Instructions that will appear on the real exam

- **DO NOT OPEN THE EXAM UNTIL YOU ARE INSTRUCTED TO.**
- **The last page of this exam is a sheet with some useful formulas and theorem statements.** Feel free to rip it off of the exam.
- Answer all of the questions as well as you can. You have **180 minutes.**
- The exam is **non-collaborative**; you must complete it on your own. If you have any clarification questions, please ask the course staff. We cannot provide any hints or help.
- This exam is **closed-book**, except for **up to three double-sided sheets of paper** that you have prepared ahead of time. You can have anything you want written on these sheets of paper.
- **Please DO NOT separate pages of your exam** (other than the reference sheet on the last page). The course staff is not responsible for finding lost pages, and you may not get credit for a problem if it goes missing.
- There are a few pages of extra paper at the back of the exam in case you run out of room on any problem. If you use them, please clearly indicate on the relevant problem page that you have used them, and please clearly label any work on the extra pages.
- Please make sure to sign out of the roster when handing in your completed exam to the teaching team.
- **Please do not discuss the exam until after solutions are posted!**

General Advice

- If you get stuck on a question or a part, move on and come back to it later. The questions on this exam have a wide range of difficulty, and you can do well on the exam even if you don't get a few questions.
- Pay attention to the point values. Don't spend too much time on questions that are not worth a lot of points.
- There are **100** total points on this exam.

Name and SUNet ID (please print clearly):

SOLUTION

This page intentionally blank. Please do not write anything you want graded here.

Honor Code

The Honor Code is an undertaking of the Stanford academic community, individually and collectively. Its purpose is to uphold a culture of academic honesty. Students will support this culture of academic honesty by neither giving nor accepting unpermitted academic aid on this examination.

This course is participating in the proctoring pilot overseen by the Academic Integrity Working Group (AIWG), therefore proctors will be present in the exam room. The purpose of this pilot is to determine the efficacy of proctoring and develop effective practices for proctoring in-person exams at Stanford.

Unpermitted Aid on this exam includes but is not limited to the following: collaboration with anyone else; reference materials other than your cheat-sheet (see below); and internet access.

Permitted aid on this exam includes a "cheat-sheet:" three double-sided sheets of paper with anything written on them, which you have prepared yourself ahead of time.

I acknowledge and attest that I will abide by the Honor Code:

[signed] ____

Exam Break Sign-out

I pledge that during my exam break:

- I will not bring any paper, electronic devices (phone, smart watch, smart glasses, etc), or aid (permitted or unpermitted) *out of or into* the exam room.
- I will not communicate with anyone other than the course instructional staff about the content of the exam.

Signature Confirming Honor Code Pledge	Exit Time	Return Time	Proctor Initial	Length (min)

If you are feeling unwell and are not able to complete the exam, please connect with the proctor to discuss options.

Good Luck!

This page intentionally blank. Please do not write anything you want graded here.

1. **(20 pt.)** For each of the parts below, select the *best bound* you can to correctly fill in the blank.

   (a) **(5 pt.)** Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables in $\{0,1\}$, with $\mathbb{E}[Y_i] = p$, and let $S = \sum_{i=1}^n Y_i$. Suppose that $\Pr[S \geq pn + t] \leq 0.1$. Then t can be as small as ______.

   (A) $O(\sqrt{\log n})$ (B) $O(\log n)$ (C) $O(\sqrt{n})$ (D) $O(n)$

   SOLUTION:
   The answer is (C), $O(\sqrt{n})$. To see this, by Hoeffding's inequality we have that $\Pr[S \geq pn + t] \leq 2\exp(-\Omega(t^2/n))$, so if we choose $t \approx \sqrt{n}$, this will be small. (Notice that you get the same answer with Chebyshev's inequality).

   (b) **(5 pt.)** Let $Y_1, \ldots, Y_n$ be i.i.d. random variables, uniformly distributed in $[0,1]$. Let $Y = \sqrt{\sum_{i=1}^n Y_i^2}$. Then

   $$\Pr[|Y - \mathbb{E}Y| \geq t] \leq \_\_\_\_\_$$

   (A) $O(1/\sqrt{t})$ (B) $O(1/t)$ (C) $\exp(-\Omega(t^2))$ (D) $\exp(-\Omega(t^2/n))$

   SOLUTION:
   The answer is (D). To see this, we can apply the Azuma-Hoeffding inequality to the Doob martingale $Z_j = \mathbb{E}[Y|Y_1, \ldots, Y_j]$. Notice that we have $|Z_j - Z_{j-1}| \leq 1$, as changing $Y_i \in [0,1]$ can only change Y by ≤ 1. (This follows from the triangle inequality, since $\|\vec{Y}\|_2 = \sqrt{\sum_i Y_i^2}$ is a norm), so we can apply the "method of bounded differences" approach from class, with $c_i = 1$.

   (c) **(5 pt.)** Let $Y_1, Y_2, \ldots, Y_k \in \{0,1\}$ be pairwise independent random variables with $\mathbb{E}[Y_i] = 1/2$. (Recall that this means that for any $i \neq j$, Y_i and Y_j are independent). Then

   $$\Pr\left[\sum_{i=1}^k Y_i = 0\right] \leq \_\_\_\_\_$$

   (A) $\exp(-\Omega(k))$ (B) $O(1/k)$ (C) $O(1/\sqrt{k})$ (D) $O(1)$

   SOLUTION:
   The answer is (B). This follows from the second moment method: Let $S = \sum_{i=1}^k Y_i$. Then

   $$\Pr[S = 0] \leq \frac{\mathrm{Var}(S)}{(\mathbb{E}S)^2} = \frac{(k/4)}{(k/2)^2} = 1/k.$$

Above, we have used that the Y_i's are pairwise independent to compute the variance.

Note that if the Y_i's were i.i.d., then (A) would be the correct answer (since $\sum_i Y_i = 0$ only if all of the Y_i are 0, which happens with probability $1/2^k$). But there are examples of pair-wise random variables where $O(1/k)$ is tight. For example, let $\ell = \log_2 k$ and let $X_1, \ldots, X_\ell$ be iid Bernoulli-$1/2$ random variables. Then let $Y_j = \sum_{i \in \mathrm{supp}(j)} X_i \mod 2$, where $\mathrm{supp}(j) \subseteq \{1, 2, \ldots, \ell\}$ denotes the support of j as a binary vector. Then you can check that the Y_j are pairwise independent, but the probability that they are all zero is at least the probability that all the X_ℓ's are zero, which is $1/2^\ell = 1/k$.

(d) **(5 pt.)** Let G be a graph with n vertices and m edges, where $m = 20n$. Then there is an independent set in G of size at least ______.

(A) $\Omega(1)$ (B) $\Omega(\log n)$ (C) $\Omega(\sqrt{n})$ (D) $\Omega(n)$

SOLUTION:
The answer was (D). In fact, this is essentially the same as a quiz question that we had for Class 9, and follows from a theorem from that lecture. (If you forgot the statement of the theorem from class, you could also re-prove it: remember that the key was to remove each vertex independently with probability p; and then for each remaining edge to remove one of its endpoints arbitrarily. Then optimize p.)

2. **(20 pt.)** Let $G = (V, E)$ be a simple graph (that is, an unweighted, undirected graph with no self-loops and no parallel edges). Let $D \geq 1$ be an integer. Suppose that each vertex $v \in V$ is associated with a set $S(v)$ of colors of size exactly $10D$. Suppose also that for each $v \in V$ and $c \in S(v)$, there are at most D neighbors u of v so that $c \in S(u)$. (For example, if "blue" is in $S(v)$, then v has at most D neighbors that also like the color "blue").

(a) **(10 pt.)** Prove that there is a way to properly color G so that for all $v \in V$, v's color is in $S(v)$. (Recall that a *proper* coloring is a coloring of the vertices so that no two neighboring vertices have the same color).

[**HINT:** *Try the LLL, and consider a bad event $A_{e,c}$ for each edge e and each color c.*]

SOLUTION:

Color each vertex v randomly with a color in $S(v)$. Let $A_{e,c}$ be the bad event that both endpoints of an edge e are colored the same color c. Then

$$\Pr[A_{e,c}] \leq \left(\frac{1}{10D}\right)^2 =: p.$$

Next, we compute the "d" parameter in the LLL. Let $e = \{u, v\}$ be an edge. Then for $c \in S(u) \cup S(v)$, consider the set $\mathcal{S}$ of events $A_{f,c'}$ where either:

- $f = \{u, w\}$ includes u, and $c' \in S(u) \cap S(w)$; or
- $f = \{v, w\}$ includes v, and $c' \in S(v) \cap S(w)$.

Then $A_{e,c}$ is mutually independent of all the events $A_{f,c'}$ outside of $\mathcal{A}$. (That's because if f and e don't intersect, these events don't have anything to do with each other; and if they do intersect, they only have something to do with each other if it's possible to color both endpoints of f the same color). So we can take

$$d = |\mathcal{S}| = 2 \times 10D \times D = 20D^2.$$

Above, the factor of 2 is for the two vertices $u, v \in e$. The factor of $10D$ is because for each of those vertices (say, u), there are at most $10D$ colors $c \in S(u)$. The factor of D is because, for each pair (u, c), there are at most D neighbors w of u so that $c \in S(u) \cap S(w)$.

So then $pd = \frac{20D^2}{100D^2} = \frac{1}{5} < \frac{1}{4}$, so the LLL applies. Thus, there is a way of coloring V so that none of the bad events occur.

another part on next page

(b) **(5 pt.)** Give a randomized algorithm to find such a coloring, given $G = (V, E)$. Your algorithm should have expected running time polynomial time in n, the number of vertices in V, and should succeed with probability 1.

You should clearly describe your algorithm so that someone who has not taken this class (but has taken, say, a basic programming/algorithms course) should be able to code it up without thinking too hard. You do not need to explain why your algorithm is correct.

SOLUTION:
We use the algorithmic LLL. In this case, that looks like this:

- For each $v \in V$, color v a uniformly random color $S(v)$.
- While there is some edge $e = \{u, v\}$ with both endpoints colored the same color:
  - Re-color u to a random color in $S(u)$, and re-color v to a random color in $S(v)$.
- Return the final coloring.

3. **(20 pt.)** Let $\{X_t\}$ be an irreducible, aperiodic, time-homogeneous Markov chain on states $\{1, \ldots, n\}$, so that for all $j \in \{1, \ldots, n\}$, $\Pr[X_{t+1} = 1 | X_t = j] \geq \varepsilon$ for some $\varepsilon > 0$.

   (a) **(5 pt.)** Let π be the stationary distribution of $\{X_t\}$. Fill in the blank with the best bound you can (do not use big-Oh notation). Justify why your answer is correct (you don't need to justify why it is the best possible).

$$\pi(1) \geq \_\_\_\_$$

   Justification:

SOLUTION:
By the definition of a stationary distribution,

$$\pi(1) = \sum_j \pi(j) \Pr[X_{t+1} = 1 | X_t = j] \geq \sum_j \pi(j)\varepsilon = \varepsilon.$$

   (b) **(15 pt.)** Show that there is a constant $C > 0$ (which depends on neither n nor ε) so that the mixing time τ_{mix} of $\{X_t\}$ satisfies $\tau_{mix} \leq \frac{C}{\varepsilon}$.
   [**HINT:** *Set up a coupling.*]
   [**HINT:** *Depending on how you do the problem, it might be helpful that* $\ln\left(\frac{1}{1-x}\right) = x + x^2/2 + x^3/3 + x^4/4 + \cdots$ *for any* $x \in (-1, 1)$.]

SOLUTION:
We will set up a coupling $\{X_t\}, \{Y_t\}$ as follows. Informally, since the probability of transitioning to state 1 is at least ε, we will coordinate $\{X_t\}$ and $\{Y_t\}$ so that they make that step towards state 1 together. Then, at each step they have at least an ε chance of colliding, so they will probably collide pretty quickly.

Formally, we set this up as follows.

- If $X_t = Y_t$, then they evolve together. (That is, sample X_{t+1} according to the transition matrix for $\{X_t\}$, and then set $Y_{t+1} = X_{t+1}$).
- If $X_t \neq Y_t$:
  - With probability ε, $X_{t+1} = Y_{t+1} = 1$.
  - With probability $1 - \varepsilon$, each chain independently takes a step according to the following probability distribution: The probability of transitioning from a state i to a state j is

$$\Pr[i \to j] = \begin{cases} \frac{P_{i,j} - \varepsilon}{1-\varepsilon} & j = 1 \\ \frac{P_{i,j}}{1-\varepsilon} & \text{else} \end{cases}$$

where above P is the transition matrix for $\{X_t\}$.

Note that we chose the rule above to make sure that each of $\{X_t\}$ and $\{Y_t\}$ follow the right distribution. Indeed, for each of them, the probability of transitioning from any state i to another state j is

$$\Pr[i \to j] = \begin{cases} 1 \cdot \varepsilon + \left(\frac{P_{ij}-\varepsilon}{1-\varepsilon}\right)(1-\varepsilon) = P_{ij} & j = 1 \\ 0 \cdot \varepsilon + \left(\frac{P_{ij}}{1-\varepsilon}\right)(1-\varepsilon) = P_{ij} & j \neq 1 \end{cases}$$

So this is a legitimate coupling.

Next, at each timestep, the probability that the two chains couple is at least ε, so

$$\Pr[\text{not coupled by time } t] \leq (1-\varepsilon)^t \leq \exp(-t\varepsilon).$$

Thus, if we choose $t = \log(2e)/\varepsilon$, we see that this is at most $1/2e$. So $\Delta\left(\frac{\log(2e)}{\varepsilon}\right) \leq \frac{1}{2e}$, which by a theorem from class implies that $\tau_{mix} \leq \frac{2e}{\varepsilon}$.

4. **(20 pt.)** Let $G = (V, E)$ be a directed, weighted graph with positive edge weights and no self-loops; for $(u, v) \in E$, let $w(u, v) > 0$ denote the edge weight of (u, v). Let $W = \sum_{e \in E} w(e)$ be the total weight of all the edges.

   Let $\pi : V \to \{1, \ldots, n\}$ be a bijection, and think of it as an ordering of the vertices (so, $\pi(v) = 1$ means that v is the first vertex, $\pi(w) = 2$ means that w is the second vertex, and so on). For such an ordering π, define

$$\mathrm{val}(\pi) = \sum_{(u,v) \in E} \mathbf{1}[\pi(u) < \pi(v)] w(u, v).$$

   That is, $\mathrm{val}(\pi)$ is the total weight of the edges that are going "forward" in the ordering.

   (a) **(5 pt.)** Prove that there exists an ordering π so that $\mathrm{val}(\pi) \geq W/2$.

SOLUTION:
We use the probabilistic method. Choose an ordering π uniformly at random. Then

$$\mathbb{E}[\mathrm{val}(\pi)] = \sum_{(u,v) \in E} \Pr[\pi(u) < \pi(v)] w(u, v) = \frac{1}{2} \sum_{(u,v) \in E} w(u, v) = W/2,$$

where above we have used that for a random ordering, $\Pr[\pi(u) < \pi(v)] = 1/2$ by symmetry. Thus, there exists some π with value at least $W/2$.

   (b) **(5 pt.)** Let π be a uniformly random ordering. Fix $t > 0$, and condition on the values of $P_t := (\pi^{-1}(1), \pi^{-1}(2), \ldots, \pi^{-1}(t))$. That is, we are conditioning on the choices of the first through t'th vertices. Let

$$U = V \setminus \{\pi^{-1}(1), \pi^{-1}(2), \ldots, \pi^{-1}(t)\}$$

   be the vertices that have not been chosen yet. Explain why there exists a $u \in U$ so that

$$\mathbb{E}[\mathrm{val}(\pi) | P_t, \pi(u) = t + 1] \geq \mathbb{E}[\mathrm{val}(\pi) | P_t]$$

SOLUTION:
We have

$$\begin{aligned}\mathbb{E}[\mathrm{val}(\pi) | P_t] &= \sum_{u \in U} \mathbb{E}[\mathrm{val}(\pi) | P_t, \pi(u) = t + 1] \cdot \Pr[\pi(u) = t + 1 | P_t] \\ &= \frac{1}{|U|} \sum_{u \in U} \mathbb{E}[\mathrm{val}(\pi) | P_t, \pi(u) = t + 1],\end{aligned}$$

as it's equally likely that we choose any of the remaining vertices $u \in U$ as the $t + 1$'st vertex in the ordering π. Thus, $\mathbb{E}[\mathrm{val}(\pi) | P_t]$ is the average over all of the values $[\mathrm{val}(\pi)|P_t, \pi(u) = t + 1]$ for $u \in U$, and so there must be at least one of those that's at least $\mathbb{E}[\mathrm{val}(\pi)|P_t]$.

more parts on next page

(c) **(5 pt.)** With the same notation as in part (b), explain how to efficiently and *deterministically* find a $u \in U$ so that

$$\mathbb{E}[\text{val}(\pi)|P_t, \pi(u) = t+1] \geq \mathbb{E}[\text{val}(\pi)|P_t].$$

Your answer should not use any probabilistic language, and should contain a clear description of a decision procedure that runs in time polynomial in n. You don't need to explain why your answer is correct.
[**HINT:** *Consider* $\mathbb{E}[\text{val}(\pi)|P_t, \pi(u) = t+1] - \mathbb{E}[\text{val}(\pi)|P_t]$.]

SOLUTION:
For $u \in U$, let $\text{Out}_U(u) = \sum_{v \in U:(u,v)\in E} w(u,v)$, and let $\text{In}_U(u) = \sum_{v \in U:(v,u)\in E} w(v,u)$. That is, $\text{Out}_U(u)$ is the total weight of edges going *out* of u, into U; and $\text{In}_U(u)$ is the total weight of edges going *in* to u, out of U. Our algorithm is the following:

- Compute $\text{Out}_U(u)$ and $\text{In}_U(u)$ for each $u \in U$.
- Return the u that maximizes $\text{Out}_U(u) - \text{In}_U(u)$.

To see why this works (not required), we follow the hint and consider $\mathbb{E}[\text{val}(\pi)|P_t, \pi(t+1) = u] - \mathbb{E}[\text{val}(\pi)|P_t]$. By linearity of expectation, this is equal to

$$\sum_{(w,v)\in E} w(w,v) \left(\Pr[\pi(w) < \pi(v)|P_t, \pi(t+1) = u] - \Pr[\pi(w) < \pi(v)|P_t]\right).$$

Notice that most of these terms cancel. The only ones that don't are when both $v, w \in U$, and when either $v = u$ or $w = u$. Thus, this simplifies to

$$\sum_{v \in U:(u,v)\in E} w(u,v)(1 - 1/2) + \sum_{w \in U:(w,u)\in E} w(u,v)(0 - 1/2),$$

where above we have used that $\Pr[\pi(w) < \pi(v)|P_t] = 1/2$ for any $v, w \in U$. We have also used that $\Pr[\pi(u) < \pi(v)|P_t, \pi(t+1) = u] = 1$, since conditioning on choosing u as the $t+1$'st vertex, we know that $v \in U$ will be at least the $t+2$'nd vertex, and similarly $\Pr[\pi(w) < \pi(u)|P_t, \pi(t+1) = u] = 0$. We can then re-write this expression as

$$\frac{1}{2}\left(\text{Out}_U(u) - \text{In}_U(u)\right),$$

so we just want to pick some u so that this is non-negative. We know that it exists by part (b), so choosing u to maximize this quantity will work.

(d) **(5 pt.)** Put together your answers from the previous parts to give a deterministic algorithm to find an ordering π so that $\text{val}(\pi) \geq W/2$. Your algorithm should run in time polynomial in n. Your algorithm should be clear enough that someone who has not taken this class can code it up without thinking too hard. You do not need to explain why it is correct.

SOLUTION:
This follows our framework for derandomization via conditional expectation. The algorithm is:

- Initialize $U \leftarrow V$.
- Initialize $\pi = \{\}$ (an empty lookup table)
- For $t = 1, 2, \ldots, n$:
  - Compute $\mathrm{Out}_U(u)$ and $\mathrm{In}_U(u)$ for each $u \in U$.
  - Choose $u^* \in U$ to maximize $\mathrm{Out}_U(u) - \mathrm{In}_U(u)$
  - Set $\pi[u^*] = t$
  - Set $U \leftarrow U \setminus \{u^*\}$
- Return π.

5. **(20 pt.)** Suppose that there are n people, each of whom has a hat. Everyone takes off their hat at a party and puts them in a pile. When the party is over, people try to find their hats by grabbing random hats from the pile. More precisely, what happens is this:

- A uniformly random permutation maps the n hats to the n people.
- If a person ends up with their own hat, they take it and go home.

Suppose that X_1 people find their hats this way and leave. Now there are $n - X_1$ people left and they do it again:

- A uniformly random permutation maps the remaining $n - X_1$ hats to the remaining $n - X_1$ people.
- If a person ends up with their own hat, they take it and go home.

This goes on for R rounds until everyone has left. Let X_i be the number of people who find their hats and leave in the i'th round. (That is, we end up with $X_1, X_2, \ldots, X_R$ so that $X_R > 0$ and $\sum_{i=1}^R X_i = n$).

(a) **(5 pt.)** Let $Y_j = \sum_{i=1}^j (X_i - \mathbb{E}[X_i | X_1, \ldots, X_{i-1}])$. Show that $\{Y_j\}$ is a martingale with respect to $\{X_j\}$.

SOLUTION:
First we check that Y_j is a function of $X_1, \ldots, X_j$ (it is), and that $\mathbb{E}|Y_j|$ is finite. To see the latter, note that

$$\mathbb{E}[|Y_j|] \leq \left| \sum_{i=1}^j X_i \right| + \left| \sum_{i=1}^j \mathbb{E}[X_i | X_1, \ldots, X_{i-1}] \right|$$

by the triangle inequality. This is at most $2n$, which is finite.

Next, we verify the martingale property. We have

$$\begin{aligned}\mathbb{E}[Y_j | X_1, \ldots, X_{j-1}] = &\sum_{i=1}^{j-1} \mathbb{E}\left[X_i - \mathbb{E}[X_i | X_1, \ldots, X_{i-1}] | X_1, \ldots, X_{j-1}\right] \\ &+ \mathbb{E}\left[X_j - \mathbb{E}[X_j | X_1, \ldots, X_{j-1}] | X_1, \ldots, X_{j-1}\right].\end{aligned}$$

The first term is

$$\sum_{i=1}^{j-1} (X_i - \mathbb{E}[X_i | X_1, \ldots, X_{i-1}]) = Y_{j-1},$$

since $\mathbb{E}[X_i | X_1, \ldots, X_{j-1}] = X_i$ for $i \leq j - 1$.

The second term is

$$\mathbb{E}[X_j|X_1,\ldots,X_{j-1}] - \mathbb{E}[X_j|X_1,\ldots,X_{j-1}] = 0.$$

So together, we get

$$\mathbb{E}[Y_j|X_1,\ldots,X_{j-1}] = Y_{j-1} + 0 = Y_{j-1},$$

as desired.

(b) **(5 pt.)** Show that for any i, $\mathbb{E}[X_i|X_1,\ldots,X_{i-1}] = 1$.

SOLUTION:
$\mathbb{E}[X_i|X_1,\ldots,X_{i-1}]$ is the expected number of people who find their hat at step i, conditioned on there being $m = n - \sum_{j=1}^{i-1} X_j$ people left. (Note that all that matters is how many people were left, not about how that was divided up between $X_1,\ldots,X_{i-1}$).
By linearity of expectation, this is the same as

$$\sum_{\ell=1}^{m} \Pr[\text{ person } \ell \text{ finds their hat } | \text{there are } m \text{ people left}] = \sum_{\ell=1}^{m} \frac{1}{m} = 1.$$

Above, we have used that $1/m$ is the probability that person ℓ finds their hat out of m possibilities.

more parts on next page

Continued from hat problem on previous page

(c) **(5 pt.)** Use the Martingale stopping theorem to compute $\mathbb{E}[R]$, the expected number of rounds until everyone has found their hat. You do not (yet) need to prove that the Martingale stopping theorem applies.

SOLUTION:
We use the Martingale stopping theorem to the stopping time R and the martingale $\{Y_i\}$. Since we can assume that it applies, we have $\mathbb{E}[Y_0] = \mathbb{E}[Y_R]$. Note that $\mathbb{E}[Y_0] = 0$ (as it is the empty sum). Thus,

$$\begin{aligned} 0 &= \mathbb{E}[Y_R] \\ &= \mathbb{E}\left[\sum_{i=1}^{R}(X_i - \mathbb{E}[X_i|X_1,\ldots,X_{i-1}])\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{R}(X_i - 1)\right] \qquad \text{by part (b)} \\ &= \mathbb{E}[n-R] \qquad \text{since } \sum_{i=1}^{R} X_i = n \\ &= n - \mathbb{E}[R]. \end{aligned}$$

We conclude that $\mathbb{E}[R] = n$.

(d) **(5 pt.)** Prove that the Martingale Stopping Theorem applies to R.

SOLUTION:
We use item 3 from the statement of the theorem, so we need to show that $\mathbb{E}[Y_{i+1} - Y_i|X_0,\ldots,X_i] < c$ for some constant c, and that $\mathbb{E}R < \infty$. For the first thing, notice that

$$\begin{aligned} \mathbb{E}[|Y_{i+1} - Y_i||X_0,\ldots,X_i] &= \mathbb{E}[|X_{i+1} - \mathbb{E}[X_{i+1}|X_0,\ldots,X_i]||X_0,\ldots,X_i] \\ &= \mathbb{E}[|X_{i+1} - 1||X_0,\ldots,X_i] \qquad \text{by part (b)} \\ &\le \mathbb{E}[X_{i+1}|X_0,\ldots,X_i] + 1 \qquad \text{by the triangle inequality} \\ &= 1 + 1 \qquad \text{by part (b) again} \\ &= 2. \end{aligned}$$

Second, we also need to show that $\mathbb{E}[R] < \infty$. To see this, note that the probability that a particular person doesn't find their hat by time t is at most $(1-1/n)^t$, since they always have at least a $1/n$ chance of finding their hat. Thus, $\Pr[R \ge t] \le$

$n(1-1/n)^t$ by the union bound. Then

$$\mathbb{E}R \le \sum_t \Pr[R \ge t] \le n \sum_t (1-1/n)^t = n \cdot \left(\frac{1}{\frac{1}{1-1/n} - 1} \right) = n(n-1) < \infty.$$

This is the end!

This is the end of the exam! You can use this page for extra work on any problem. **Keep this page attached** to the exam packet (whether or not you use it), and if you want extra work on this page to be graded, clearly label which question your extra work is for.

This page is for extra work on any problem. **Keep this page attached** to the exam packet (whether or not you use it), and if you want extra work on this page to be graded, clearly label which question your extra work is for.

This is the end of the exam!

This is the end of the exam! You can use this page for extra work on any problem. **Keep this page attached** to the exam packet (whether or not you use it), and if you want extra work on this page to be graded, clearly label which question your extra work is for, and make a note on the problem page itself.

This page is for extra work on any problem. **Keep this page attached** to the exam packet (whether or not you use it), and if you want extra work on this page to be graded, clearly label which question your extra work is for, and make a note on the problem page itself.

This page is for extra work on any problem. **Keep this page attached** to the exam packet (whether or not you use it), and if you want extra work on this page to be graded, clearly label which question your extra work is for, and make a note on the problem page itself.

This page is for extra work on any problem. **Keep this page attached** to the exam packet (whether or not you use it), and if you want extra work on this page to be graded, clearly label which question your extra work is for, and make a note on the problem page itself.

Some useful inequalities, definitions and theorem statements

Note: We have not always stated full theorems here, just the quantitative punchlines. You are responsible for knowing when each theorem applies.

Inequalities and Series

- $1 - x \le e^{-x}$ for any x.
- $(n/k)^k \le \binom{n}{k} \le (en/k)^k$ for all $k \le n$.
- $\binom{n}{k} \le \frac{n^k}{k!}$ for all $k \le n$.
- $\sum_{i=1}^n 1/i = \Theta(\log n)$
- $\sum_{i=1}^n 1/i^c = O(1)$ for all $c > 1$.

Definitions

- $f(n) = O(g(n))$ means $\exists$ constants $c, n_0 > 0$ so that for all $n \ge n_0$, $f(n) \le cg(n)$.
- $f(n) = \Omega(g(n))$ means $\exists$ constants $c, n_0 > 0$ so that for all $n \ge n_0$, $f(n) \ge cg(n)$.
- $f(n) = o(g(n))$ means that $\frac{f(n)}{g(n)} \to 0$ as $n \to \infty$.
- $f(n) = \omega(g(n))$ means that $\frac{f(n)}{g(n)} \to \infty$ as $n \to \infty$.
- If $X \sim \text{Poi}(\lambda)$, then $\Pr[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}$.
- If $X \sim N(\mu, \sigma^2)$, then $\Pr[X = x] = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$
- If $X \sim \text{Ber}(p)$, then $X \in \{0, 1\}$ and $\Pr[X = 1] = p$.

Concentration Inequalities

- Markov's inequality: For a non-negative random variable X, $\Pr[X > t] \le \frac{\mathbb{E}X}{t}$.
- Chebyshev's inequality: For any random variable X, $\Pr[|X - \mathbb{E}X| > t] \le \frac{\text{Var}(X)}{t^2}$.
- A few Chernoff bounds: For independent $X_i \in \{0, 1\}$, if $X = \sum_{i=1}^n X_i$, then:
  - For $\delta > 0$, $\Pr[X \ge (1+\delta)\mathbb{E}[X]] \le \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}[X]}$. If $\delta \in (0, 1]$ this is $\le \exp(-\delta^2\mathbb{E}[X]/3)$.
  - For $\delta \in (0, 1]$, $\Pr[X \le (1-\delta)\mathbb{E}[X]] \le \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mathbb{E}[X]}$. If $\delta \in (0, 1]$, this is $\le \exp(-\delta^2\mathbb{E}[X]/2)$.

- For $c \geq 6$, $\Pr[X \geq c\mathbb{E}X] \leq 2^{-c\mathbb{E}X}$.
- Hoeffding's inequality: if $X_i \in [a_i, b_i]$, $\Pr\left[|X - \mathbb{E}X| \geq t\right] \leq \exp\left(\frac{-2t^2}{\sum_i (a_i - b_i)^2}\right)$.

- Tail bound for Poisson random variables: If $X \sim \text{Poi}(\lambda)$, then for any $c > 0$, $\Pr[|X - \lambda| \geq c] \leq 2\exp\left(\frac{-c^2}{2(c+\lambda)}\right)$.

- Azuma-Hoeffding Inequality: Let $\{Z_t\}$ be a martingale with respect to $\{X_t\}$, and suppose $|Z_i - Z_{i-1}| \leq c_i$ for all $i \leq n$. For any $\lambda > 0$, $\Pr[|Z_n - Z_0| \geq \lambda] \leq 2\exp\left(\frac{-\lambda^2}{2\sum_{i=1}^n c_i^2}\right)$.

Probabilistic Method

- Second moment method: for real-valued X, $\Pr[X = 0] \leq \frac{\text{Var}[X]}{(\mathbb{E}[X])^2}$.

- LLL: Let $A_1, \ldots, A_n$ be events so that $\Pr[A_i] \leq p$ for all i, and where each A_i is mutually independent of all but d other events. Then:
  - If $pd \leq 1/4$, then $\Pr[\bigcap_i \overline{A_i}] > 0$
  - If $p(d+1) \leq 1/e$, then $\Pr[\bigcap_i \overline{A_i}] > 0$.

Markov Chain / Martingale Theorems

- Fundamental theorem of Markov chains: Let $\{X_t\}$ be an irreducible aperiodic Markov chain over a finite state space with transition matrix P. Then there is a unique stationary distribution π so that $\Pr[X_t = i | X_0 = j] \to \pi_i$ for all states i, j. Further, $1/\pi_i$ is the expected return time of state i, and $\pi P = \pi$.

- Let $\{X_t\}$ be a finite irreducible aperiodic Markov chain with a coupling $\{(X_t, Y_t)\}$. Then $\Delta(t) \leq \max_{s,s'} \Pr[X_t \neq Y_t | X_0 = s, Y_0 = s']$.

- Let $\{X_t\}$ be a finite irreducible aperiodic Markov chain and let T be a strong stationary stopping time. Then $\Delta(t) \leq \Pr[T > t]$.

- The *Doob Martingale* for a quantity A is $Z_t = \mathbb{E}[A | X_0, \ldots, X_t]$. Theorem: it is a martingale.

- Martingale stopping theorem: Let $\{Z_t\}$ be a martingale with respct to $\{X_t\}$. Let T be a stopping time for $\{X_t\}$. Then $\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$ if at least one of the following holds:

  1. There is a constant c s.t. $|Z_i| \leq c$ for all i.

  2. There is a constant c s.t. $T < c$ with probability 1.

  3. $\mathbb{E}[T] < \infty$ and there is a constant c s.t. for all i, $\mathbb{E}[|Z_{i+1} - Z_i| | X_0, \ldots, X_i] < c$.