resulting BST will have completely degenerate structure?
    b. Using your expression from part (a), determine the smallest value of n for which the probability of forming a completely degenerate BST is less than 0.001 (i.e., 0.1%).

11. Say a hacker has a list of n distinct password candidates, only one of which will successfully log her into a secure system.
    a. If she tries passwords from the list at random, deleting those passwords that do not work, what is the probability that her first successful login will be (exactly) on her k-th try?
    b. Now say the hacker tries passwords from the list at random, but does **not** delete previously tried passwords from the list. She stops after her first successful login attempt. What is the probability that her first successful login will be (exactly) on her k-th try?

12. Say a university is offering 3 programming classes: one in Java, one in C++, and one in Python. The classes are open to any of the 100 students at the university. There are:
    - a total of 32 students in the Java class;
    - a total of 24 students in the C++ class;
    - a total of 21 students in the Python class;
    - 12 students in both the Java and C++ classes (note: these students are also counted as being in each class in the numbers above);
    - 10 students in both the Java and Python classes;
    - 7 students in both the C++ and Python classes; and
    - 3 students in all three classes (note: these students are also counted as being in each pair of classes in the numbers above).

    a. If a student is chosen randomly at the university, what is the probability that the student is not in any of the 3 programming classes?
    b. If a student is chosen randomly at the university, what is the probability that the student is taking *exactly one* of the three programming classes?
    c. If two different students are chosen randomly at the university, what is the probability that at least one of the chosen students is taking at least one of the programming classes?

13. A binary string containing M 0's and N 1's (in arbitrary order, where all orderings are equally likely) is sent over a network. What is the probability that the first r bits of the received message contain exactly k 1's?

14. A computer generates two random integers in the range 1 to 12, inclusive, where each value in the range 1 to 12 is equally likely to be generated.

    a. What is the probability that the second randomly generated integer has a value that is (strictly) greater than the first? (For part (a), do not use a simulation or an approximation.)
    b. **[Coding]** Write a function in the programming language of your choice that takes in a number of trials, runs that many simulations of this experiment, and returns the fraction of the trials in which the second integer is greater. In Python 3, your function signature might look like:

    ```python
    def simulation(num_trials : int) -> float:
        ...  # Your code here
    ```

    Include your code in your submission, and report the fraction returned for `num_trials = 100000`.