Lisa Yan   Problem Set #1
CS 109   September 25, 2019

Problem Set #1
Due: 1:00pm on Friday, October 4th

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.

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

    `https://forms.gle/LbfB6K9GDUm1vJ2j8`

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

    a. There are no other restrictions?
    b. The letters Q and U must be next to each other (but in any order)?
    c. All five vowels must be next to each other?
    d. No two vowels can be next to each other?

3. 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?

4. In how many ways can n identical server requests ("identical balls") be distributed among r servers ("urns") so that the ith server receives at least m_i requests, for each $i = 1, 2, \ldots, r$? You can assume that $n \geq \sum_{i=1}^{r} m_i$.

5. Determine the number of vectors $(x_1, x_2, \ldots, x_n)$ such that each x_i is a non-negative integer and $\sum_{i=1}^{n} x_i \leq k$, where k is some constant non-negative integer. Note that you can think of n (the size of the vector) and k as constants that can be used in your answer.

6. Imagine you have a robot (Θ) that lives on an $n \times m$ grid (it has n rows and m columns):
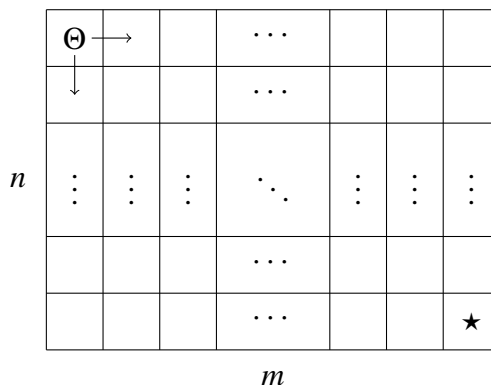
The robot starts in cell $(1, 1)$ and can take steps either to the right or down (**no left or up steps**). How many distinct paths can the robot take to the destination ($\star$) in cell (n, m):

   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. If we assume that all possible poker hands (comprised of 5 cards from a standard 52 card deck) are equally likely, what is the probability of being dealt:

   a. a flush? (A hand is said to be a flush if all 5 cards are of the same suit. Note that this definition means that *straight flushes* (five cards of the same suit in numeric sequence) are also considered flushes.)
   b. two pairs? (This occurs when the cards have numeric values a, a, b, b, c, where a, b and c are all distinct.)
   c. three of a kind? (This occurs when the cards have numeric values a, a, a, b, c, where a, b and c are all distinct.)

8. Say we roll a six-sided die six times. What is the probability that:

   a. We will roll two different numbers, *thrice* (three times) each?
   b. We will roll some number *exactly* three times?

9. 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 27 students in the Java class;
   - a total of 28 students in the C++ class;
   - a total of 19 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);
   - 4 students in both the Java and Python classes;
   - 11 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?

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. Suppose that m strings are hashed (randomly) into N buckets, assuming that all N^m arrangements are equally likely. Find the probability that exactly k strings are hashed to the first bucket.

12. 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?

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. Say we send out a total of 20 distinguishable emails to 12 distinct users, where each email we send is equally likely to go to any of the 12 users (note that it is possible that some users may not actually receive any email from us). What is the probability that the 20 emails are distributed such that there are 4 users who receive exactly 2 emails each from us and 3 users who receive exactly 4 emails each from us?

15. A computer generates two random integers in the range 1 to 10, inclusive, where each value in the range 1 to 10 is equally likely to be generated. What is the probability that the second randomly generated integer has a value that is (strictly) greater than the first?

16. **[Coding]** Consider a game, which uses a generator that produces independent random integers between 1 and 100, inclusive. The game starts with a sum $S = 0$. The first player adds random numbers from the generator to S until $S > 100$, at which point they record their last random number x. The second player continues by adding random numbers from the generator to S until $S > 200$, at which point they record their last random number y. The player with the highest number wins; e.g., if y > x, the second player wins. Is this game fair? Write a program to simulate 100,000 games. Based on your simulations, what is the probability estimate that the second player wins? You should report this value to three decimal places. For extra credit, calculate the exact probability (without sampling).

    Please include any code that you wrote. You can take a screenshot, print with your code editor's interface, or even copy the text into your LaTeX document, but please make sure that your code is readable (has indentation, is not cut off at the end of the line or at the end of the page). For LaTeX, we recommend the verbatim environment, or (for code-highlighting) the minted package (https://www.sharelatex.com/learn/Code_Highlighting_with_minted) with the breaklines option.