Solutions will be posted on the web on Monday, April 7.

Practice Midterm 2

- This will be a **closed book** exam. Calculators **are** permitted. One single-sided sheet of notes is also permitted.
- This practice midterm consists of 10 questions. The first seven questions are multiple choice; the remaining three require written answers. (The actual midterm will be shorter.)
- Answer the multiple choice questions by circling the correct answer. You should be able to answer all of these from memory, by inspection, or with a very small calculation. Incorrect answers attract a negative score, so if you do not know the answer you should **not** guess.
- Write your answers to the other questions in the spaces provided below them. None of these questions requires a long answer, so you should have enough space; if not, continue on the back of the page and state clearly that you have done so. **Show all your working**.
- The questions vary in difficulty: if you get stuck on some part of a question, leave it and go on to the next one.

---

1. A multiple choice exam has six possible answers for each question, only one of which is correct. A correct answer receives 4 points, while an incorrect answer incurs a penalty of b points. If we wish to ensure that the expected score for a student who randomly guesses on every question is zero, we should set b to be

$$\tfrac{1}{6} \qquad \tfrac{4}{5} \qquad \tfrac{5}{6} \qquad 1 \qquad \tfrac{6}{5} \qquad 5$$

---

2. Let X be a non-negative random variable.

   (a) Markov's inequality says that $\Pr[X \geq 3\mathbf{E}[X]]$ is at most

$$\frac{3}{\mathbf{E}[X]} \qquad \frac{1}{3} \qquad \frac{2}{3} \qquad \frac{\mathbf{E}[X]}{3} \qquad \frac{1}{3\mathbf{E}[X]}$$

   (b) Chebyshev's inequality says that $\Pr[|X - \mathbf{Var}[X]| \geq 3\sqrt{\mathbf{Var}[X]}]$ is at most

$$\frac{1}{9} \qquad \frac{1}{3} \qquad \frac{\mathbf{Var}[X]}{9} \qquad \frac{\mathbf{Var}[X]}{3} \qquad \frac{1}{3\mathbf{Var}[X]}$$

3. You are given a randomized algorithm that always outputs either 'yes' or 'no'. When it outputs 'yes', this answer is always correct; when it outputs 'no', this answer is correct with probability at least $\frac{1}{n}$, where n is the input size.

   (a) The number of repeated trials of this algorithm you would have to perform in order to guarantee a correct answer with probability at least $1 - \frac{1}{\mathrm{e}}$ is about

   e $\qquad$ $\ln n$ $\qquad$ n $\qquad$ n^2 $\qquad$ e^n

   (b) Let T denote the answer to part (a). The number of repeated trials of the original algorithm you would have to perform in order to guarantee a correct answer with probability at least $1 - \frac{1}{\mathrm{e}^{100}}$ is about

   $T + 100$ $\qquad$ $\mathrm{e}T$ $\qquad$ $(\ln 100)T$ $\qquad$ $100T$ $\qquad$ $\mathrm{e}^{100}T$

---

4. Let $\phi = (x_1 \vee x_2) \wedge (x_3) \wedge (x_2 \vee x_4)$ be an input for MaxSat. The *expected* number of clauses of ϕ that are satisfied by a *random* assignment is exactly

   3 $\qquad$ $\frac{3}{2}$ $\qquad$ $\frac{17}{8}$ $\qquad$ $\frac{3}{4}$ $\qquad$ $\frac{7}{8}$ $\qquad$ 2

---

5. A fair coin is tossed n times, where n is large. With probability about $\frac{2}{3}$, the *proportion* of heads obtained will deviate from the expected value $\frac{1}{2}$ by less than

   a constant $\qquad$ $\frac{1}{2\ln n}$ $\qquad$ $\frac{1}{2\sqrt{n}}$ $\qquad$ $\frac{1}{2n}$ $\qquad$ $\frac{1}{2^n}$

---

6. Let $\mathcal{H}$ be a 2-universal family of hash functions from a universe U to a table T. Let $x, y \in U$ such that $x \neq y$. The *number* of functions $h \in \mathcal{H}$ for which $h(x) = h(y)$ is at most:

   $\frac{|U|}{|T|}$ $\qquad$ $\frac{|\mathcal{H}|}{2}$ $\qquad$ $\frac{|T|}{|\mathcal{H}|}$ $\qquad$ $\frac{1}{|T|}$ $\qquad$ $\frac{|\mathcal{H}|}{|T|}$ $\qquad$ $\frac{|\mathcal{H}|}{|U|}$

---

7. X and Y are (not necessarily independent) random variables with $\mathbf{E}[X] = \mathbf{E}[Y] = 0$ and $\mathbf{Var}[X] = 1$, $\mathbf{Var}[Y] = 4$. Circle those *three* of the following statements that *must* be true about X and Y:

   $\Pr[X > 0] > 0$ $\qquad$ $\Pr[Y > X] > 0$ $\qquad$ $\Pr[\,|Y| > |X|\,] > 0$

   $\mathbf{Var}[X + Y] = 5$ $\qquad$ $\Pr[X > 0] = \Pr[X < 0]$ $\qquad$ $\Pr[\,|Y| \geq 8\,] \leq \frac{1}{4}$

---

8. **Hash Functions** Let $U = \{0, 1, \ldots, m-1\}$ be a universe of size m and $T = \{0, 1, \ldots, n-1\}$ a hash table of size $n < m$. Let $\mathcal{H}$ be a family of functions from U to T.

   (a) Define what it means for the family $\mathcal{H}$ to be *2-universal*.

   [**Note**: You should specify clearly the sample space used in your definition.]

   (b) Give an example of a 2-universal family of hash functions that has size only $O(m^2)$.

   [Note: you are **not** required to prove that your family is 2-universal.]

   (c) Assume that $m = p = n$ is prime. Define the function $h_a : \mathbf{Z}_p \rightarrow \mathbf{Z}_p$ by $h_a(x) = ax \bmod p$, and let $\mathcal{H}$ be the family of all functions h_a for $a \in \mathbf{Z}_p$. Show that $\mathcal{H}$ is 2-universal.

9. **Random Walks**

Recall that, in an undirected graph $G = (V, E)$, the commute time between two vertices u, v is the sum of the hitting times $h_{uv} + h_{vu}$, and it is also equal to $2mR_{uv}$, where $m = |E|$ and R_{uv} is the effective resistance between u and v.

(a) Let $G = (V, E)$ be a clique of size four, that is, $|V| = 4$ and for every $u, v \in V$, $)u, v) \in E$. Compute the commute time between any two vertices.

(b) Let $G = (V, E)$ be a cycle of length n. Compute the commute time between any two vertices.

(c) Let $G = (V, E)$ be a rooted tree where all leaves have depth d. Compute the commute between the root and each of the leaves.

10. **Finding a group of strangers** Consider a group of n people. Suppose that the total number of acquaintanceships in the group (i.e., pairs of people who know each other) is $2n$. We may model this situation as an undirected graph $G = (V, E)$, in which the vertices are people and we draw an edge between two people if and only if they know each other. The total number of edges is $2n$. In this problem, our aim is to find a fairly large group of people *all of whom are strangers*. Such a group of people is called an *independent set*.

    (a) Suppose we construct a random subset S of people as follows: for each person in the entire group, we independently flip a coin with heads probability $1/4$; if it comes up heads, we put the person in S. What is the value of $\mathbf{E}[|S|]$?

    (b) Let the r.v. X denote the number of edges *inside* the set S (i.e., edges of G that connect pairs of people in S). What is the value of $\mathbf{E}[X]$?

    [**Hint:** Consider each edge e of G separately; what is the probability that both endpoints of e belong to S?]

    (c) Now for each edge e inside S, choose an endpoint of e arbitrarily and remove it (if it hasn't been removed already). Let S' be the set of remaining vertices. Obviously S' must be an independent set (you should check that you understand why).

    Use parts (a) and (b) to deduce that $\mathbf{E}[|S'|] \geq \frac{n}{8}$.

    (d) Deduce from parts (a), (b) and (c) that there must exist an independent set of at least $\frac{n}{8}$ people in the original group.

    (e) Give an efficient randomized algorithm that, with probability at least $\frac{1}{15}$, finds an independent set of at least $\frac{n}{16}$ people. Justify the success probability of your algorithm.