Mathematical Toolkit **Fall 2026**

Homework 4

Due: November 18, 2026

Note: *You may discuss these problems in groups. However, do not look up solutions online or "look up" solutions using AI. Also, you must write up your own solutions and mention the names of the people in your group. Also, list any books, papers or other sources you refer to. It is recommended that you typeset your solutions in LaTeX. Homeworks are due by 11:59PM on the due date.*

1. **Perfect hashing.**
   Given a set S of N elements from some universe U, an ideal hashing scheme would have no collisions among elements of S and use only $O(N)$ space total to store S. Here is a way to use universal hashing to achieve this goal.

   (a) First prove that if we hash into a table of size $M = N^2$, then with universal hashing there is at least a $1/2$ probability of zero collisions. That is $\mathbb{P}_{h \leftarrow H}[\exists s \neq s' \in S : h(s) = h(s')] \leq 1/2$.

   (b) Now we want to reduce the storage needed. We will do this through a 2-level hashing scheme.

      i. First, for a hash function h, let $Q(h, i) = |\{s \in S : h(s) = i\}|$. Prove that if we use $M = N$ then $\mathbb{P}_{h \leftarrow H}[\sum_i Q(h, i)^2 > 4N] < 1/2$. Hint: first prove that $\mathbb{E}[\sum_i Q(h, i)^2] < 2N$ by writing $\sum_i Q(h, i)^2$ as a sum of simpler indicator variables $X_{ss'}$ denoting the event $h(s) = h(s')$.

      ii. Now, consider the following 2-level hashing scheme. First, choose a hash function $h^* : U \to \{0, 1, \ldots, N-1\}$ such that $\sum_i Q(h^*, i)^2 \leq 4N$. From part (i), we can do this by repeatedly trying $h \leftarrow H$, and in expectation we'll have to try at most twice. Now, for each $i \in \{0, 1, \ldots, N-1\}$, choose a second-level hash function h_i that hashes into a table T_i of size $Q(h^*, i)^2$ and has zero collisions among $\{s \in S : h^*(s) = i\}$. Finally, output the 2-level scheme that hashes s into $T_i[h_i(s)]$ for $i = h^*(s)$. Explain how we can efficiently find $h_1, ..., h_{N-1}$ with zero collisions among the relevant elements.

2. **One sided Chebyshev.**

   Recall that for a real-valued random variable X, Chebyshev's inequality shows that for any $c > 0$

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq c\right] \leq \frac{\mathsf{Var}[X]}{c^2}.$$

Note that the above bound does not say anything when c is smaller than the standard deviation $\sigma = \sqrt{\mathsf{Var}[X]}$. Prove the following one-sided variant of Chebyshev's inequality that is meaningful even when c is smaller than the standard deviation of the random variable:

$$\mathbb{P}[X - \mathbb{E}[X] \geq c] \;\leq\; \frac{\mathsf{Var}[X]}{c^2 + \mathsf{Var}[X]}.$$

(**Hint**: First bound the probability that $\mathbb{P}[X + t - \mathbb{E}[X] \geq c + t]$. Then plug in $t = \mathsf{Var}[X]/c$, which is the value of t that minimizes the right-hand-side.)

3. **Biased and unbiased coins 1.**
Suppose you have a fair coin (a coin of bias $1/2$) and you want to implement a coin of bias p for some given $0 < p < 1$. E.g., maybe $p = 1/7$. Give an algorithm for using the fair coin to implement the coin of bias p that requires making only two flips of your fair coin in expectation. (The number of flips you use will be a random variable, but its expectation should be ≤ 2.) Assume you have access to the binary representation of p (yes, this is a hint).

4. **Biased and unbiased coins 2.**
Let's now consider the opposite direction. Suppose we have access to a coin of bias p for some *unknown* value $p \in (0,1)$. Show how we can use this to implement a fair coin (bias $1/2$). What is the expected number of flips of your coin of bias p you use to output one fair bit? (Note: your answer will depend on p and will increase as p gets closer to 0 or 1.)

5. **(Due to T. Cover and M. Rabin).** Consider the following game. A friend writes down two numbers x and y ($x \neq y$) on two different slips of paper and then randomly puts one in one hand and the other in the other hand. You get to pick a hand and see the number in it. You then can either keep the number you saw or else return it and get the other number. You win if you end up with the larger number and you lose if you end up with the smaller number.

   For a given (possibly randomized) strategy S, let $P_S(x,y)$ denote your probability of winning using strategy S, given that the two numbers are x and y.

   (a) Consider the strategy S = "if the first number I see is ≥ -17, then I keep it, else I switch." What is $P_S(x,y)$? (Remember, the first number you are shown is equally likely to be x or y.) Your answer should be a function of x and y.

   (b) Give a randomized strategy S such that $P_S(x,y) > 1/2$ for all $x \neq y$.