Midterm 2

12:30–2:00pm, 19 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 are 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. A fair coin is tossed n times, where n is large. With high probability, the number of heads deviates from the expected value $\frac{n}{2}$ by at most (up to a constant factor)

a constant $\qquad \ln\ln n \qquad \ln n \qquad \sqrt{n} \qquad n$

---

2. Let x be a random n-bit number. The probability that x is prime is approximately (up to a small constant factor)

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

---

3. You are given a randomized algorithm with the following properties. The algorithm 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 $\frac{1}{2}$. The number of repeated trials of the algorithm that are necessary to reduce the error probability to at most ϵ is approximately

$$\frac{1}{\epsilon} \qquad \frac{1}{\log_2 \epsilon} \qquad 2^{1/\epsilon} \qquad \log_2\left(\frac{1}{\epsilon}\right) \qquad \frac{1}{\epsilon^2}$$

[continued overleaf]

4. Let G be a random graph chosen according to the model $\mathcal{G}_{n,p}$.

(a) Suppose you wanted to ensure that the expected number of edges in G is asymptotically equal to $10n$ as $n \to \infty$. Which of the following values of p would you use?

$$\frac{20}{n} \qquad \frac{20}{n^2} \qquad \frac{10}{n} \qquad \frac{10}{n^2} \qquad \frac{1}{10}$$

(b) Suppose you wanted to ensure that the expected degree of any given vertex is asymptotically equal to $\ln n$ as $n \to \infty$. Which of the following values of p would you use? (Recall that the *degree* of a vertex is the number of edges incident to it.)

$$\frac{1}{\ln n} \qquad \frac{1}{n \ln n} \qquad \frac{\ln n}{n^2} \qquad \frac{1}{(\ln n)^2} \qquad \frac{\ln n}{n}$$

---

5. In the Boolean circuit model of computation, as $n \to \infty$ almost every function $f : \{0,1\}^n \to \{0,1\}$ requires circuits of size at least approximately (up to constant factors)

$$n! \qquad \frac{2^n}{n} \qquad n^2 \qquad 2^{2^n} \qquad n^{2n}$$

---

6. A random variable X has expectation $\mathrm{E}(X) = 0$ and variance $\mathrm{Var}(X) = 1$. Circle those <u>three</u> of the following statements that <u>must</u> be true about X:

$\Pr[X \le 3] \le \frac{1}{9}$ $\qquad$ $\Pr[X \le -3] \le \frac{1}{9}$ $\qquad$ $\Pr[|X| \ge 3] \le \frac{1}{9}$

$\Pr[X \ge 3] \le \frac{1}{18}$ $\qquad$ $\Pr[X = 0] < 1$ $\qquad$ $\Pr[X = 0] > 0$

---

7. Suppose that Schwartz's test is applied to a single-variable polynomial $Q(x)$ of total degree d, by picking values for x randomly from the set $S = \{1, 2, \ldots, n\}$.

(a) If $Q \not\equiv 0$, the probability that the test makes an error is at most

$$\frac{1}{d} \qquad \frac{1}{nd} \qquad \frac{d}{n} \qquad \frac{d}{2n} \qquad \frac{d}{n^2}$$

(b) For the polynomial $Q(x) = x^3 - 3x^2 + 3x - 1$, the probability of error is *exactly*

$$\frac{1}{2} \qquad \frac{1}{n} \qquad \frac{2}{n} \qquad \frac{3}{n} \qquad \frac{1}{n^2}$$

[continued overleaf]

8. The art gallery problem

An art gallery may be modeled as an undirected graph $G = (V, E)$, in which the vertices are rooms and the edges are doors connecting rooms. We will assume that there are n rooms, and that every room has doors to at least *three* other rooms.

Our task is to place guards in the rooms in such a way that every room is visible to at least one guard: we call such a placement of guards *safe*. When we place a guard in a room, we assume that she is able to see that room and all rooms connected to it, i.e., that vertex of the graph and all its neighbors. Of course, we would like to minimize the number of guards required. We will use the probabilistic method to find a solution with not too many guards.

(a) Let $S \subseteq V$ be any subset of the rooms, and let Y be the set of rooms that are neither in S nor have a neighbor in S. Show that placing guards in the set of rooms $S \cup Y$ is safe.

(b) Now suppose that we pick the subset S at random, as follows: for each room $v \in V$ independently, we place v in S with probability p. What is $\mathrm{E}(|S|)$, the expected size of S?

(c) For the random S as in part (b), show that the corresponding set Y has expected size $\mathrm{E}(|Y|) \leq n(1-p)^4$. [Hint: Recall that every room has at least three neighbors.]

(d) Now set $p = \frac{\ln 4}{4}$. Show that, with this value of p, we have $\mathrm{E}(|S| + |Y|) \leq \frac{n}{4}(\ln 4 + 1) \approx 0.6n$. [Hint: Recall that $(1 - \frac{t}{m})^m \leq \mathrm{e}^{-t}$ for all $m, t > 0$.]

[continued overleaf]

Q8 continued

(e) Deduce from part (d) that there must exist a safe placement that uses at most $\frac{n}{4}(\ln 4 + 1)$ guards.

(f) Give an efficient algorithm that, with probability at least $1 - \frac{1}{1+\epsilon}$, finds a safe placement of guards in G using at most $\frac{n}{4}(\ln 4 + 1)(1 + \epsilon)$ guards, for any specified $\epsilon > 0$. Justify the error probability of your algorithm.

(g) Explain how to modify your algorithm of part (f) so that it finds such a placement with probability at least $1 - 2^{-100}$. What is the increase in running time?

[continued overleaf]

9. Let $X_1, \ldots, X_n$ be independent, identically distributed random variables with $\Pr[X_i = 1] = \Pr[X_i = -1] = \frac{1}{2}$. Let $S_n = X_1 + \cdots + X_n$.

(a) What are the values of $\mu = \mathrm{E}(X_i)$ and $\sigma^2 = \mathrm{Var}(X_i)$?

(b) What are the values of $\mathrm{E}(S_n)$ and $\mathrm{Var}(S_n)$?

(c) Use Chebyshev's inequality to show that,

$$\Pr[|S_n| \geq 2\sqrt{n}\,] \leq \frac{1}{4} \qquad \text{for all } n.$$

(d) Use the Central Limit Theorem to show that

$$\Pr[|S_n| \geq 2\sqrt{n}\,] \to \epsilon \qquad \text{as } n \to \infty,$$

where $\epsilon \approx 0.046$. You may assume that, if Z has the standard Normal distribution, then $\Pr[Z \leq 2] \approx 0.977$.

(e) Comment briefly on the relationship between the results of parts (c) and (d).

[continued overleaf]

10. Removing duplicates

Consider the following problem. You are given n m-bit strings (not all distinct) and are asked to remove any duplicates amongst them (leaving exactly one copy of each distinct string). An obvious algorithm for this problem would first sort the strings, using any standard $O(n \log n)$ algorithm, and then remove duplicates in one pass through the sorted list. The total running time would be $O(mn \log n)$, where the factor of m arises from the cost of comparing two m-bit strings.

Give a more efficient randomized algorithm for this problem that runs in time $O(mn + n \log n)$, where we assume that arithmetic operations and comparisons on numbers with $O(\log(nm))$ bits can be performed in constant time. Your algorithm should take an additional parameter t as input, and should have an error probability of at most $\frac{1}{t}$. You should carefully justify the running time and error probability of your algorithm.