CS91T Lab Assignment 4

This homework is due at 11:59PM on Wednesday, April 5. For the written portion of this assignment, write your solution using LaTeX. Submit this homework in a file named `hw4.tex` using **github**.

This is a partnered assignment, but you may choose to work on this solo, assuming no students in the class need a partner. You should primarily be discussing this assignment with your partner. It's ok to discuss approaches at a high level. However, you should not reveal specific details of a solution, nor should you show your written solution to anyone else.

The main **learning goals** of this assignment are

Part 1: Written Problems

1. (**Mitzenmacher-Upfal Problem 5.12.**) The following problem models a simple distributed system wherein agents contend for resources but "back off" in the face of contention. Balls represent agents, and bins represent resources.

   The system evolves over a number of rounds. Every round, balls are thrown independently and uniformly into n bins. Any ball that lands in a bin by itself is *served* and removed from consideration. The remaining balls are thrown again in the next round. We begin with n balls in the first round, and we finish when every ball is served.

   (a) If there are b balls at the start of a round, what is the expected number of balls at the start of the next round?

   (b) Suppose that every round, the number of balls served was exactlythe expected number of balls to be served. Show that all the balls would be served in $O(\log\log(n))$ rounds.

   Hint: If x_j was the expected number of balls left after j rounds, show that $x_{j+1} \leq \frac{x_j^2}{n}$, and use this inequality to show there will be $O(\log\log(n))$ rounds.

2. **Probabilistic Method.** In this problem, you will prove the following fact using the probabilistic method.

   Fact 1 *Let n, m be positive integers. Let $0 < p < 1$ and $q := 1 - p$. Then,*

   $$(1 - p^m)^n + (1 - q^n)^m \geq 1 \ .$$

   (a) First, let M be an $n \times m$ matrix with 0/1 entries. Independently fill in each cell $M[i, j]$ with a 1 with probability p, and with a 0 with probability q. Let A_r be the event that row r contains at least one 0 cell, and let B_c be the event that column c contains at least one 1 cell.[1] *Briefly* argue that the events $\{A_r\}$ are mutually independent, as are $\{B_c\}$.

   (b) What is $\Pr[A_r]$? What is $\Pr[B_c]$?

   (c) Now, let $A^* := \bigcap_r A_r$ and $B^* := \bigcap_c B_c$. What is $\Pr[A^*]$? What is $\Pr[B^*]$?

   (d) Next, argue that A^*, B^* cannot both *not* happen, i.e., that $\Pr[A^* \cup B^*] = 1$.

[1] Formally, A_r is the event that $M[r, c] = 0$ for some c. B_c is the event that $M[r, c] = 1$ for some r.

(e) Using the union bound, conclude the fact must hold.

3. **Probabilistic Method**. You're given a unit square[2]. Inside the square lie a number of circles. The total circumference of all the circles is 10. You're not given how many circles there are, nor how large each circle is (aside from the fact that the total circumference is 10), nor where they are (aside from the fact that each circle lies completely inside the unit square). Show that it is possible to draw a straight line that intersects at least four of the circles.

   Hint: Let L be a uniform vertical line. Argue that $\Pr[L \text{ intersects four circles}] > 0$.

4. **Square-free matrices.** Call an $n \times n$ binary matrix M *square-free* if there is no 3×3 submatrix with all 1 entries. For example, the following 4×4 matrix is square-free

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

   whereas the following 4×4 matrix is *not* square-free because the each entry is a 1 in the submatrix $\{1, 3, 4\} \times \{1, 2, 4\}$:

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

   There are applications in theoretical computer science where we'd like binary matrices to be square-free and have as many one-entries as possible. Let $Z(n)$ be the maximum number of one entries in a square-free $n \times n$ binary matrix.

   Use the alteration technique to give a lower bound on $Z(n)$ — your answer should state that "there exists a square-free binary matrix with X one entries". Express X as a function of n and make it as large as you can.

Part 2: Research Papers One of the things graduate students and more experienced researchers spend lots of time on is reading research papers. This can be a frustrating experience for beginning researchers, especially in theoretical computer science and algorithm design. Part of the problem is that research papers are technical. **Reading research papers productively is a skill**. Just like any skill, it takes practice. Fortunately, with practice and patience, you can learn how to read research papers productively. One of the first things to know about reading research papers is that it is an iterative process. To read a research paper effectively, it helps to read it iteratively in multiple passes.

- First, read the paper for a high-level overview. What are the main contributions of the paper? Does it introduce a new model, technique, or research problem? Does it solve a well-known open problem?

[2] A unit square is a square of length 1 on all sides

- Second, read through the paper more thoroughly, but skip over most of the proofs. By the end of this pass, you should have (i) a mostly complete understanding of what the paper's results and contributions are, (ii) the structure of how they prove their results, and (iii) an idea of what open problems the paper poses.
- Finally, read through the paper linearly. Try to understand each proof completely.

The purpose of this part of your assignment is to get an initial taste reading research papers.

1. First, read the paper "How to Read a Paper" by S. Keshav. This short paper gives a similar three-pass description of reading research papers.
2. Second, make a **first pass** over the paper "The Space Complexity of Approximating the Frequency Moments" By Alon, Matias, and Szegedy. Do not attempt to fully understand this paper! Just understand what the results are and what the main computation model is.
3. Third, make a partial **second pass** over the AMS paper. [You will not have time to complete a second pass.] Try to understand the streaming model and one of their theorem.
4. Write a three paragraph description of what you learned from this paper. (This should include but is not limited to: what problem(s) does this paper consider? What are the main result(s)? What techniques do they use?)

You should spend no more than three hours on this part of the assignment.

Part 3: Extra Challenge Problems In many of the assignments, there will be a few extra challenge problems. These problems are completely optional.

1. **(extra challenge problem).** Complete your second pass on the AMS paper. Then, pick one open problem and do a literature search to see if it has been solved since publication of the AMS paper.
2. **(extra challenge problem).** Get better bounds for $R(5)$.
3. **(extra challenge problem).** Get better bounds for $R(k)$.