Sample Midterm 2

6:00–8:00pm, 18 November

Read these instructions carefully

*1. This is a **closed book** exam. Calculators **are** permitted.*

2. This midterm consists of 10 questions. The first seven questions are multiple choice; the remaining three require written answers.

*3. Answer the multiple choice questions by circling the correct answer (or the best answer if more than one is correct). 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 **do not** guess.*

4. Write your answers to the other questions in the spaces provided. 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**.

5. 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. Let z be a random number less than k. The probability that z is prime is (up to a small constant factor)

$$\frac{1}{k} \qquad \frac{1}{\ln k} \qquad \frac{\ln k}{k} \qquad \frac{1}{k^2} \qquad \frac{1}{2^k}$$

---

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

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

$$\frac{1}{9} \qquad \frac{1}{3} \qquad \frac{2}{3} \qquad \frac{\mathrm{E}\,(X)}{3} \qquad \frac{1}{3\mathrm{E}\,(X)}$$

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

$$\frac{1}{9} \qquad \frac{1}{3} \qquad \frac{2}{3} \qquad \frac{\mathrm{Var}\,(X)}{3} \qquad \frac{1}{3\mathrm{Var}\,(X)}$$

---

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 $1 - \frac{1}{e}$ is about

$$\mathrm{e} \qquad \ln n \qquad n \qquad n^2 \qquad \mathrm{e}^n$$

(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 $1 - \frac{1}{e^{100}}$ is about

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

[continued overleaf]

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

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

5. Let $X_1, \ldots, X_n$ be independent, identically distributed random variables with expectation μ and variance σ^2, and let $S_n = X_1 + \cdots + X_n$. The Central Limit Theorem says that the distribution of one of the following quantities approaches the standard normal distribution as $n \to \infty$. Which one?

$$\frac{S_n - \mu}{\sigma} \qquad \frac{S_n - n\mu}{\sigma\sqrt{n}} \qquad \frac{S_n - n\mu}{\sigma/\sqrt{n}} \qquad \frac{S_n - n\mu}{\sigma} \qquad \frac{S_n - n\mu}{\sigma n}$$

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

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

7. 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 <u>number</u> of functions $h \in \mathcal{H}$ for which $h(x) = h(y)$ is at most:

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

8. 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 <u>all of whom are strangers</u>. Such a group of people is called an <u>independent set</u>.

(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 $\frac{1}{4}$; if it comes up heads, we put the person in S. What is the value of $E(|S|)$?

(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 $E(X)$? [**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 $E(|S'|) \geq \frac{n}{8}$.

(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.

9. **A threshold for isolated vertices**

Let G be a random graph in the $\mathcal{G}_{n,p}$ model with $p \le \frac{1}{2}$. A vertex v of G is said to be isolated if it is not connected to any other vertex of G. In this problem, we will show that $p = \frac{\ln n}{n}$ is a threshold value for the existence of an isolated vertex in G; i.e., if $p \ll \frac{\ln n}{n}$ then $\Pr[G$ contains an isolated vertex$] \to 1$, and if $p \gg \frac{\ln n}{n}$ then $\Pr[G$ contains an isolated vertex$] \to 0$.

(a) For a given vertex v, what is $\Pr[v$ is isolated$]$ as a function of n and p?

(b) Let the r.v. X be the number of isolated vertices in G. What is the value of $E(X)$, as a function of n and p?

(c) Show that if $p \gg \frac{\ln n}{n}$ then $E(X) \to 0$ as $n \to \infty$. [**Hint:** Show that $\ln E(X) \to -\infty$ using the fact that $\ln(1-z) \le -z$ for all $z \in [0,1]$.]

(d) Show that if $p \ll \frac{\ln n}{n}$ then $E(X) \to \infty$ as $n \to \infty$. [**Hint:** Show that $\ln E(X) \to \infty$ using the fact that $\ln(1-z) \ge -2z$ for all $z \in [0, \frac{1}{2}]$.]

(e) Deduce from part (c) that, if $p \gg \frac{\ln n}{n}$, then $\Pr[G$ contains an isolated vertex$] \to 0$ as $n \to \infty$.

(f) Now assume that $\frac{\mathrm{Var}(X)}{E(X)^2} \to 0$ as $n \to \infty$ when $p \ll \frac{\ln n}{n}$. Use Chebyshev's inequality and part (d) to deduce that in this case $\Pr[G$ contains an isolated vertex$] \to 1$ as $n \to \infty$.

(f) [**Extra Credit Only: answer on back**] Prove the fact that was assumed in part (e), i.e., prove that $\frac{\mathrm{Var}(X)}{E(X)^2} \to 0$ as $n \to \infty$ when $p \ll \frac{\ln n}{n}$.

10. Testing equality of multisets

Let $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$ be <u>multisets</u> over the universe $U = \{0, \ldots, m-1\}$, i.e., each of X and Y consists of n not necessarily distinct elements of U, given in some arbitrary order. Suppose we want to test whether $X = Y$, in the sense that X and Y both contain exactly equal numbers of all elements in U. Clearly this problem can be solved in time $O(n \log n)$ by <u>sorting</u> the members of X and Y and then comparing them. Here we investigate a more efficient randomized algorithm.

(a) By considering the polynomial $Q_X(z) = (z - x_1)(z - x_2) \ldots (z - x_n)$ (and a similar polynomial $Q_Y(z)$ for Y), design a randomized algorithm based on the Schwartz-Zippel technique for testing whether $X = Y$. Your algorithm should always output "yes" when $X = Y$, and should output "no" with probability at least $\frac{1}{2}$ when $X \neq Y$. You should specify the values of all quantities used in your algorithm (in terms of n).

(b) Assuming that each arithmetic operation can be performed in <u>constant</u> time, what is the running time of your algorithm as a function of n?

(c) The assumption of part (b) is only justified if the integers that appear during your computation are fairly small. How would you modify your algorithm to ensure that the number of bits required to perform the computation is only $O(\log n + \log \log m)$, at the cost of a small additional probability of error? Justify your answer with a rough calculation.