CS 537: *Randomness in Computing* Prof. Sofya Raskhodnikova
Boston University September 20, 2024

Homework 4 – Due Thursday, September 26, 2024

Page limit You can submit **at most** 2 pages per problem, even if the problem has multiple parts. If you submit a longer solution for some problem, only the first 2 pages will be graded.

Reminder Collaboration is permitted, but you must write the solutions *by yourself without assistance*, and be ready to explain them orally to the instructor if asked. You must also identify your collaborators and whether you gave help, received help, or worked something out together. Getting solutions from outside sources such as the Web or students not enrolled in the class is strictly forbidden.

Exercises Please practice on exercises in Chapter 2 of Mitzenmacher-Upfal.

Problems

1. (**Geometric distribution and related distributions**)

   (a) (**In parallel**) Two faulty machines M_1 and M_2 are repeatedly run synchronously in parallel (i.e., both machines execute one run, then both machine execute a second run, and so on.) On each run, M_1 fails with probability p_1 and M_2 fails with probability p_2, all failure events are independent. Let the random variables X_1 and X_2 denote the number of runs until the first failure of M_1 and M_2, respectively. Let $X = \min\{X_1, X_2\}$ denote the number of runs until the first failure of *either* machine. What distributions do X_1, X_2 and X have? Give a formal justification for your answer for X.

   (b) (**007 style**) James Bond is imprisoned in a cell with three *possible* ways to escape: an air-conditioning duct, a sewer pipe, and the door (which is unlocked). The air-conditioning duct leads him on a two-hour trip after which he falls through a trap door on his head, much to the amusement of his captors. The sewer pipe is similar, but takes five hours to traverse. Each fall produces temporary amnesia (in fiction, people get this a lot) and he is returned to his cell immediately after his fall. Assume he always chooses one of these three exits with probability 1/3. On average, how long does it take before he realizes that the door is unlocked and he just escapes?

   (c) (**k-sided die**) You have a fair k-sided die, where $k \geq 4$. The sides of the die are labeled with $[k]$. You roll the die until you get a **pair** of consecutive 4s. Find the expected number of rolls by using the following method. Break down the process into two phases, where the first phase lasts until you obtain the first 4. Define appropriate random variables. Use the law of total expectation to understand the second phase.

2. (**Fish**) In the ocean around Boston there are n kinds of fish. Each catch comes uniformly at random from the n kinds.

   (a) What is the expected number of fish you must catch to get all n kinds?

   (b) If you catch $2n$ fish, what is the expected number of kinds of fish that you did not get?

   (c) If you catch $3n$ fish, what is the expected number of kinds of fish that you got exactly once?

   (d) What is the expected number of fish you must catch to get $n/2$ kinds? (You may assume that n is even.)

You have a pond with k fish in your yard.

(e) Every morning you catch one fish uniformly at random and cook it. By the evening, one new fish is born. What is the expected age (in days) of the oldest fish in the pond, assuming that you have been doing this for a really long time?

You discovered a kind of fish that can reproduce asexually. One fish of this kind produces one offspring with probability p_1, two offsprings with probability p_2, and zero offsprings with probability $1 - p_1 - p_2$.

(f) You caught one fish of this kind and put it in your fish tank. Assuming no fish die, what is the expected number of fish your tank will eventually have? For which values of p_1 and p_2 is it bounded?

3. (**Randomized Quicksort**)

You are a Teaching Fellow for a class with n students, where n is huge. Your fellow TFs and you have to sort n graded exams. Inspired by your recent CS 537 lecture, you decide to use randomized quicksort. Your part of the work is the following: you choose a (uniformly random) pivot, perform all comparisons for that pivot and then recurse on the exams preceding the pivot (in alphabetical order). (The remaining recursive calls are handled by other TFs.) In each recursive call, you do the same thing as before: choose a pivot, perform all comparisons for that pivot and then recurse only on exams preceding the pivot.

(a) What is the expected number of comparisons you have to perform?

(b) You decide to analyze the expected number of pivots you get to choose. Your friend Hurry Sorter claims that it is $\log_2 n$ because, after processing each pivot, on average you give half of the remaining exams to other TFs. What's wrong with his argument?

(c) What is the expected number of pivots you get to choose?

4* (**Optional, no collaboration**). Tim asks for your help with sorting exams in alphabetical order. You are trying to convince him that the exams are already nearly sorted. You pull out a set of assignments (of your choice) in order they appear, and Tim checks that they are indeed in alphabetical order. Suppose there are n assignments, and they are actually in a uniformly random order. Let L be the largest number of assignments for which the check that Tim performs passes. (For example, if the assignments are in order Divya, Annie, Spyros, Tolya, Calvin, Noah, Erick, then L is 3: Annie, Spyros, and Tolya appear in the correct order, but there is no larger set of assignments that are in alphabetical order.) Prove that the expectation of L is at least $\sqrt{n}$. (For simplicity, you may assume that $\sqrt{n}$ is an integer, and don't actually play this trick on Tim!)