University of Washington
Department of Computer Science and Engineering
CSEP 521, Winter 2021

January 12, 2021

Homework 2, Due Thursday, January 21, 2021

Problem 1 (10 points):

Suppose that there are n office workers who are sent home to spend more time with family during the pandemic. When the pandemic ends, management assigns each worker to a random desk, without the requirement that only one person is assigned to a desk. In other words, worker i is assigned to desk j with probability $\frac{1}{n}$, independent of the assignment of other people to desks.

a) What is the expected number of people getting their original desk back?

b) What is the probability of having zero people assigned to a particular desk?

c) What is the probability of having exactly one person assigned to a particular desk?

d) What happens to the probabilities from parts b and c and n gets large?

e) What is the probability that no two people are assigned to the same desk?

Problem 2 (10 points):

In the last homework assignment, you implemented the selection algorithm using different pivot selection strategies. The motivation behind median-of-three and median-of-five pivot heuristics is that by getting a pivot that splits the input in half, you will more quickly find the value. This problem is to explore the limit of this technique by assuming that you always a perfect split. Show that if the pivot is always the median, then the selection algorithm makes (about) $2n$ comparisons. (The code, as presented in class, returns the pivot if it is the element we are looking for. To avoid this issue, you can show that if the pivot is always the median, then the selection algorithm takes $2n$ comparisons to find the maximum element.)

Problem 3 (10 points):

The undirected minimum $S-T$ cut problem is given an undirected graph $G=(V,E)$ with distinguished vertices s and t, find a partition of the vertices (S,T) (where $S \cup T = V$ and $S \cap T = \emptyset$) with $s \in S$ and $t \in T$ which minimizes the number of edges between S and T. Karger's Algorithm solves a similar problem, where we do not have specific vertices to anchor the cut, and look for a partition of V into a pair of non-empty subsets S and S' that minimizes the number of edges in the the cut.

Show that Karger's algorithm does not work for the $S-T$ mincut problem. Karger's algorithm "works" because a minimum cut $C=(S,S')$ has probability at least $\frac{2}{n(n-1)}$ of not having any edges contracted by the random contraction process. Show that this does not hold for the $S-T$ cut problem.

Hint: What you need to show is that for the $S - T$ mincut problem, you can have a minimum cut C which has a high probability of having an edge selected during a contraction phase. (High probability means that the probability of not being contracted is $c^{-\alpha n}$ for some constants $c > 1$ and $\alpha > 0$. Note that this depends on n, so you will need to come up with an example that applies to a family of graphs.)

Programming Problem 4 (20 points):

The *Coupon Collector* problem is: There are n types of coupons. Each time you get a coupon, you are given a coupon of a random type (with equal probability of receiving each type of coupon). The question is how many coupons do you expect to receive, on the average, before you have collected the full set of coupons.

Your programming assignment is to write a simulator of the Coupon Collector Problem, and run simulations to see how many coupons you receive before you have completed the set. You should compare your results with the theoretical analysis, which is that the expected number of coupons is $nH_n \approx n \ln n + 0.57n$ where H_n is the n-th harmonic number.

The analysis of the coupon collector problem is to look at the expected time to get a new coupon if there are k coupons that remain to be collected. If k coupons remain to be collected, then the probability of getting a new coupon is $\frac{k}{n}$, so the expected number of coupons until you get a new coupon is $\frac{n}{k}$.

Let X_k^n be the random variable for the number of coupons needed to get a new coupon when k coupons remain. Experimentally evaluate the expected values of X_k^n. The purpose of this problem is to compare the theoretical analysis of the coupon collector with an experimental version (see Wikipedia), so you will need to identify suitable parameters in terms of range of values and number of repetitions.