CS–174 Combinatorics & Discrete Probability, Fall 96

Final Examination

12:30–3:30pm, 17 December

Read these instructions carefully

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

2. This exam consists of 15 questions. The first nine questions are multiple choice; Q10 requires two diagrams; the remaining five 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 multiple choice exam has five possible answers for each question, only one of which is correct. A correct answer receives 1 point, while an incorrect answer receives 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

$$0 \qquad \frac{1}{5} \qquad \frac{1}{4} \qquad 1 \qquad \text{none of these}$$

---

2. Let p be a prime, and let $\mathbb{Z}_p$ denote the set $\{0, 1, \ldots, p-1\}$. Suppose that a is chosen uniformly at random from the set $\mathbb{Z}_p - \{0\}$, and b is chosen uniformly at random from the set $\mathbb{Z}_p$, independently of a.

(a) For any fixed $x \in \mathbb{Z}_p$, the probability that $ax + b = 0 \pmod p$ is

$$0 \qquad \frac{1}{2} \qquad \frac{1}{p^2} \qquad \frac{1}{p(p-1)} \qquad \frac{1}{p}$$

(b) For any fixed $x, y \in \mathbb{Z}_p$ with $x \neq y$, the probability that both $ax + b = 0 \pmod p$ and $ay + b = 1 \pmod p$ is

$$0 \qquad \frac{1}{2} \qquad \frac{1}{p^2} \qquad \frac{1}{p(p-1)} \qquad \frac{1}{p}$$

---

3. A graph with n vertices is constructed by repeatedly inserting edges chosen independently at random from among those not already inserted. As $n \to \infty$, the expected number of edges that are inserted before the graph becomes connected is approximately (up to small constant factors)

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

4. Let $G = (V, E)$ be a connected undirected graph with n vertices and m edges. Let C_v denote the expected cover time of G starting from vertex v, and let $C_G = \max_v C_v$.

(a) The best general upper bound on C_v is

$O(n + m)$ $O(nm)$ $O(n^2m)$ $O(nm^2)$ $O(n^3)$

(b) If we wish to guarantee that a random walk starting from v visits every vertex with probability at least $\frac{1}{2}$, it suffices to take a walk of length

C_v $2C_v$ $4C_v$ C_v^2 none of these

(c) If we wish to guarantee that a random walk starting from v visits every vertex with probability at least $1 - 2^{-100}$, it suffices to take a walk of length

C_G $100C_G$ $200C_G$ $2^{100}C_G$ none of these

5. Let X and Y be random variables on the same sample space, with the property that $X \geq Y$ at all sample points. Circle those, if any, of the following statements that **must** be true about X and Y.

$\mathrm{E}(X) \geq \mathrm{E}(Y)$ $\mathrm{Var}(X) \geq \mathrm{Var}(Y)$ X, Y are not independent

6. Let a and b be n-bit numbers with $a \neq b$. Suppose we want to pick a random m-bit prime p such that $\Pr[a = b \pmod{p}]$ is less than some small constant. Up to constant factors, the value of m we should choose is closest to

a constant $\ln n$ $\sqrt{n}$ n 2^n

7. Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed random variables with expectation $\mathrm{E}(X_i) = 0$ and variance $\mathrm{Var}(X_i) = \sigma^2$. Let $S_n = \sum_{i=1}^{n} X_i$. Circle those **three** of the following statements that **must** hold as $n \to \infty$.

$\Pr[S_n = 0] \to 1$ $\Pr[\frac{S_n}{\sqrt{n}} = 0] \to 1$ $\Pr[\frac{S_n}{n} = 0] \to 1$

$\Pr[S_n > 0] - \Pr[S_n < 0] \to 0$ $\Pr[S_n > \sigma\sqrt{n}] \to$ a constant $\Pr[S_n \geq 0] \to \frac{1}{2}$

8. $3n$ balls are tossed at random into n bins.

(a) As $n \to \infty$, the probability that the first bin is empty is

0 $\frac{1}{n}$ $\frac{1}{3n}$ e^{-1} e^{-3} $\frac{1}{3}$

(b) Let P denote the solution to part (a). As $n \to \infty$, the probability that the first bin contains exactly one ball is

P $\frac{1}{3}P$ $3P$ $9P$ $\frac{9}{2}P$ $\frac{1}{\mathrm{e}}P$

9. We have a group of five people. Each person picks, uniformly and independently at random, a number between 1 and 100. For each pair of distinct people i, j, let X_{ij} be the indicator r.v. of the event "i, j both choose the same number", and for each triple of distinct people i, j, k, let X_{ijk} be the indicator r.v. of the event "i, j, k all choose the same number." Circle those of the following pairs of r.v.'s that are independent (assuming that the indices i, j, k, ℓ, m are all distinct):

X_{ij} & $X_{k\ell}$ $\qquad$ X_{ij} & X_{ik} $\qquad$ X_{ijk} & X_{ij}

X_{ijk} & $X_{i\ell}$ $\qquad$ X_{ijk} & $X_{ij\ell}$ $\qquad$ X_{ijk} & $X_{i\ell m}$

10. Consider the following skip list over the universe of integer keys, that stores the set $\{1, 2, 4, 10\}$.

(a) Draw the new skip list that results from inserting the element 7 with level 2.

(b) Draw the new skip list that results from deleting the element 1 from the *original* list.

[continued overleaf]

11. Universal 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. 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*, and explain briefly why this is a desirable property in applications to the dynamic dictionary problem.

(b) If $\mathcal{H}$ is the family of *all* functions from U to T, then $\mathcal{H}$ is 2-universal. Why is this family nevertheless not a good choice in such applications?

(c) Give an example of a 2-universal family of hash functions of size only $O(m^2)$. [Note: you are **not** required to prove that your family is 2-universal.]

(d) The family $\mathcal{H}$ is said to be *unbiased* if it satisfies, for all $x \in U$ and $z \in T$,

$$\Pr_{h \in \mathcal{H}}[h(x) = z] = \frac{1}{n}.$$

Give a simple example of a small family of efficiently computable hash functions that is unbiased but which has disastrous behavior in hashing applications.

[continued overleaf]

12. Estimating a failure probability

You are given a device which has a tendency to fail randomly. Each time you activate it, it either functions correctly or fails; the behavior on different activations is assumed to be independent. Your task is to estimate the failure probability p to good accuracy, using only these activation tests (i.e., you are not permitted to take the device apart or do other such things).

(a) Suppose you perform a single test. Define the r.v.

$$X = \begin{cases} 1 & \text{if the test fails;} \\ 0 & \text{otherwise.} \end{cases}$$

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

(b) Now suppose you perform n tests. Define the r.v.

$$X_i = \begin{cases} 1 & \text{if the } i\text{th test fails;} \\ 0 & \text{otherwise.} \end{cases}$$

Then you output as your estimate of p the value $Y = \frac{1}{n}\sum_{i=1}^{n} X_i$. What are the values of $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$? [Hint: Recall that, for any r.v. Z, $\mathrm{Var}(cZ) = c^2\mathrm{Var}(Z)$.]

(c) Now suppose you want to choose the sample size n so that your estimate is within a ratio $(1 \pm \epsilon)$ of p with probability at least $1 - \delta$, where ϵ and δ are parameters. In other words, you want to choose n so that

$$\Pr[\,|Y - p| \geq \epsilon p] \;\leq\; \delta.$$

Show, using Chebyshev's inequality, that it suffices to take $n \geq \frac{1}{p\epsilon^2\delta}$. [Note: You should state Chebyshev's inequality precisely.]

[continued overleaf]

Q12 continued

(d) Recall that Chernoff's bound states that, for a sequence of independent coin tosses in which the expected total number of heads is μ, the number of heads X satisfies $\Pr[|X - \mu| \geq d\mu] \leq 2\mathrm{e}^{-d^2\mu/3}$, for $0 < d < 1$. Use this to show that the lower bound on n in part (c) can be reduced to $\frac{3}{p\epsilon^2}\ln(\frac{2}{\delta})$.

(e) In practice, of course, you would not know the value of p in advance, so the bounds on n from parts (c) and (d) cannot be used directly. Suggest how you would overcome this problem in practice.

(f) What do the bounds in parts (c) and (d) tell you about the difficulty of estimating p for a device that fails extremely rarely?

[continued overleaf]

13. Random walks

(a) In *any* connected undirected graph G with n vertices, for *any* two vertices u, v, show that the expected hitting time $\mathcal{H}_{uv}$ from u to v in a random walk on G satisfies $H_{uv} \leq n^3$. [Hint: consider the upper bound on the cover time proved in class.] Also, give an example of a graph for which this bound is tight up to a constant factor. [Note: you are **not** required to prove this tightness.]

The remainder of this question concerns the following undirected graph G, which has n vertices and $m = 2n - 3$ edges (consisting of a line of length n plus an edge from each vertex to vertex 1):

(b) Show that $H_{n1} \leq 3$. [Hint: compare this hitting time to a suitable coin-tossing experiment.]

(c) Show that $H_{1n} \leq 6n$. [Hint: use a coin-tossing argument similar to part (a), but considering *two* steps of the walk at a time.]

(d) Show that the effective resistance R_{1n} between 1 and n is at least $\frac{1}{2}$. Hence deduce from part (b) that $H_{1n} \geq 2n - 6$ (and hence, together with part (c), H_{1n} must be linear in n).

[continued overleaf]

14. Knowing all the right people

Consider the following scenario. We have a large group of n people, some of whom know each other. We call a person *influential* if he/she knows at least $\frac{n}{100}$ other people (i.e., one person in 100). Since we would like to have access to all influential people, we'd like to find a small set S of people so that, for every influential person, there is someone in S who knows him/her. We'll call such a set S a *covering set.*

(a) Suppose we construct S at random as follows: for each of the n people independently, flip a coin with heads probability p. If the coin comes up heads, put that person in S. What is $\mathrm{E}(|S|)$, the expected size of the set S?

(b) Let x be some particular influential person. For S constructed randomly as in part (a), show that the probability that nobody in S knows x is at most $(1-p)^{n/100}$.

(c) Determine a value for p such that the probability in (b) is at most $\frac{1}{3n}$. [Hint: recall that $(1-\frac{t}{m})^m \leq \mathrm{e}^{-t}$ for all $m, t > 0$.]

(d) Deduce that, with this value of p, the set S is a covering set with probability at least $\frac{2}{3}$.

(e) Deduce from parts (a) and (d) that there exists a covering set S of size at most $200\ln(3n)$. [Hint: Apply Markov's inequality to the r.v. $|S|$.]

[continued overleaf]

15. Randomness and Security

In this question, we first establish (in part (a)) a simple fact about random integers modulo some fixed m. Then we explore two applications of this fact to problems arising in security: storing passwords safely (part (b)), and sharing limited information (part (c)).

(a) For a positive integer m, let $\mathbb{Z}_m = \{0, 1, \ldots, m-1\}$ denote the set of integers mod m. Let $a \in \mathbb{Z}_m$ be a fixed integer mod m, and let r be chosen uniformly at random from $\mathbb{Z}_m$. Prove that the number $d = (a - r) \bmod m$ is a uniform random element of $\mathbb{Z}_m$. [Note: aim for a precise and convincing proof—avoid woolly arguments.]

(b) You choose a new 8-character password w for your computer account. You want to record it somewhere in case you forget it, but this is dangerous since it might fall into the wrong hands. Here is a better idea: choose an 8-character word r uniformly at random. Viewing w and r as 8-digit numbers in base b (where b is the size of the character set), compute the 8-digit number $d = (w - r) \bmod m$, where $m = b^8$. Then write down the numbers r and d in separate places.

Show that this scheme has the following properties:

(i) If you forget your password, you can easily recover it from the recorded numbers r and d.

(ii) If a malicious intruder obtains either the number r or the number d, but not both, then he gains *no information* at all about your password w. [Hint: Use part (a).]

[continued overleaf]

Q15 continued

(c) Three students, S_1, S_2, S_3, are given grades g_1, g_2, g_3 in the range $[0..100]$. They want to compute the *sum* of their grades without revealing their actual grades to one another. They come up with the following protocol, **which you should read and digest carefully before proceeding**.

```
set $m = 301$
each student $S_i$ selects $r_i \in \mathbb{Z}_m$ independently and uniformly at random
    $S_1$ passes $t_1 = r_1$ to $S_2$
    $S_2$ passes $t_2 = (t_1 + r_2) \bmod m$ to $S_3$
    $S_3$ passes $t_3 = (t_2 + r_3) \bmod m$ to $S_1$
each student $S_i$ computes $d_i = (g_i - r_i) \bmod m$
    $S_1$ passes $t_4 = (t_3 + d_1) \bmod m$ to $S_2$
    $S_2$ passes $t_5 = (t_4 + d_2) \bmod m$ to $S_3$
    $S_3$ passes $t_6 = (t_5 + d_3) \bmod m$ to $S_1$
$S_1$ announces “sum of grades $= t_6$”
```

Show that this protocol solves the problem, by verifying the following two properties:

(i) The number announced by S_1 is correct.

(ii) At the end of the protocol, no student has any information, except for the sum of the grades, that he could not have computed himself at the beginning of the protocol. [Note that if a student simply receives a random number from somebody else then this is no new information, since he could have generated a random number himself.]