CS–174 Combinatorics & Discrete Probability, Spring 2025

Final Exam Solutions

11:30am–2:30pm, 14 May

Read these instructions carefully

1. **Write your name and SID number on the front page, and your SID number on every page!**
2. This is a **closed book** exam, but you are allowed one two-sided cheat sheet and blank scratch paper. No phones, calculators or other electronic equipment.
3. The exam consists of 13 questions. The first 8 questions are multiple choice; the remaining 5 require written answers.
4. Approximate point totals for each question part are indicated in the margin. The maximum total number of points is 100.
5. **Multiple choice questions 1–7:** Answer these by **filling in** the **single** circle adjacent to the correct answer; there is **no** penalty for incorrect answers (unless you select more than one answer).
6. **Multiple choice question 8:** Answer this by **filling in** the squares adjacent to **all** the answers that apply. Points will be allocated both for selecting correct answers and for not selecting incorrect answers, so you should select only those that apply.
7. **Other questions 9–13:** Write your answers to these in the spaces provided below them. None of these questions requires a very long answer, so you should have enough space—if not you are probably writing too much. **Always show your working for these questions!**
8. The questions vary in difficulty: if you get stuck on some part of a question, leave it and go on to the next one. Point totals and space provided are not necessarily an indication of difficulty.

Your First Name: **Your Last Name:**

Your SID Number:

[exam starts on next page]

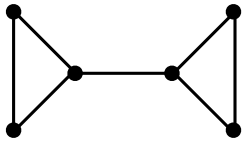
1. John is waiting for a bus, whose arrival time T has a geometric distribution with parameter $\frac{1}{10}$. The bus picks up, on average, one passenger every two units of time, so that when it arrives the number of passengers has a Poisson distribution with parameter $\frac{T}{2}$. The expected number of passengers on the bus when John boards is *3pts*

● 5; ○ 10; ○ 20; ○ $e^{1/5}$; ○ $\frac{1}{20}$

Let X denote the number of passengers when John boards. Then $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|T]] = \mathrm{E}[\frac{T}{2}] = \frac{1}{2}\mathrm{E}[T] = 5$.

---

2. Consider the simple network below.



If each link (edge) fails independently with probability $\frac{1}{2}$, the probability that the network remains connected is *3pts*

● $\frac{1}{8}$; ○ $\frac{1}{4}$; ○ $\frac{39}{128}$; ○ $\frac{1}{3}$; ○ $\frac{1}{2}$

The event that the network is connected is equivalent to the middle edge not failing and at most one of the edges in each of the triangles failing. These events are all independent, and each occurs with probability $\frac{1}{2}$. So the overall probability is $(\frac{1}{2})^3 = \frac{1}{8}$.

---

3. Suppose n balls are thrown randomly into 1000 bins.

(a) What is the smallest among the following values of n for which you would expect some bin to have more than one ball? *3pts*

○ 10; ● 40; ○ 120; ○ 280; ○ 500

This is the birthday problem, with $n = 1000$ possible birthdates. We'd expect a collision to occur when the number of people is a small constant factor times $\sqrt{n}$, so 40 is the best choice here.

(b) What is the smallest among the following values of n for which you would expect there to be no empty bins? *3pts*

○ 1500; ○ 3000; ● 7000; ○ 15000; ○ 1000000

This is coupon collecting with $n = 1000$ coupons. We'd expect to collect all coupons after buying about $n \ln n$ boxes, which is around 7000.

---

4. Let Z be a discrete uniform random variable taking integer values in the set $\{a, a+1, \ldots, b-1\}$. The MGF $\mathcal{M}_Z(t)$ of Z is *3pts*

○ $\frac{(a+b-1)t}{2}$; ● $\frac{e^{bt} - e^{at}}{(b-a)(e^t - 1)}$; ○ $\frac{e^{(b-a)t}}{(b-a)(e^t-1)}$; ○ $\frac{e^{bt} - e^{at}}{(b-a)t}$

By definition, $\mathcal{M}_Z(t) = \mathrm{E}[e^{tZ}] = \frac{1}{b-a}\sum_{i=a}^{b-1} e^{ti} = \frac{e^{ta}}{b-a} \cdot \frac{e^{t(b-a)}-1}{e^t-1} = \frac{e^{bt}-e^{at}}{(b-a)(e^t-1)}$.

---

5. Consider two random variables, X and Y on a common probability space, whose distributions are as follows:

$$\Pr[X=0]=\tfrac{1}{3} \quad \Pr[X=1]=\tfrac{1}{3} \quad \Pr[X=2]=\tfrac{1}{3};$$
$$\Pr[Y=0]=\tfrac{1}{4} \quad \Pr[Y=1]=\tfrac{1}{2} \quad \Pr[Y=2]=\tfrac{1}{4}.$$

(a) If X and Y are coupled independently, then $\Pr[X=Y]$ is *3pts*

○ $\frac{1}{12}$; ○ $\frac{1}{4}$; ● $\frac{1}{3}$; ○ $\frac{1}{2}$; ○ $\frac{2}{3}$

If X, Y are independent, then $\Pr[X=Y]=\sum_i \sum_j \Pr[X=i]\Pr[Y=j]=\frac{1}{3}\sum_j \Pr[Y=j]=\frac{1}{3}$.

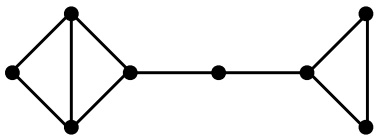
(b) If X and Y are coupled optimally, then $\Pr[X=Y]$ is *3pts*

○ $\frac{1}{2}$; ○ $\frac{2}{3}$; ○ $\frac{3}{4}$; ● $\frac{5}{6}$; ○ 1

$\Pr[X=Y]=\sum_i \min\{\Pr[X=i],\Pr[Y=i]\}=\frac{1}{4}+\frac{1}{3}+\frac{1}{4}=\frac{5}{6}$.

---

6. Consider the graph H below.



(a) According to a first moment analysis, the candidate value of $p=p(n)$ that is the threshold for the existence of a (not necessarily induced) copy of H in a random graph $G \in \mathcal{G}_{n,p}$ is *3pts*

○ n^{-10}; ○ n^{-1}; ○ $n^{-2/3}$; ● $n^{-4/5}$; ○ 1

H has 8 vertices and 10 edges, so by the usual counting argument $\mathrm{E}[X]=\Theta(n^8p^{10})$, where X is the number of copies of H in G. This is $\Theta(1)$ when $p=\Theta(n^{-4/5})$, so the first moment analysis suggests that this is the threshold.

(b) Let the r.v. X denote the number of copies of H in $G \in \mathcal{G}_{n,p}$. Which of the following additional properties guarantees that the value you chose in part (a) is indeed a threshold? *3pts*

○ $\frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2}\to\infty$; ● $\frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2}\to 0$; ○ $\frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2}\to 1$; ○ $\frac{\mathrm{E}[X^2]}{\mathrm{E}[X]^2}\to 0$

To verify that $p(n)$ is a threshold, we need to show that the second moment is suitably bounded, or specifically that $\frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2}\to 0$. None of the other options implies this.

---

7. Suppose we specify a Markov chain on the non-negative integers using three parameters p, q, r, as follows. For all $i > 0$, parameter p denotes the transition probabilities $P(i, i+1)$; q denotes $P(i, i-1)$; and r denotes $P(i, 0)$. Always we have $P(0, 1) = 1$. For each of the following cases, specify whether the Markov chain is transient, null recurrent or positive recurrent. Shade exactly one circle in each case.

	transient	null recurrent	pos. recurrent
(a) $p = q = \frac{1}{2};\ r = 0$	○	●	○
(b) $p = \frac{2}{3};\ q = 0;\ r = \frac{1}{3}$	○	○	●
(c) $r = 0;\ (p, q) = \begin{cases} (\frac{1}{3}, \frac{2}{3}) & \text{if } 0 < i \leq 100; \\ (\frac{2}{3}, \frac{1}{3}) & \text{if } i > 100. \end{cases}$	●	○	○

All three properties are class properties (i.e., the same for all communicating states), so it's enough to consider the behavior of walks starting at 0. (a) is standard symmetric random walk (with reflecting barrier at 0), which we have seen in class is null recurrent. (b) is random walk with a drift away from 0, but with a "reset" back to 0 from every state: since this walk has constant probability of returning to 0 from every state, it is clearly positive recurrent. (c) has drift towards 0 up to state 100, and drift away from 0 thereafter; once this walk reaches state 101 (which happens with probability 1), there is a non-zero probability that it never returns, so it is transient.

---

8. Note: In Question 8 only, shade the squares adjacent to all solutions that apply. Points will be awarded both for selecting the correct solutions and for not selecting the incorrect ones. *6pts*

Let $X_1, ..., X_n$ be *pairwise independent* 0-1 valued random variables with $\Pr[X_i = 1] = \frac{1}{4}$ for all i, and let $Y = \sum_{i=1}^{n} X_i$. Which of the following statements **must** be true? [Choose all that apply.]

□ $\Pr[X_1 = X_2 = X_3] = \frac{7}{16}$; ■ $\Pr[Y \geq n/2] \leq \frac{3}{n}$; □ $\Pr[Y \geq n/2] \leq e^{-n/12}$;

■ $\Pr[X_1 = X_2 = X_3 = 1] \leq \frac{1}{16}$; □ $\Pr[Y = 0] = (\frac{3}{4})^n$; ■ $\Pr[Y \leq n/8] \leq \frac{6}{7}$

$\Pr[Y \geq n/2] \leq \frac{3}{n}$ follows from Chebyshev, observing that $\mathrm{E}[Y] = \frac{n}{4}$. $\Pr[X_1 = X_2 = X_3 = 1] \leq \Pr[X_1 = X_2 = 1] = \frac{1}{16}$ by pairwise independence of X_1, X_2. And $\Pr[Y \leq n/8] = \Pr[(n - Y) \geq 7n/8] \leq \frac{6}{7}$ by Markov, observing that $(n - Y)$ is a non-negative r.v. with expectation $\frac{3n}{4}$.

All the other statements assume more than pairwise independence: $\Pr[X_1 = X_2 = X_3] = \frac{7}{16}$ assumes at least 3-wise independence; $\Pr[Y \geq n/2] \leq e^{-n/12}$ is a Chernoff bound, which assumes mutual independence; and similarly, $\Pr[Y = 0] = (\frac{3}{4})^n$ assumes mutual independence.

[continued overleaf]

9. Counting Particles [13pts total]

In a physics experiment, there is a large reservoir of particles, some of which are so-called ζ-*particles*. Your task is to estimate the proportion p of ζ-particles in the reservoir. You are provided with a detection device that works as follows: particles are streamed, one per unit time, each chosen u.a.r. (with replacement) from the reservoir, until the first ζ-particle appears in the stream. The device outputs the time of that first appearance and then shuts off.

(a) Let $X_1, \ldots, X_t$ be a sequence of independent outputs from the device, and let $X = \sum_{i=1}^{t} X_i$. What is the expectation $\mathrm{E}[X]$? Briefly explain your answer. *2pts*

$\mathrm{E}[X] = \frac{t}{p}$. Each X_i is Geometric with parameter p, so $\mathrm{E}[X_i] = \frac{1}{p}$ and $\mathrm{E}[X] = \sum_i \mathrm{E}[X_i] = \frac{t}{p}$.

(b) For any positive integer m, show that $\Pr[X > m] = \Pr[Y < t]$, for a suitable binomial r.v. Y. You should also specify the parameters of Y. *2pts*

The event $X > m$ corresponds to there being more than m trials in total before the tth success, where the success probability for each independent trial is p. This is equivalent to there being fewer than t successes in the first m trials, which in turn is equivalent to a Binomial(m, p) r.v. being smaller than t. Hence we get the desired equality for $Y \sim \mathrm{Bin}(m, p)$.

(c) Let μ denote the expectation in part (a). Use part (b) to show that $\Pr[X > (1+\varepsilon)\mu] \leq \exp(-\frac{\varepsilon^2 t}{6})$. [HINT: For a binomial r.v. Z with $\mathrm{E}[Z] = \nu$, you may assume the Chernoff bound in the form $\Pr[Z < (1-\alpha)\nu] \leq \exp(-\frac{\alpha^2 \nu}{3})$.] *5pts*

By part (b), we have $\Pr[X > (1+\varepsilon)\mu] = \Pr[Y < t]$, where $Y \sim \mathrm{Bin}((1+\varepsilon)\mu, p)$. Since $\mu = \frac{t}{p}$ by part (a), we have that $\mathrm{E}[Y] = (1+\varepsilon)t$. Note that the event $Y < t$ corresponds to Y being below its mean by εt, which is $\frac{\varepsilon t}{(1+\varepsilon)t} = \frac{\varepsilon}{1+\varepsilon}$ times $\mathrm{E}[Y]$. Thus, applying the lower tail of the Chernoff bound to Y with $\nu = \mathrm{E}[Y] = (1+\varepsilon)t$ and $\alpha = \frac{\varepsilon}{1+\varepsilon}$, we get that

$$\Pr[X > (1+\varepsilon)\mu] = \Pr[Y < t] \leq \exp\Big(-\big(\frac{\varepsilon}{1+\varepsilon}\big)^2 \frac{(1+\varepsilon)t}{3}\Big) = \exp\Big(-\frac{\varepsilon^2 t}{3(1+\varepsilon)}\Big) \leq \exp\Big(-\frac{\varepsilon^2 t}{6}\Big).$$

Note: There was some confusion about how to compute the deviation parameter $\alpha = \frac{\varepsilon}{1+\varepsilon}$ for use in the Chernoff bound. This requires you to compute the expectation $\mathrm{E}[Y] = (1+\varepsilon)t$ and then use the equation $t = (1-\alpha)\mathrm{E}[Y]$ to compute α.

(d) Assuming the same bound as in part (c) for the lower tail $\Pr[X < (1-\varepsilon)\mu]$, explain how to use t independent observations from the detection device to produce an estimate $\hat{p}$ of p satisfying the condition *3pts*

$$\Pr[|\hat{p} - p| \geq \varepsilon p] \leq \delta, \qquad (*)$$

for user-specified parameters $\varepsilon, \delta \in (0, 1/2]$. You should specify the number of observations, t, as a function of ε and δ. [HINT: You may assume that if $\hat{Z}$ estimates Z within a factor $(1 \pm \varepsilon)$ then $\frac{1}{\hat{Z}}$ estimates $\frac{1}{Z}$ within a factor $(1 \pm 2\varepsilon)$ when $\varepsilon \in (0, 1/2]$.]

By part (a), $\mathrm{E}[X] = \frac{t}{p}$. Therefore our estimator of p will be $\hat{p} = \frac{t}{X}$. By part (c) and the given bound on the lower tail, if we set $t \geq \frac{6}{\varepsilon^2} \ln(\frac{2}{\delta})$ then with probability at least $1 - \delta$, X will lie within $(1 \pm \varepsilon)$ of its mean, $\frac{t}{p}$. This latter fact ensures that $\hat{p}$ lies within $(1 \pm 2\varepsilon)$ of p (since $(1-\varepsilon)^{-1} < 1 + 2\varepsilon$ (assuming $\varepsilon \leq \frac{1}{2}$) and $(1+\varepsilon)^{-1} \geq 1 - \varepsilon$). So replacing ε by $\varepsilon/2$ in the earlier analysis achieves the required condition $(*)$ (and increases our lower bound on t to $t \geq \frac{24}{\varepsilon^2} \ln(\frac{2}{\delta})$).

Note: Some students omitted to observe that we have to work with the estimator $\frac{t}{X}$ rather than X itself.

(e) What is the total expected amount of time taken by the experiment (measured as total number of time steps used by the detection device)? *1pt*

From part (d) we need $t = \frac{24}{\varepsilon^2}\ln(\frac{2}{\delta})$ observations. Since each observation takes expected time $\frac{1}{p}$, the total expected time is $\frac{24}{p\varepsilon^2}\ln(\frac{2}{\delta})$.

[continued overleaf]

10. Two-dimensional robot [12pts total]

A robot walks on a rectangular grid in two dimensions, starting at position $(X_0, Y_0) = (0, 0)$. At each step, the robot chooses one of the four possible grid directions u.a.r. and takes one step in that direction. Let (X_t, Y_t) denote the position of the robot at time t. The robot stops after the first step t at which its squared distance $X_t^2 + Y_t^2$ from its starting point is at least R^2. Your goal is to compute the expected number of steps until this happens.

(a) Show that the process $Z_t := X_t^2 + Y_t^2 - t$ is a martingale w.r.t. the sequence $((X_0, Y_0), \ldots, (X_t, Y_t))$. *4pts*
[HINT: It is much easier to work with the martingale differences $D_t = Z_t - Z_{t-1}$.]

Clearly Z_t is finite for all t, and the value of Z_t depends only on X_t, Y_t. For the main martingale property, as suggested in the hint we work with the difference $D_t = X_t - X_{t-1}$. Note in particular that if the robot makes a move in the increasing X-direction, then $D_t = (X_{t-1}+1)^2 - X_{t-1}^2 - 1 = 2X_{t-1}$, and similarly for the other three moves. We therefore have

$$\mathrm{E}[D_t | X_0, Y_0, \ldots, X_{t-1}, Y_{t-1}] = \frac{1}{4}\Big[2X_{t-1} - 2X_{t-1} + 2Y_{t-1} - 2Y_{t-1}\Big] = 0,$$

as required.

Several students forgot to mention the facts that Z_t must be a function of only $(X_0, Y_0), ..., (X_t, Y_t)$ and that $\mathrm{E}[|Z_t|]$ must be finite.

(b) Let T denote the first step at which $X_T^2 + Y_T^2 \geq R^2$. Use the Optional Stopping Theorem to derive an expression for $\mathrm{E}[T]$ in terms of $\mathrm{E}[X_T^2 + Y_T^2]$. [NOTE: Be sure to justify why the OST applies in this situation.] *4pts*

Clearly T is a valid stopping time. To justify using the OST, we will show that $\mathrm{E}[T]$ is bounded and that the martingale differences $|D_t|$ are bounded. Note that T is the time for a random walk on a finite graph to hit a certain set of vertices, which we know has finite expectation. To see that the differences are bounded, note from part (a) that $D_t \in \{\pm 2X_{t-1}, \pm 2Y_{t-1}\}$, so $|D_t| \leq 2R$.

Now the OST implies that $\mathrm{E}[Z_T] = Z_0 = 0$, and hence $\mathrm{E}[T] = \mathrm{E}[X_T^2 + Y_T^2]$.

Note: Several students claimed that the sequence $|Z_t|$ is bounded, but this is not true because Z_t contains the term t, which is not bounded.

(c) Finally, use part (b) to show that $\mathrm{E}[T] = \Theta(R^2)$. [NOTE: This requires you to show both an upper and a lower bound for $\mathrm{E}[T]$.] *4pts*

From part (b), it suffices to show that $X_T^2 + Y_T^2 = \Theta(R^2)$. The lower bound is immediate, since by definition $X_T^2 + Y_T^2 \geq R^2$. For the upper bound, note that the position of the robot at time $T-1$ must satisfy $X_{T-1}^2 + Y_{T-1}^2 < R^2$. Moreover, we know that the next move can increase this expression by at most $2\max\{|X_{T-1}|, |Y_{T-1}|\} + 1 \leq 2R + 1$. Hence we deduce that $X_T^2 + Y_T^2 < R^2 + 2R + 1$. Combining this with the lower bound gives $X_T^2 + Y_T^2 = \Theta(R^2)$ (or more precisely $R^2 + O(R)$).

An alternative argument for the upper bound is that, at the stopping time, we must have both $X_T \leq R$ and $Y_T \leq R$ (else we would have stopped earlier). This implies that $X_T^2 + Y_T^2 \leq 2R^2$, which is a weaker bound than that above but still $O(R^2)$.

[continued overleaf]

11. MaxCut in Random Graphs [10 pts total]

Let $G \in \mathcal{G}_{n,1/2}$ be a random graph in the $\mathcal{G}_{n,p}$ model with $p = \frac{1}{2}$, and let $f(G)$ denote the size of a maximum cut in G. (Recall that a maximum cut is a partition of the vertices of G into two parts that maximizes the number of edges between vertices in the two parts.) Let $\mathrm{E}[f]$ denote the expected value of $f(G)$ over the choice of random graph G.

(a) Suppose we view $f(X_1, \ldots, X_n)$ as a function of variables X_i, one for each vertex i, which specifies the edges between i and vertices $j > i$. Using the associated Doob martingale, derive a bound based on Azuma's inequality for the probability $\Pr[|f - \mathrm{E}[f]| \geq \lambda]$. Explain clearly how your bound is derived. You will not be penalized for incorrect constants. *4pts*

We define the Doob martingale $Z_i := \mathrm{E}[f \mid X_1, \ldots, X_i]$ with $Z_0 = \mathrm{E}[f]$ and $Z_n = f(G)$. We claim that f is n-Lipschitz, since changing all edges incident on a single vertex can affect the size of the maximum cut by at most $\pm n$. Azuma's inequality with bounded differences $c_i = n$ then gives

$$\Pr[f - \mathrm{E}[f] \geq \lambda] \leq \exp\Big(-\frac{\lambda^2}{2\sum_{i=1}^{n} c_i^2}\Big) = \exp\Big(-\frac{\lambda^2}{2n^3}\Big).$$

Note: A point was deducted for failing to define the Doob martingale.

(b) Now view $f(Y_1, \ldots, Y_{\binom{n}{2}})$ as a function of variables Y_j, one for each possible edge j, which specifies the presence or absence of that edge. Repeat part (a) for the Doob martingale associated with this representation of f. *4pts*

We use the Doob martingale $Z_i := \mathrm{E}[f \mid Y_1, \ldots, Y_i]$. In this case the function f is 1-Lipschitz, since changing a single edge in G can only change the size of the maximum cut by ± 1. Azuma's inequality with all $c_1 = 1$ now gives

$$\Pr[f - \mathrm{E}[f] \geq \lambda] \leq \exp\Big(-\frac{\lambda^2}{2\sum_{i=1}^{\binom{n}{2}} c_i^2}\Big) = \exp\Big(-\frac{\lambda^2}{n(n-1)}\Big).$$

(c) It can be shown that $\mathrm{E}[f] \leq \frac{n^2}{8} + cn^{3/2}$, for some constant $c > 0$. Are either of the bounds in parts (a) or (b) useful as a large-deviation bound for f? Briefly explain your answer. *2pts*

Part (a) implies that deviations from the mean of size $\omega(n^{3/2})$ are unlikely for f, while part (b) implies the same for deviations of size $\omega(n)$. Since we are told that the value of $\mathrm{E}[f]$ is known up to factors of order $n^{3/2}$, the bound from part (a) is not useful while that from part (b) does imply concentration of f.

For full credit on this part, you were required to specify the width of the concentration bound in both parts (a) and (b). Just claiming that the bound in (b) is better than that in (a) is not enough.

[continued overleaf]

12. Pairwise independence [15pts total]

A certain randomized algorithm requires a supply $X_1, X_2, \ldots, X_t$ of random bits with $\Pr[X_i = 1] = p$ for some p in the range $\frac{1}{4t} \le p \le \frac{3}{4t}$, which satisfy the following property:

- $\Pr[\sum_{i=1}^{t} X_i > 0] \ge c$ for some constant c (independent of t).

(Note that the bits are not required to be independent.)

(a) Show that if the X_i are *mutually independent* random bits with $\Pr[X_i = 1] = p$, then they satisfy the above property. *3pts*

If the X_i are mutually independent, then

$$\Pr[\sum_i X_i > 0] = 1 - (1-p)^t \ge 1 - \exp(-pt) \ge 1 - \exp(-1/4).$$

So the required property holds with $c = 1 - \exp(-1/4)$.

(b) Explain how you would use a source of mutually independent *uniform* random bits to construct the sequence $X_1, \ldots, X_t$ in part (a). Your construction should require only $O(t)$ bits in expectation. *3pts*

The idea is to take a sequence of $c = \Theta(\log t)$ uniform random bits and set $X_i = 1$ iff all the bits are 1. Then $\Pr[X_i = 1] = 2^{-c} = \Theta(\frac{1}{t})$. To ensure that p lies in the desired range, we can take $c = \lceil \log(\frac{4t}{3}) \rceil$. For the implementation, we can cut off the sequence (and set $X_i = 0$) as soon as the first 0 appears, so the expected number of bits needed to produce each X_i is 2, for a total $O(t)$ expected bits.

Note: Many students assumed that the value of p is given; the intended meaning of the problem is that we are free to use any p in the given range. This doesn't change the problem much, except that the worst-case number of bits may be larger if p requires many bits to specify it–but the expected number of uniform bits needed is still only 2 per biased bit. Students were not penalized for this alternative interpretation.

(c) A software vendor offers to supply random bits $X_1, \ldots, X_t$ with $\Pr[X_i = 1] = p$ that are only *pairwise* independent, at a much lower price than mutually independent bits. Show that these pairwise independent bits also satisfy the above property $\Pr[\sum_{i=1}^{t} X_i > 0] \ge c$, for a possibly different constant c. [HINT: Recall the principle of inclusion-exclusion: for events $E_1, E_2, \ldots$, we have $\Pr[\cup_i E_i] = \sum_i \Pr[E_i] - \sum_{i<j} \Pr[E_i \cap E_j] + \sum_{i<j<k} \Pr[E_i \cap E_j \cap E_k] - \cdots$.] *6pts*

By inclusion-exclusion, we have

$$\begin{aligned} \Pr[\sum X_i > 0] &\ge \sum_i \Pr[X_i = 1] - \sum_{i<j} \Pr[X_i = 1 \wedge X_j = 1] \\ &= t \cdot p - \binom{t}{2} \cdot p^2 = tp\Big(1 - \frac{(t-1)p}{2}\Big) \ge tp\Big(1 - \frac{tp}{2}\Big). \end{aligned}$$

Here we used pairwise independence to deduce $\Pr[X_i = 1 \wedge X_j = 1] = p^2$. Note that $tp \in [\frac{1}{4}, \frac{3}{4}]$. The function $z(1 - \frac{z}{2})$ is monotonically increasing in the interval $[\frac{1}{4}, \frac{3}{4}]$, so it is minimized at $z = \frac{1}{4}$. Hence the above calculation gives $\Pr[\sum_i X_i > 0] \ge \frac{7}{32}$, so the property holds with $c = \frac{7}{32}$.

Note: Many students attempted to apply inclusion exclusion to the events $X_i = 0$ instead of $X_i = 1$. This doesn't work as you are asked to bound the probability of the event $\sum_i X_i > 0$, which is equivalent to the union of the events $X_i = 1$. Another, less common error was to try to use Markov's or Chebyshev's inequality, neither of which is strong enough for this problem.

(d) Describe how the software vendor might produce the t pairwise independent random bits with bias in the range $\frac{1}{4t} \leq p \leq \frac{3}{4t}$ from a much smaller number of *mutually* independent *uniform* random bits. [NOTE: You do **not** need to justify your construction, but you should specify it clearly.] *3pts*

Following a construction we saw in class, we can take $m = \Theta(\log t)$ mutually independent uniform bits $x_1, \ldots, x_m$ and convert them into $2^m - 1 = \Theta(t)$ pairwise independent uniform bits y_i via $y_i := \oplus_{j \in S_i} x_i$, where S_i ranges over all non-empty subsets of $\{1, \ldots, m\}$. Using the same technique as in part (b), we can convert these into random bits with bias $p \in [\frac{1}{4t}, \frac{3}{4t}]$ with overhead only $O(\log t)$ (or constant in expectation). Thus the entire construction requires only $m = O(\log t)$ mutually independent bits.

An alternative construction is the following. Choose a prime q in the range $\frac{4t}{3} < q < 4t$, and two independent random integers $a, b \bmod q$. Then set $Z_i = ai + b \mod q$ for $0 \leq i \leq q - 1$. We know from class that the Z_i will be pairwise independent and uniform in $\{0, \ldots, q-1\}$. For our random bits, we will use the first t of the Z_i, and set $X_i = 1$ iff $Z_i = q - 1$ (and 0 otherwise). This ensures that $\Pr[X_i = 1] = \frac{1}{q}$, which lies in the desired range $[\frac{1}{4t}, \frac{3}{4t}]$ by choice of q.

[continued overleaf]

13. A Markov chain on subsets [14pts total]

Suppose we have a set of n items labeled $\{1, \ldots, n\}$, and we want to choose a random subset of k items. In this problem we explore how to do this using the following Markov chain whose state space is Ω, the set of all k-subsets of $\{1, \ldots, n\}$. At each step, the Markov chain does the following, where $X \in \Omega$ is the current state:

- pick a pair of items (i, j) u.a.r. with replacement from $\{1, \ldots, n\}$ (note that $i = j$ is possible);
- if $i \in X$ and $j \notin X$, move to the state $(X \setminus \{i\}) \cup \{j\}$ (which for simplicity we denote $X - i + j$), else do nothing.

(a) Show that this Markov chain converges to the uniform distribution on Ω. In your answer, state clearly which properties you need to verify and why each of them holds. *3pts*

We need to show that the chain is irreducible and aperiodic, and that the uniform distribution is stationary.

- For irreducibility, we can build a path between any two sets X, Y as follows. Let $S_X = X \setminus Y$ and $S_Y = Y \setminus X$. Note that $|S_X| = |S_Y|$ since $|X| = |Y|$, so we can pair up the elements of S_X, S_Y in an arbitrary way, say $(x_1, y_1), \ldots, (x_\ell, y_\ell)$. The path from X to Y exchanges each x_i with y_i in sequence.
- For aperiodicity, we just note that the chain contains self-loops (since, e.g., we could pick $i = j$).
- For uniformity, we observe that the transition matrix is symmetric: any transition corresponding to a pair of elements i, j can be reversed by choosing the pair j, i, and all transitions have the same probability $\frac{1}{n^2}$.

Now let (X_t) and (Y_t) be two copies of the Markov chain, and consider the coupling in which X_t, Y_t both pick the same pair (i, j) and then move accordingly. Suppose the current states are $X_t = X$ and $Y_t = Y$, and define $S_X = X \setminus Y$, $S_Y = Y \setminus X$, $I = X \cap Y$ and $R = \{1, \ldots, n\} \setminus (X \cup Y)$. Note that S_X, S_Y, I, R are disjoint and partition $\{1, \ldots, n\}$: you may find it helpful to draw a Venn diagram before proceeding! Also, let $d_t = |S_X| = |S_Y|$; we will use d_t as a measure of the difference between X_t and Y_t.

(b) Show that $d_{t+1} = d_t - 1$ if either $i \in S_X$ and $j \in S_Y$, or $i \in S_Y$ and $j \in S_X$. *2pts*

If $i \in S_X$ and $j \in S_Y$, then X moves to $X - i + j$ (because $i \in X$ and $j \notin X$) while Y does not change (because $i \notin Y$). This decreases the distance between X and Y by 1. If $i \in S_Y$ and $j \in S_X$, the same holds with the roles of X and Y reversed.

[continued overleaf]

(c) Show that in all other cases $d_{t+1} = d_t$. [HINT: By symmetry, it is enough to consider moves that change X. Note that X changes if and only if $i \in X$ and $j \notin X$ (and vice versa for Y). Do a case analysis based on the three cases: (i) $i \in S_X, j \in R$; (ii) $i \in I, j \in S_Y$; (iii) $i \in I, j \in R$. The fourth case $i \in S_X, j \in S_Y$ is covered by part (b).] *3pts*

Following the hint, we consider only moves in which X changes. The situation for moves in which Y changes will be entirely symmetrical. And if neither X nor Y changes then clearly d_t doesn't change.

For moves that change X, note that the three cases in the hint cover all cases except for that covered in part (b).

- Case (i). In this case X moves to $X - i + j$ and Y does not move; but the move of X exchanges one element not in Y with another, so d_t does not change.
- Case (ii). In this case again X moves to $X - i + j$ and Y does not move; but the move of X exchanges one element in Y for another, so again d_t does not change.
- Case (iii). In this case X moves to $X - i + j$ and Y moves to $Y - i + j$; but both X and Y exchange an element in their intersection for an element that is in neither of them, so again d_t does not change.

(d) Deduce from parts (b) and (c) that the expected time for d_t to decrease from d to $d - 1$ is $\frac{n^2}{2d^2}$. *3pts*

From the analysis in parts (b) and (c), we know that d_t is non-increasing, and that it decreases by 1 only when $i \in S_X$ and $j \in S_Y$, or $i \in S_Y$ and $j \in S_X$. This event happens with probability $\frac{1}{n^2}(2|S_X||S_Y|) = \frac{2d_t^2}{n^2}$. Thus the expected time for d_t to decrease from d to $d - 1$ is $\frac{n^2}{2d^2}$.

(e) Deduce from part (d) that the mixing time of this Markov chain is $O(n^2)$. *3pts*

We know from class that the mixing time of the chain is bounded above by the expected time until $X_t = Y_t$ under any coupling. This condition is equivalent to $d_t = 0$. Since $d_0 \leq \frac{n}{2}$, part (d) implies that the expected time for d_t to reach 0 is at most

$$\sum_{d=1}^{n/2} \frac{n^2}{2d^2} = \frac{n^2}{2} \sum_{d=1}^{n/2} \frac{1}{d^2} = O(n^2),$$

since $\sum_{d=1}^{\infty} \frac{1}{d^2}$ is finite. Hence the mixing time is $O(n^2)$.

Note: Several students claimed that the mixing time bound of $O(n^2)$ follows immediately from the result in part (d). This is not the case because you need to calculate the expected time to reduce the distance from $n/2$ to 0, not just to reduce the distance by 1 as in part (d).

[The End]