Homework 11

We may grade a **subset of the assigned questions**, to be determined after the deadline, so that we can provide better feedback on the graded questions.
Unless otherwise stated, each question requires sufficient justification to convince the reader of the correctness of your answer.
For bonus questions, we will not provide any insight during office hours or Piazza, and we do not guarantee anything about the difficulty of these questions.
We strongly encourage you to typeset your solutions in LaTeX.
If you collaborated with someone, you must state their name(s). You must write your own solution for all problems and may not look at any other student's write-up.

0. If applicable, state the name(s) and uniqname(s) of your collaborator(s).

   Solution:

1. Last month, while attempting to take a picture of the wild turkeys across the FXB, Daphne got chased by some very angry turkeys. As a result, her doctor said that she is at risk of contracting meleagrisphobia (a phobia of turkeys). Unfortunately, the scientists at Michigan Medicine who studied meleagrisphobia don't have a very good way to determine if someone has meleagrisphobia: on each trial, their test produces a false negative with probability $\frac{1}{3}$ and a false positive with probability $\frac{1}{3}$. Assume that each trial is completely randomized and independent of other trials.

   (a) Having recognized the danger of contracting meleagrisphobia due to her scary encounter with the turkeys, Daphne asks the scientists to run the test on her n times. Let A be the number of times the test comes back positive (i.e., it says "Daphne has meleagrisphobia"). Assuming Daphne does **not** have meleagrisphobia, compute $\mathrm{E}[A]$.

   Solution:

   (b) Suppose that Daphne **does** have meleagrisphobia. Let Y be a random variable for the number of trials that come back positive, out of n trials.

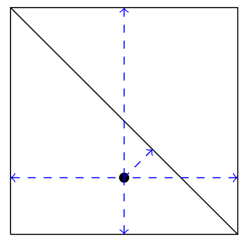
   i. Find a closed-form expression for $\mathrm{E}[Y]$ in terms of n.

   Solution:

   ii. To handle the false negatives, the scientists decide to run n tests and give a diagnosis based on the result which happens strictly more than $n/2$ times. Using the lower-tail Chernoff bound, find the minimum (odd) value of n such that $Y > n/2$ with probability at least 95%, given that Daphne **does** have meleagrisphobia.

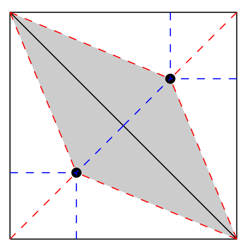
   Solution:

2. One way to estimate $\sqrt{2}$ is by "throwing darts" uniformly at random into a 1×1 square.



(a) Let p be the probability that a uniformly random point Q in the unit square is closer to the diagonal between $(0, 1)$ and $(1, 0)$ than to all of the sides of the square. Prove that $p = \sqrt{2} - 1$.

Hint: Consider the diagram below. The region of points that are closer to the diagonal than to every side of the square form a rhombus with vertices $(0, 1)$, $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $(1, 0)$, $(1 - \frac{1}{\sqrt{2}}, 1 - \frac{1}{\sqrt{2}})$. (You do not have to prove this.)



Solution:

(b) Using the previous part, design a randomized algorithm that works by throwing n darts randomly in the square, and outputs a value depending on where the darts land. Let τ be the random variable representing the output of your algorithm. Prove that the expected value of τ is $\sqrt{2}$.

Solution:

(c) Let τ be the random variable representing the output of your algorithm from part (b). Use the appropriate Chernoff bound to find the minimum number of darts that should be thrown to ensure each of the following accuracies for τ with at least 99% confidence.

i. $|\tau - \sqrt{2}| < 0.1$

Solution:

ii. $|\tau - \sqrt{2}| < 0.01$

Solution:

iii. $|\tau - \sqrt{2}| < 0.001$

Solution:

3. In class we proved that the *expected* time of RQuickSort is $O(n \log n)$, but this does not necessarily imply that its running time is $O(n \log n)$ with probability close to 1. In this problem you'll prove that it does, in fact, run in $O(n \log n)$ time with high probability.

   In RQuickSort, call a pivot p "good" if Partition$(A[1..n], p)$ returns (L, R) with $|L| \leq (2/3)n$ and $|R| \leq (2/3)n$. The probability that p is a good pivot is at least $1/3$; any pivot in the middle third of the sorted order is a good pivot.

   (a) Consider all the recursive calls to RQuickSort containing a particular input element e. Prove that there are at most $\log_{3/2} n$ calls to Partition of the form Partition$(B[1..m], p)$ where $e \in B[1..m]$, p is a good pivot, and $B[1..m]$ is a subarray of A.

   Solution:

   (b) Define $X = X_1 + \cdots + X_t$, where X_i's are independent indicator random variables with $\Pr[X_i = 1] = 1/3$. For any positive integer t, argue that

   $$\Pr[e \text{ is involved in } \geq t \text{ recursive calls}] \leq \Pr[X \leq \log_{3/2} n].$$

   Note: If an event p implies an event q, then $\Pr[p] \leq \Pr[q]$, since q could either happen because p happened, or q could happen some other way.

   Solution:

   (c) Using parts (a) and (b), show that $\Pr[e \text{ is involved in } \geq t \text{ recursive calls}] < 1/n^2$, where $t = c \ln n$ and c is some constant. You can make c as large a constant as you like.

   Solution:

   (d) Using part (b), argue that with probability at least $1 - 1/n$, *every* element e is (simultaneously) involved in less than some $t = c \ln n$ recursive calls.

   Solution:

   (e) Conclude that with probability at least $1 - 1/n$, RQuicksort takes $O(n \log n)$ time.

   Solution: