Homework 1 (due 1/29)

DS-563 / CD-543 @ Boston University

Spring 2024

Before you start…

Collaboration policy: You may verbally collaborate on required homework problems, however, you must write your solutions independently. If you choose to collaborate on a problem, you are allowed to discuss it with **at most three** other students currently enrolled in the class.

The header of each assignment you submit must include the field "Collaborators:" with the names of the students with whom you have had discussions concerning your solutions. A failure to list collaborators may result in credit deduction.

You may use external resources such as textbooks, lecture notes, and videos to supplement your general understanding of the course topics. You may use references such as books and online resources for well known facts. However, you must always cite the source.

You may **not** look up answers to a homework assignment in the published literature or on the web. You may **not** share written work with anyone else.

Submitting: Solutions should be submitted via Gradescope (entry code: 6G4V6G). Your solutions should be typed. It is strongly suggested to use LaTeX.

Grading: Whenever we ask for an algorithm (or bound), you may receive partial credit if the algorithm is not sufficiently efficient (or the bound is not sufficiently tight).

Questions (10 points each)

1. Read the handout about useful probabilistic inequalities on the course webpage. Which one is your favorite and why?

   Note: For the rest of this homework, please be very explicit what probabilistic inequalities you are using when you apply them.

2. In minimization problems, the goal is to output a solution that minimizes some value as much as possible. Examples include finding the shortest path from x to y, finding the shortest tour that passes by all mailboxes for your friend who is a mail carrier, or finding the most time efficient way of assembling cars in a factory.

   Suppose that you have a randomized algorithm for a minimization problem. For a given input instance x, let $\mathrm{OPT}(x) \geq 0$ be the optimal value. The algorithm outputs a solution that *in expectation* has value at most $4\,\mathrm{OPT}(x)$, where the expectation is taken over its internal randomness. How much can you narrow the following probabilities, compared to the obvious range $[0, 1]$? Explain.

(a) The probability of outputting a solution of value at most $3\,\mathrm{OPT}(x)$.

(b) The probability of outputting a solution of value at most $4\,\mathrm{OPT}(x)$.

(c) The probability of outputting a solution of value at most $7\,\mathrm{OPT}(x)$.

3. Suppose that you have a randomized algorithm that estimates some quantity. It can err by providing an estimate that is too high or too low, which we will refer to as a *bad estimate*. However, with probability at least $2/3$ (taken over the algorithm's internal randomness), it provides a good estimate, i.e., one that is in the desired range. Given any $\delta \in (0, 1/3)$, how can you improve the algorithm so that it errs with probability at most δ?

   Hint: How about taking the median of a number of independent estimators? How can you bound the probability that half or more of the estimates are outside of the desired range?

4. Consider an experiment in which an event $\mathcal{E}$ occurs with probability p. Why does it suffice to run the experiment independently $\Omega(1/p)$ times to see $\mathcal{E}$ occur at least once with probability $1/2$?

   Hint: What is the probability that you repeat the experiment k times and it does not occur? You may find the following inequality useful: $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$.

5. How much time (approximately) did you spend on this homework? Was is too easy/too hard?