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 **150 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:" two 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. **(12 pt.)** For each of the parts below, select the *best bound* you can to correctly fill in the blank.

   (a) **(4 pt.)** Let G be a finite group and let H be a subgroup of G, so that $H \neq G$. Choose $g \in G$ uniformly at random. Then $\Pr[g \in H] \leq$ ______.

   (A) 1 (B) $1 - 1/|G|$ (C) $1/2$ (D) $1/|S|$

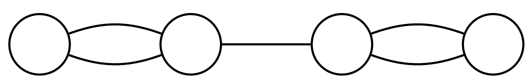
   SOLUTION:
   The answer is (C). This is the reason why our primality testing algorithm worked! (Aka, Lagrange's theorem).

   (b) **(4 pt.)** Let $f(x, y, z)$ be a 3-variate polynomial of total degree at most d, and suppose that f is not identically zero. Choose a, b, c independently and uniformly at random from the set $\{1, 2, \ldots, 10\}$. Then $\Pr[f(a, b, c) = 0] \leq$ _____.

   (A) 1 (B) $d/10$ (C) $d/10^3$ (D) $d^3/10^3$

   SOLUTION:
   The answer is (B). This follows from the Schwartz-Zippel lemma.

   (c) **(4 pt.)** Let G be the following graph:

   Then Pr[One run of Karger's algorithm succeeds starting from this graph] $\geq$ _____

   (A) 0 (B) $8/15$ (C) $2/3$ (D) $4/5$

   SOLUTION:
   The answer is (B). The probability that it succeeds in the first step is $4/5$. Conditioned on succeeding, we are left with a graph with three (mega)-vertices, A,B,C, with a double edge between A and B and a single edge between C, and so the probability of succeeding in the next step is $2/3$, at which point we are down to only two nodes and we are done. So the probability of success is $4/5 \cdot 2/3 = 8/15$.

2. **(26 pt.)** For all of the parts of this problem, suppose that $Y_1, Y_2, \ldots, Y_n$ are random variables (not necessarily independent) **that take on values in** $\{0, 1\}$**.**

   For each of the following scenarios, select the smallest correct way, *among the options provided*, to fill in the blank in the following expression

$$\Pr\left[\sum_{i=1}^{n} Y_i > n/2\right] < \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_},$$

and also identify which inequality/theorem you would use to show this. It's okay to identify the inequality/theorem by name or uniquely identifying description; you don't need to write out a formula. In this problem, all the big-Oh notation below is as $n \to \infty$, so in particular you can assume n is sufficiently large.

For parts (a)-(c), no justification is required (or will be considered), beyond the name of the inequality.

(a) **(6 pt.)** For each i, $\mathbb{E}[Y_i] = 1/5$. (And you know nothing else about them other than that they are binary).

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

I would use the inequality: ____

SOLUTION:
The answer is A. I would use Markov's inequality.

(b) **(6 pt.)** For each i, $\mathbb{E}[Y_i] = 1/5$, and the Y_i are *pairwise independent.* That is, for all $i \neq j$, Y_i and Y_j are independent.

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

I would use the inequality: ____

SOLUTION:
The answer is C. I would use Chebyshev's inequality.

(c) **(6 pt.)** For each i, $\mathbb{E}[Y_i] = 1/5$, and the Y_i are fully independent.

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

I would use the inequality: ____

SOLUTION:
The answer is E. I would use a Chernoff bound.

(d) **(8 pt.)** The Y_i are defined as follows. We drop $30 \cdot n$ balls into n bins, and Y_i is 1 if bin i has at least 100 balls in it and is 0 otherwise.

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

In a few sentences, explain what you would do to establish your answer. That is, say what inequalities/theorems would you use, and clearly define the random variables you would use them on. You don't need to work out the details. You don't need to justify why your answer is tight (e.g., if your answer is (A), you should explain how to show that (A) is true, but you don't need to explain how to show that (B) is not true).

SOLUTION:

The answer is E. To see this, we approximate the occupancies by Poisson random variables. In more detail, let $k \sim Poi(40n)$, and consider dropping k balls into n bins. Let Z_i be the occupancy of the i'th bin, so $Z_i \sim Poi(40)$. Then let W_i be 1 if $Z_i \geq 100$ and 0 otherwise. Thus, $\Pr[W_i = 1] =: p$ is some constant, and by our tail bound for Poisson random variables,

$$p \leq 2\exp(-60^2/(2 \cdot 100)),$$

which is very small and certainly less than 1/2.

Now the W_i are fully independent and we can use a Chernoff bound to show that $\Pr[\sum_i W_i > n/2] \leq \exp(-\Omega(n))$. To de-Poissonify, we observe that the probability that $k < 40n$ is $\exp(-\Omega(n))$ by our tail bound for Poisson random variables. If that happens, then

$$\Pr[\sum_i Y_i > n/2] \leq \Pr[\sum_i W_i > n/2 | k \geq 40n],$$

because the Y_i correspond to a situation where fewer balls have been dropped, so it is less likely that there are many bins with at least 100 balls. Finally, we can use the fact that

$$\Pr[\sum_i W_i > n/2 | k \geq 40n] \leq \frac{\Pr[\sum_i W_i > n/2]}{\Pr[k \geq 40n]} \leq \exp(-\Omega(n)).$$

Grading note: We would not require so much detail, especially about de-Poissonification, since we only asked for a sentence or two. But we've included it here to make the answer clear.

3. **(20 pt.)** For each of the tasks below, explain briefly how you would do them. You can *(and, **HINT**, probably should)* use any algorithm we have seen as a black box.

(a) **(10 pt.)** Let $\mathcal{X} = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$. Fix $T > 0$. You want to split $\mathcal{X}$ into two disjoint parts, $\mathcal{X} = \mathcal{Y}_1 \cup \mathcal{Y}_2$, so that there are few pairs x_i, x_j in different parts that are close to each other. Formally, the goal is to minimize $\sum_{x_i \in \mathcal{Y}_1, x_j \in \mathcal{Y}_2} \mathbf{1}[\|x_i - x_j\|_2 \leq T]$. Give a randomized algorithm that runs in time $\mathrm{poly}(n)$ and finds a minimizing $\mathcal{Y}_1, \mathcal{Y}_2$ with probability at least 0.9.

SOLUTION:
Let $G = (V, E)$ be a graph with $V = \mathcal{X}$, and so that $\{x, y\} \in E$ if and only f $\|x - y\|_2 \leq T$. The run Karger's algorithm on G (repeating enough times to ensure a success probability of at least 0.9), and return the partition given by the minimum cut.

(b) **(10 pt.)** Let $\mathcal{X} = \{x_1, x_2, \ldots, x_n\} \subseteq \mathbb{R}^d$. Fix $T > 0$. Suppose you have black-box access to a deterministic algorithm $\mathcal{A}$ that runs on any such set $\mathcal{X}$ and partitions it into k disjoint parts, $\mathcal{X} = \mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \cdots \cup \mathcal{Y}_k$, so that the diameter of each $\mathcal{Y}_i$ is at most T.[1] (The *diameter* of a set $\mathcal{Y}$ is $\mathrm{diam}(\mathcal{Y}) = \max_{y \neq y' \in \mathcal{Y}} \|y - y'\|_2$). The algorithm $\mathcal{A}$ runs in time polynomial in n, k, and 2^d.

Give a randomized algorithm that runs in time polynomial in n, k, and d (not 2^d) that does the following. Given such a set $\mathcal{X}$ and a parameter T, your algorithm should return a partition $\mathcal{Y}_1, \ldots, \mathcal{Y}_k$ of $\mathcal{X}$ so that each part $\mathcal{Y}_i$ has diameter at most $1.1 \cdot T$. Your algorithm should be correct with probability 0.9, and should use $\mathcal{A}$ as a black box.

SOLUTION:
Let $\Phi \in \mathbb{R}^{m \times d}$ be a JL transform with distortion $\varepsilon < 1/100$ and target dimension $m = O(\log n)$. Then let $y_i = \Phi x_i$ for all i. Run $\mathcal{A}$ on the points $\{y_i : i \in [n]\}$ to obtain a partition $\tilde{\mathcal{Y}}_1, \ldots, \tilde{\mathcal{Y}}_k$. Then return $\mathcal{Y}_1, \ldots, \mathcal{Y}_k$, where

$$\mathcal{Y}_j = \left\{ x_i : y_i \in \tilde{\mathcal{Y}}_j \right\}.$$

To see why this works (which is not required for credit), note that

$$\mathrm{diam}(\mathcal{Y}_j) = \max_{x \neq x' \in \mathcal{Y}_j} \|x - x'\|_2 \leq \max_{x \neq x' \in \mathcal{Y}_j} \|\Phi x - \Phi x'\|_2 \left(\frac{1}{1 - \varepsilon} \right) \leq \frac{T}{1 - \varepsilon} \leq 1.1 \cdot T.$$

[1]For the purposes of this problem, suppose that such a partition always exists.

4. **(21 pt.)** There are n users, each of whom have an unlimited number of jobs to complete. Each job takes one timestep to complete. At each timestep, you independently choose a uniformly random user and complete a job for them. Note that you choose the users with replacement, so it's possible that over time you could complete multiple jobs for the same user.

   (a) **(5 pt.)** Let T be the first time you have completed at least one job for each of the n users. What is $\mathbb{E}[T]$? Circle the best (smallest) answer that is true, and briefly justify it. You do not need to do a formal proof, and you can appeal to anything we have seen in class.

   (A) $O(n)$ (B) $O(n \log n)$ (C) $O(n^2)$ (D) $2^{O(n)}$

   Justification:

   SOLUTION:
   The answer is (B). This is coupon collecting.

   (b) **(8 pt.)** Let S be the first time you have completed a job for 99% of the n users. What is $\mathbb{E}[S]$? Circle the best (smallest) answer that is true, and briefly justify it. You do not have to give a formal proof, and you can appeal to anything we have seen in class.

   (A) $O(n)$ (B) $O(n \log n)$ (C) $O(n^2)$ (D) $2^{O(n)}$

   Justification:

   SOLUTION:
   The answer is (A). To see this, we can use the same logic we used when deriving the answer to the coupon collector's problem:

$$\begin{aligned}\mathbb{E}S &= \sum_{i=1}^{0.99n} \mathbb{E}[\text{time to see the } i\text{'th new user after you see the } i+1\text{'st}] \\ &= \sum_{i=1}^{0.99n} \frac{n}{n-i+1}.\end{aligned}$$

   Now, notice that $n - i + 1 \geq 0.01n + 1$, so $\frac{n}{n-i+1} \leq 100$. Thus, the above is at most $\sum_{i=1}^{0.99n} 100 \leq 100n = O(n)$.

[Another part on next page!]

[Continued from previous page]

(c) **(8 pt.)** Suppose you run for $t \sim \mathrm{Poi}(n^2)$ timesteps. Prove that with probability at least 0.99, no user has had more than $1.1n$ jobs run. You may assume that n is sufficiently large.

SOLUTION:
Let X_i be the number of jobs for user i. Then $X_i \sim \mathrm{Poi}(n)$ for all i. Thus,

$$\begin{aligned}\Pr[X_i > 1.1n] &\leq \Pr[|X_i - n| \geq 0.1n] \\ &\leq 2\exp\left(\frac{-(0.1n)^2}{1.1n}\right) \\ &\leq 2\exp(-\Omega(n)),\end{aligned}$$

where above we used our tail bound for Poisson random variables. Union bounding over all $i = 1, \ldots, n$, we have

$$\Pr[\exists i, X_i > 1.1n] \leq 2n\exp(-\Omega(n)) \leq 0.01$$

for large enough n.

5. **(21 pt.)** Suppose you are interested in computing a function $f : D_1 \to D_2$, for some domains D_1 and D_2.

(a) **(8 pt.)** Let $\varepsilon \in (0, 1/2)$, and suppose $D_2 = \{0, 1\}$. Suppose you have a randomized algorithm A so that, for any $x \in D_1$, $A(x) = f(x)$ with probability at least $\frac{1}{2} + \varepsilon$. You decide to make a more robust algorithm, $\tilde{A}$, that just runs A independently T times and returns the most frequent answer. Show that $\tilde{A}$ is correct with probability 0.99 for some value T that is $O(1/\varepsilon^2)$.

SOLUTION:
Fix $x \in D_1$. $\tilde{A}(x)$ will fail if the number of times that $A(x)$ fails to output $f(x)$ is greater than $T/2$. Let X_i be the indicator variable that is 1 if $A(x) \neq f(x)$ on the i'th trial. Then, $\mathbb{E}X_i = 1 - p_1 \leq 1/2 - \varepsilon$. By Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}\left\{\sum_i X_i \geq T/2\right\} &= \mathbb{P}\left\{\left|\sum_i (X_i - \mathbb{E}X_i)\right| \geq \varepsilon T\right\} \\
&\leq \frac{\mathbb{E}\left(\sum_i (X_i - \mathbb{E}X_i)\right)^2}{\varepsilon^2 T^2} \\
&\leq \frac{T p_1 (1 - p_1)}{\varepsilon^2 T^2} \\
&= \frac{p_1(1 - p_1)}{\varepsilon^2 T}
\end{aligned}$$

Thus, if we choose $T = \Omega(1/\varepsilon^2)$ sufficiently large, this will be smaller than 0.01.

(b) **(8 pt.)** Now suppose that $D_2 = \{0, 1, \ldots, n\}$, instead of $\{0, 1\}$. Suppose that for any $x \in D_1$, $f(x) = A(x)$ with probability at least $p \geq \frac{C \log n}{n \log \log n}$, where C is some constant that you get to choose. Further suppose that for any *incorrect* answer $i \in \{0, 1, \ldots, n\}$, the probability that A outputs i is at most $1/n$. Prove that there exists some C so that, for sufficiently large n, $\tilde{A}$ (which still returns the most frequent answer out of the T trials) is correct with probability at least 0.99 when $T = n$. You may use anything we have seen in class as a black box.

SOLUTION:

Imagine that we have a bin for each element of D_2, and a ball for each trial. We will win as long as (A) the number of times that $A(x) = f(x)$ is at least $C \log n / \log \log n$ (half its expectation), and (B) the maximum load in the bins corresponding to $D_2 \setminus \{f(x)\}$ is less than $C \log n / \log \log n$. We will show that with high probability, both (A) and (B) occur.

We know that (A) will happen with high probability by a Chernoff bound (the second one on the reference sheet, with $\delta = 1/2$). In more detail, letting X_i be the event that $A(x) = f(x)$ and letting $\mu = pT$, we have

$$\begin{aligned}\mathbb{P}\{\text{ not } (A)\} &= \mathbb{P}\left\{\sum_i X_i < \frac{\mu}{2}\right\} \\ &\leq \exp(-\mu/8) \\ &= \exp(-pT/8) \\ &= \exp(-C \log n / \log\log n) = o(1).\end{aligned}$$

In particular, this is less than 0.05 for large enough n.

To analyze (B), consider the situation where each of $\Pr[A(x) = y]$ is exactly equal to $1/n$ for all $y \in D_2 \setminus \{f(x)\}$, and where none of the balls land in the bin corresponding to $f(x)$. The maximum load in this situation will only be higher than in our situation, so without loss of generality we may examine this balanced situation, where we are dropping n balls uniformly into n bins.

We saw in class that, with probability $1 - o(1)$, the maximum load in this situation is $\Theta(\log n / \log \log n)$, for sufficiently large n. In particular, there is some constant C' so that the max load is at most $C' \log n / \log \log n$ with probability $1 - o(1)$. Choose $C = 2C'$, and let n be large enough that the $o(1)$ failure probability is at most 0.05. Then (B) happens with probability at least $1 - 0.05$.

Finally, by the union bound that both (A) and (B) occur (and thus (1) holds) with probability at least 0.99.

(c) **(5 pt.) Note: this one may be more difficult, and is only worth 5 points.** As above, suppose $D_2 = \{0, 1, \dots, n\}$. Now suppose that A is correct with probability $1/4$, and can output any particular incorrect answer i with probability at most $1/8$. How small can you take T to still allow the guarantee that $\tilde{A}$ is correct with probability at least 0.99?

SOLUTION:
You can do it with $T = O(1)$. To see this, fix $x \in D_2$. For $y \in D_2$, let $p(y) = \mathbb{P}\{A(x) = y\}$. Partition $D_2 \setminus f(x)$ into at most 16 chunks, so that for each chunk $S \subseteq D_2 \setminus \{f(x)\}$,

$$\sum_{y \in S} p(y) \in [1/16, 3/16].$$

You can do this by the greedy algorithm: start taking elements of $D_2 \setminus \{f(x)\}$ until you have at least $1/16$ probability mass, then move on to the next chunk. Because all of the items in $D_2 \setminus \{f(x)\}$ have probability at most $1/8$, the maximum mass you'll accrue in any chunk is at most $1/16 + 1/8 = 3/16$. Now, we can have at most 16 of these chunks, because otherwise the total probability mass across the chunks would be more than 1. Next, we observe that the maximum number of votes for any element y in a chunk S is upper bounded by the number of votes for any element in S.

Fix a chunk $S \subseteq D_2 \setminus \{f(x)\}$. The probability that the number of votes for S is more than $(3/16 + 1/32)T$ is $\exp(-\Omega(T))$ by a Chernoff bound. Similarly, the probability that the number of votes for $f(x)$ is less than $3/16$ is also $\exp(-\Omega(T))$ by a Chernoff bound. Union bounding over the ≤ 17 bad events (one for each chunk S, and one for $f(x)$), we see that if we take $T = O(1)$ to be a large enough constant, the probability that any of these bad events occur is small.

If none of the bad events occur, then $\tilde{A}$ returns the correct answer.

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

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

Definitions

- $f(n) = O(g(n))$ means that there are some constants $c, n_0 > 0$ so that for all $n \geq n_0$, $f(n) \leq cg(n)$.
- $f(n) = \Omega(g(n))$ means that there are some constants $c, n_0 > 0$ so that for all $n \geq n_0$, $f(n) \geq cg(n)$.
- $f(n) = o(g(n))$ 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 $\Pr[X = 1] = p$.

More on other side

Concentration Inequalities

- Markov's inequality: For a non-negative random variable X, $\Pr[X > t] \leq \frac{\mathbb{E}X}{t}$.
- Chebyshev's inequality: For any random variable X, $\Pr[|X - \mathbb{E}X| > t] \leq \frac{\mathrm{Var}(X)}{t^2}$.
- A few Chernoff bounds: For independent $X_i \in \{0,1\}$, if $X = \sum_{i=1}^n X_i$, then:
  - For $\delta > 0$, $\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}[X]}$. If $\delta \in (0,1]$ this is $\leq \exp(-\delta^2 \mathbb{E}[X]/3)$.
  - For $\delta \in (0,1]$, $\Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mathbb{E}[X]}$. If $\delta \in (0,1]$, this is $\leq \exp(-\delta^2 \mathbb{E}[X]/2)$.
  - For $c \geq 6$, $\Pr[X \geq c\mathbb{E}X] \leq 2^{-c\mathbb{E}X}$.
- Tail bound for Poisson random variables: If $X \sim \mathrm{Poi}(\lambda)$, then for any $c > 0$, $\Pr[|X - \lambda| \geq c] \leq 2\exp\left(\frac{-c^2}{2(c+\lambda)}\right)$.

Balls and Bins

- The maximum load, when dropping n balls into n bins, is $\Theta\left(\frac{\log n}{\log \log n}\right)$ with probability $1 - o(1)$.
- Coupon collecting: $\lim_{n\to\infty} \Pr[X \geq n \log n + cn] = 1 - e^{-e^{-c}}$

Dimension Reduction

- Bourgain's Embedding: for any finite metric space (X, d) with $|X| = n$, there is an embedding of (X, d) into $\mathbb{R}^k$ under the ℓ_1 metric with distortion $O(\log n)$, where $k = O((\log n)^2)$.
- Johnson-Lindenstrauss Lemma: for any $\varepsilon \in (0,1)$, for any $X \subseteq \mathbb{R}^d$ with $|X| = n$, there is a linear map $f : \mathbb{R}^d \to \mathbb{R}^m$ with $m = O(\varepsilon^{-2} \log n)$ that embeds (X, ℓ_2) into $(\mathbb{R}^m, \ell_2)$ with distortion at most $(1+\varepsilon)$.

More on other side