CSCE 658: Randomized Algorithms – Spring 2024
Mock Final Exam

Your Name:

Problem 1. (40 points total) Multiple choice and free response.

1. (5 points) Suppose a bag contains 3 red marbles, 2 blue marbles, and 5 green marbles. Suppose we draw two marbles from the bag, without replacement. Conditioned on the first drawn marble being green, what is the probability that the second drawn marble is red? (No justification is required.)

2. (5 points) Suppose $x \in \mathbb{R}^n$ is an arbitrary vector. Which of the following is true?
   - ◯ $\|x\|_1 \geq \|x\|_2$
   - ◯ $\|x\|_1 = \|x\|_2$
   - ◯ $\|x\|_1 \leq \|x\|_2$
   - ◯ There is no fixed relationship between $\|x\|_1$ and $\|x\|_2$

3. (5 points) Let P be a linear program written in standard form and D be its dual program. Let $P(x)$ be the objective of a feasible solution to P and $D(y)$ be the objective of a feasible solution to D. Which of the following is always true?
   - ◯ $P(x) > D(y)$
   - ◯ $P(x) = D(y)$
   - ◯ $P(x) < D(y)$
   - ◯ None of the above

4. (5 points) The Misra-Gries algorithm is a streaming algorithm for L_2 heavy-hitters on insertion-deletion streams.

   A. True B. False

5. (5 points) There exists a deterministic algorithm that achieves the optimal error guarantee for the counting problem and is ε-differentially private.

   A. True B. False

6. (5 points) In the Equality communication problem, Alice and Bob are given binary strings $A, B \in \{0,1\}^n$ respectively, and their goal is to compute whether $A = B$.

   Then every deterministic two-way communication protocol that solves Equality must use $\Omega(n)$ bits of communication.

   A. True B. False

7. (5 points) There exists a deterministic algorithm that achieves the optimal regret for the online learning with experts problem.

   A. True B. False

8. (5 points) The following pseudo-code demonstrates which streaming algorithm?

```
(a) Randomly map each universe item i to one of b buckets
(b) Map each universe item to a random sign
(c) Keep a signed counter for the items in each bucket
(d) To estimate the frequency of each item, multiply the sign of the item
    by the signed counter in its assigned bucket
```

- ◯ AMS Algorithm
- ◯ CountMin
- ◯ CountSketch
- ◯ Misra-Gries

Problem 2. (20 points total) Concentration inequalities.

1. (5 points) Let X be a random variable with mean 200 and variance 100. Use Chebyshev's inequality to prove a tight upper bound on $\mathbf{Pr}\left[|X - 200| \geq 20\right]$. Describe the probability distribution of a random variable X that shows the tightness of your bound.

2. (5 points) Prove that for any random variable X, we have $\mathbb{E}\left[X^2\right] \geq \mathbb{E}\left[X\right]^2$.

3. (5 points) Let X be a non-negative random variable with $\mathbb{E}[X] > 0$. Prove that $\mathbf{Pr}[X = 0] \leq \frac{\mathbb{E}[X^2] - \mathbb{E}[X]^2}{\mathbb{E}[X]^2}$.

Problem 3. (20 points total) Max load.

In class, we showed that there exists a fixed constant $C_1 > 1$, such that after n tosses of a fair n-sided die, with high probability, no outcome occurs more than $C_1 \log n$ times. Thus perhaps after $n \log n$ tosses of a fair n-sided die, one might expect that with high probability, no outcome occurs more than $C_1 \log^2 n$ times.

Show this intuition is not tight. In particular, show that there exists a fixed constant $C_2 > 1$, such that after $n \log n$ tosses of a fair n-sided die, with high probability, no outcome occurs more than $C_2 \log n$ times.

HINT: You may use the following multiplicative Chernoff bounds:

Let $X_1, \ldots, X_n$ be independent random variables taking on values in $\{0, 1\}$. Let X denote their sum and $\mu = \mathbb{E}[X]$ denote the expected sum. Then for any $t \in (0, 1)$,

$$\mathbf{Pr}[X \geq (1+t)\mu] \leq e^{-t^2\mu/2}.$$

Problem 4. (20 points total) Lower bounds for streaming algorithms.

In the INDEX problem, Alice has a vector $x \in \{0,1\}^n$ and Bob has a position $i \in [n]$ and their goal is for Bob to determine whether $x_i = 0$ or $x_i = 1$ after receiving a message from Alice. It is known that any one-way protocol for INDEX that succeeds with probability at least $\frac{2}{3}$ requires $\Omega(n)$ communication from Alice and Bob.

Suppose a frequency vector $x \in \mathbb{R}^n$ is implicitly defined through a insertion-only data stream. Let $\mathcal{A}$ be a streaming algorithm that processes x, receives a query $i \in [n]$ *after the data stream*, and outputs whether x_i is even or odd, with probability at least $\frac{2}{3}$. Prove by a reduction from INDEX that $\mathcal{A}$ must use $\Omega(n)$ bits of space.