CS 466/666: Algorithm Design and Analysis **University of Waterloo: Fall 2025**

Lecture 5

September 18, 2025

Instructor: Sepehr Assadi

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

Topics of this Lecture

We will conclude our initial study of probabilistic analysis of algorithms and the background on concentration inequalities in this lecture. The last example we study is a topic, aptly named *power of two choices*, in the balls and bins experiments of the last lecture.

1 Power of Two Choices in Balls and Bins

The problem we consider in this lecture is as follows.

Problem 1 (**Maximum Load in Balls-and-Bins**)**.** We are give a collection of m balls and n bins (so number of balls can be different from bins here). For each ball, we *independently* sample *two* distinct bins chosen *uniformly at random*; then place this ball to the bin which has the smaller load between the two (breaking ties arbitrarily). What is the asymptotically maximum load of a bin with probability at least $2/3$?

More precisely, our goal is to find the smallest possible upper bound $T(n)$ (as a function of n) such that with probability at least $2/3$, the load of *every* bin in this process is $O(T(n))$.

Before getting to the solution, let us go over a bit of practical motivation. The balls-and-bins experiment is a natural model of *hashing* and *task allocation*, e.g., choosing a server/processor (a bin) for an arriving task (a ball) in an online fashion. In some of these cases, we cannot readily know the loads of *all* servers in advance when a task arrives to choose a server with minimal load to assign this task to. For instance, when the tasks are coming from many parallel streams, no single stream can keep track of the loads on the servers (without a costly coordination with other streams) nor it is feasible for it to "query" the load of every server when a new task arrives (without incurring a huge communication cost) – the method in Problem 1 then gives us a "light weight" way of implementing an approximately load balanced solution with no coordination between the streams and minimal communication (only querying load of two bins at a time).

Mitzenmacher [Mit01] proved that the right answer to Problem 1 for $m = n$ is a load of

$$\log\log n + O(1),$$

which is exponentially better than the original approach of the last lecture (also notice how sharp this bound is)! This is a beautiful but intricate proof that we will not cover in this course; instead, we are going to prove a weaker version of this result which gets the asymptotically same bound (but with a worse leading constant) and requires a slightly smaller number of balls than bins. This other proof is from the paper of Karp, Luby, and Meyer auf der Heide [KLMadH92] for a different problem of hashing in parallel algorithms.

Theorem 1. *In Problem 1 for $m = n/100$ balls, the maximum load is $O(\log \log n)$* with high probability.

> **Remark.** The term 'with high probability' in Theorem 1 is used to refer to a probability which is at least $1 - 1/n^c$ for some constant $c \geqslant 1$; in most cases, one do not even explicitly specifies this constant and just leave it as being more than one. Generally, we accept an event happening 'with high probability' as happening really with a high probability in practice, because n, the input size, tends to be sufficiently large.
>
> For instance, if $n = 10^4$ (which is not even that large) and $c = 4$ (again not large), then the probability of this event happening is at most 10^{-16} – you can think of this number as follows: if every person on earth tries this event once every hour, *in expectation*, it still takes more than a century for the event to happen even once in the entire world!

> **Remark.** Morally speaking, the assumption of $m = n/100$ instead of $m = n$ is without loss of generality; we can just think of the experiment with $m = n$ as 100 repetitions of the process with $m = n/100$ restarted each time, and so the maximum load remains $O(\log \log n)$; this is not entirely accurate because after each restart, the initial loads are not the same, but that should only help us because we are sending the balls to less loaded bins. Nevertheless, we will not attempt to formalize this and stick to the $m = n/100$ case for simplicity.

We will prove this theorem in the rest of this lecture. Interestingly, this can be proven using basic ideas from **random graph theory**, a beautiful area of discrete probability theory.

Consider the following distribution $\mathbb{G}(m, n)$ on n-vertex graph:

> **Definition 2. Random graph distribution $\mathbb{G}(m, n)$:** Sample an n-vertex graph $G = (V, E)$ with a fixed set of vertices, say, $V = [n]$, as follows: sample m pairs of vertices uniformly at random and independently (with repetition) and add these pairs as edges to G.

The distribution $\mathbb{G}(m, n)$ is one standard form of a random graph distribution[1] and in random graph theory, we study properties of the graphs sampled from this and similar distributions. For instance, we can prove the following results.

Lemma 3. *For $m = n/100$, the largest* connected component *in a graph $G = (V, E)$ sampled from $\mathbb{G}(m, n)$ has $O(\log n)$ vertices with high probability.*

Lemma 4. *For $m = n/100$, any graph $G = (V, E)$ sampled from $\mathbb{G}(m, n)$ with high probability satisfies the following property: for any set $S \subseteq V$ of vertices, there are at most $5 \cdot |S|$ edges between the vertices in S.*

We will prove both these properties in the next section. For now, let us see what any of these has anything to do with Problem 1.

[1] Another standard form—called the Erdos-Renyi random graphs—is a distribution wherein every pair of vertices is sampled as an edge independently with some given probability p.

1.1 Proof of Theorem 1 (assuming Lemmas 3 and 4)

We can think of the balls-and-bins experiment of Problem 1 as sampling a graph $G = (V, E)$ from $\mathbb{G}(m, n)$: the vertices in V correspond to the bins and each edge of E corresponds to a ball and shows which two bins are sampled randomly for it[2]. Can we *upper bound* the maximum load of a bin in this trial of the balls-and-bins experiment if we are only given the graph G? *Yes*, we are going to do that exactly.

Throughout this proof, we fix $m = n/100$, which corresponds to the setting of the problem we are interested in. To continue we need to define the following process for a graph $G \sim \mathbb{G}(n/100, n)$:

Definition 5. Iterative peeling process: *Peel* all vertices in the graph G with degree at most 20 (i.e., remove them and their edges) and call them group B_1; then, again peel all vertices with degree at most 20 now (after we removed the vertices in B_1, the degree of some other vertices can drop), and call them B_2, and continue like this until either all vertices have been peeled or no vertex can be peeled.

Let $B_1, B_2, \ldots,$ be these groups and define B_∞ to be the last group that contains all vertices that could have not been peeled at all.

The following lemma is the key to the proof.

Lemma 6. *Consider a balls-and-bins trial, its corresponding graph G, and its iterative peeling process. For any $i \geqslant 1$, $i \neq \infty$, the bins in B_i have load at most $20 \cdot i$.*

Proof. We prove this inductively. Consider the base case of $i = 1$. These are bins with degree at most 20 in G; this means that at most 20 ball ever even considered picking them and so even if those balls all enter these bins, the load of these bins cannot become more than 20.

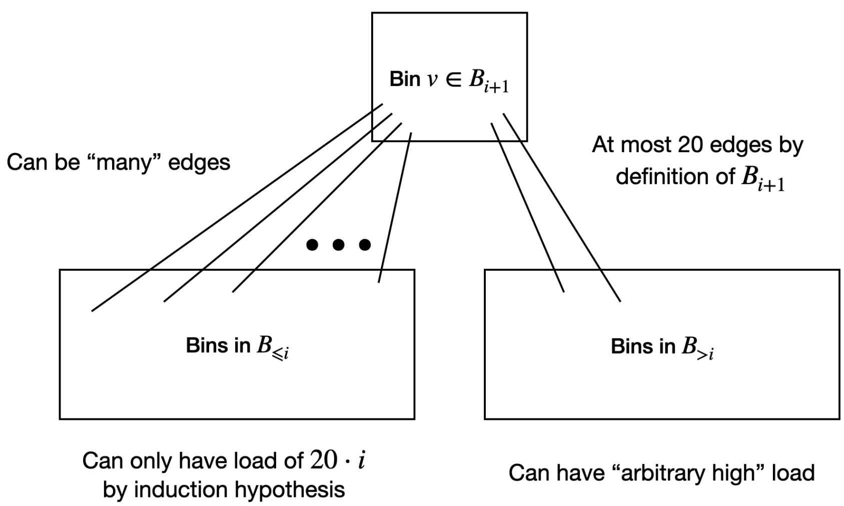

Figure 1: An illustration of the proof of the induction step. The edges/balls between v and $B_{>i}$ can (potentially) always choose v to assign their ball, but their number is few and do not increase load of v by much. The edges between v and $B_{\leqslant i}$ will only choose v to assign when its load is "small" as otherwise they can choose $B_{\leqslant i}$ instead by induction hypothesis.

[2]Note that this a *lossy* "mapping" as in, even though each trial of the balls-and-bins experiment defines a fixed graph G, given the graph only, we cannot recover *all* the information about the trial necessarily; in particular we are losing the information about the ordering of the arrival of balls now.

Now consider the induction step for $i+1$ (you can refer to Figure 1 for an illustration of this proof). All the bins in $B_{\leqslant i}$ have load at most $20 \cdot i$ by induction and a bin $v \in B_{i+1}$ only has $\leqslant 20$ edges to vertices *not* in $B_{\leqslant i}$ by definition. Consider any ball that is between $B_{\leqslant i}$ and v. If the load of v ever becomes more than $20 \cdot i$, then that ball chooses its other endpoint in $B_{\leqslant i}$ as its lower load bin by induction hypothesis. This means that the balls between $B_{\leqslant i}$ and v can only increase the load of v to $20 \cdot i$. The remaining balls may increase the load by another 20 at most, so load of v will be at most $20i + 20 = 20 \cdot (i+1)$, proving the induction hypothesis. $\square$

We are now ready to conclude the proof of Theorem 1. We argued that the corresponding graph G of a trial is sampled from $\mathbb{G}(n/100, n)$ and thus, with high probability, it satisfies both properties of Lemma 3 and Lemma 4 (we can do a union bound on the events of both lemmas). Conditioned on this, we argue that number of groups $B_1, B_2, \ldots,$ can only be $\log\log n + O(1)$ and $B_\infty = \emptyset$.

Consider any connected component C of G and we know that its size is $c = O(\log n)$ (as we conditioned on the event of Lemma 3). Moreover, we know that the average degree of vertices in C is at most 10 (as we conditioned on the event of Lemma 3 and because the sum of the degrees is twice the number of edges (the so-called handshaking lemma)). This means that at most $|C|/2$ vertices in C can have degree more than 20 otherwise average degree itself becomes more than 10 (this is also an application of Markov bound, do you see how?). So, in the first iteration of the peeling process, the size of the connected component C shrinks to at least half i.e., it is now at most $c/2$. But, again, even in this new component, the average degree is at most 10 (again by conditioning on the event of Lemma 3 that holds for *every* subset). So, in the next iteration, again size of C shrinks by another factor of half. This means that after at most $\log c$ iterations of the peeling process, this entire component is peeled.

In conclusion, as size of each connected component of G is $O(\log n)$, we get that after $\log O(\log n) = \log\log n + O(1)$ iterations of the peeling process, no vertex remains in the graph (and thus $B_\infty = \emptyset$). But then, by Lemma 6, we get that the maximum load of every bin is at most $20 \cdot \log\log n + O(1)$. As the events we conditioned on happen with high probability, we obtain this $O(\log\log n)$ bound with high probability also, concluding the proof.

2 Random Graph Theory: Proofs of Lemmas 3 and 4

We now prove the two results from random graph theory (Lemmas 3 and 4) that we used earlier.

There is a general strategy applicable to proving many random graph theory result: show that for the property to be false, a certain *witness* should exist in the graph; then, prove that each witness happens with such a lower probability in the graph that you can union bound over all its possible choices (we will shortly see two examples to help with clarifying this vague strategy). The heart of these arguments then usually is in finding the "right" witness. We will use this approach in proving both lemmas.

2.1 Lemma 3: Size of Connected Components in Random Graphs

Lemma (Restatement of Lemma 3)**.** *The largest* connected component *in a graph $G = (V, E)$ sampled from $\mathbb{G}(n/100, n)$ has $O(\log n)$ vertices with high probability.*

Proof. Consider a connected component C in a graph G. Then, since C has a spanning tree, we know that there are at least $|C|-1$ *different* edges inside C in G. Thus, for any $k \geqslant 1$, if we have a connected component of size k in G, we necessarily have a set S of k vertices with at least $k-1$ edges. We use such an S as our *witness* and prove that for some $k_0 = \Theta(\log n)$, with high probability, there is no witness of size $k \geqslant k_0$ in G.

Fix an integer $k \geqslant 1$. We have (here $\binom{S}{2}$ for a set (and not a number) S means the set of all pairs in S),

$$\Pr\left(\exists \text{ a set of size } k \text{ with} \geqslant k-1 \text{ different edges}\right) \leqslant \sum_{\substack{S \subseteq V \\ |S|=k}} \Pr\left(S \text{ has} \geqslant k-1 \text{ edges}\right) \qquad \text{(by union bound)}$$

$$\leqslant \sum_{\substack{S \subseteq V \\ |S|=k}} \sum_{\substack{F \subseteq \binom{S}{2} \\ |F|=k-1}} \Pr\left(\text{all edges of } F \text{ appear in } G\right)$$

(again, by union bound, because some fixed $k-1$ pairs inside $\binom{S}{2}$ should all be edges in G)

$$\leqslant \sum_{\substack{S \subseteq V \\ |S|=k}} \sum_{\substack{F \subseteq \binom{S}{2} \\ |F|=k-1}} \left(\frac{n/100}{\binom{n}{2}}\right)^{k-1}$$

(each edge appears in G with prob. $\leqslant m/\binom{n}{2}$ and edges are negatively correlated; see below (*))

$$= \binom{n}{k} \cdot \binom{\binom{k}{2}}{k-1} \cdot \left(\frac{n/100}{\binom{n}{2}}\right)^{k-1}$$

(by the bound on number of choices of S and F)

$$= \binom{n}{k} \cdot \binom{\frac{k\cdot(k-1)}{2}}{k-1} \cdot \left(\frac{n}{50\cdot n\cdot(n-1)}\right)^{k-1}$$

$$\leqslant \left(\frac{e\cdot n}{k}\right)^{k} \cdot \left(\frac{e\cdot k\cdot(k-1)}{2\cdot(k-1)}\right)^{k-1} \cdot \left(\frac{1}{50\cdot(n-1)}\right)^{k-1}$$

(as $\binom{a}{b} \leqslant (e\cdot a/b)^b$ proven in Lecture 4 (Fact 5))

$$\leqslant \left(\frac{e\cdot n}{k}\right)^{k} \cdot (2k)^{k-1} \cdot \left(\frac{1}{40\cdot n}\right)^{k} \cdot (40n)$$

(we used $(e/2) \leqslant 2$ and $n-1 \geqslant 4n/5$ as $n \to \infty$)

$$\leqslant \left(\frac{2e}{40}\right)^{k} \cdot 40n \leqslant \left(\frac{1}{2}\right)^{k} \cdot 40n.$$

In the equation marked (*), the edges are *negatively correlated* meaning that the probability of their joint event is upper bounded by what would have happened if they were independent; [3] because we have a fixed budget of m samples and if an edge is sampled, it can only decrease the probability that another edge is also chosen. Let $k_0 = 10\log n$. Then, for any $k \geqslant k_0$, we have,

$$\Pr\left(\exists \text{ a set of size } k \text{ with } \geqslant k-1 \text{ edges}\right) \leqslant (1/2)^k \cdot 40n \leqslant (1/2)^{k_0} \cdot 40n = (1/n)^{10} \cdot 40n \leqslant 1/n^8,$$

as $n \to \infty$. Taking another union bound over $n - k_0 + 1$ choices of k in $[k_0 : n]$, we get that

$$\begin{aligned}&\Pr\left(\exists \text{ a connected component of size } \geqslant 10\log n\right)\\ &\qquad\leqslant \Pr\left(\exists \text{ a set } S \text{ of size } \geqslant 10\log n \text{ with } \geqslant |S|-1 \text{ edges}\right) \leqslant 1/n^7,\end{aligned}$$

concluding the proof. $\square$

2.2 Lemma 4: Induced Edge Counts in Random Graphs

Lemma (Restatement of Lemma 4)**.** *Any graph $G=(V,E)$ sampled from $\mathbb{G}(n/100, n)$ with high probability satisfies: for any set $S \subseteq V$ of vertices, there are at most $5\cdot|S|$ edges between the vertices in S.*

Proof. This time, we simply take the witness to be a set S with more than $5\cdot|S|$ edges inside. Note that these edges may not necessarily be different from each other, i.e., we consider each parallel edge separately.

[3] A series of events $E_1, \ldots, E_t$ are called **negatively correlated** if

$$\Pr\left(E_1 \wedge E_2 \wedge \cdots \wedge E_t\right) \leqslant \prod_{i=1}^{t} \Pr\left(E_i\right).$$

However, the way we count this is to say that at least $5|S|$ out of m sampled edges should be inside S. So, for any fixed $S \subseteq V$ with size $|S| = k$,

$$\Pr\left(S \text{ has } \geqslant 5k \text{ edges inside}\right) \leqslant \binom{n/100}{5k} \cdot \left(\frac{\binom{k}{2}}{\binom{n}{2}}\right)^{5k};$$

here, first term chooses $5k$ of the $n/100$ samples and the second term is the probability that these samples are all in S. We can now do a union bound over all choices of S with a fixed size $k \geqslant 1$ and have,

$$\begin{aligned}
\Pr\left(\exists \text{ a set } S \text{ of size } k \text{ with} \geqslant 5k \text{ edges inside}\right) &\leqslant \binom{n}{k} \cdot \binom{n/100}{5k} \cdot \left(\frac{\binom{k}{2}}{\binom{n}{2}}\right)^{5k} \\
&\text{(the first term choose } k \text{ vertices of } S \text{ from } n \text{ vertices of } V\text{, and the rest is from above)} \\
&\leqslant \left(\frac{e \cdot n}{k}\right)^{k} \cdot \left(\frac{e \cdot n}{500 \cdot k}\right)^{5k} \cdot \left(\frac{k \cdot (k-1)}{n \cdot (n-1)}\right)^{5k} \\
&\qquad \text{(using } \tbinom{a}{b} \leqslant (e \cdot a/b)^{b} \text{ inequality)} \\
&\leqslant \left(\frac{e \cdot n}{k}\right)^{k} \cdot \left(\frac{(k-1)}{100 \cdot (n-1)}\right)^{5k} \qquad \text{(as } e/500 < 1/100\text{)} \\
&\leqslant \left(\frac{n}{k}\right)^{k} \cdot \left(\frac{k}{n}\right)^{5k} \left(\frac{e}{10^{10}}\right)^{k} \\
&\qquad \text{(as } (k-1)/(n-1) \leqslant k/n \text{ and } 100^{5} = 10^{10}\text{)} \\
&= \left(\frac{k}{n}\right)^{5k} \cdot \left(\frac{1}{10}\right)^{9k}; \qquad \text{(as } e < 10\text{)}
\end{aligned}$$

now note that either $1 \leqslant k \leqslant n^{0.1}$ or $n^{0.1} < k \leqslant n$; in the first case:

$$\left(\frac{k}{n}\right)^{5k} \cdot \left(\frac{1}{10}\right)^{9k} \leqslant \left(\frac{1}{n^{0.9}}\right)^{5k} \cdot 1 \leqslant \frac{1}{n^{4.5}};$$

in the second case:

$$\left(\frac{k}{n}\right)^{5k} \cdot \left(\frac{1}{10}\right)^{9k} \leqslant 1 \cdot \left(\frac{1}{10}\right)^{n^{0.1}} \ll \frac{1}{n^{100}},$$

for $n \to \infty$. A union bound over all n choices of k then implies the claim, concluding the proof. □

Remark. In the proofs of both Lemma 3 and Lemma 4 we have been *extremely* relaxed with the constants and inequalities; despite this, we still obtained the desired bounds (this is common in "these types of probabilistic proofs" where we leave 'these types' part entirely undefined on purpose–you know it, when you see it!). In principle, we could have tried to optimize the bounds and that would have, in the limit, allowed us to prove the same result as in Theorem 1 for $m = (1-\varepsilon) \cdot n$ for any constant $\varepsilon > 0$.

Nevertheless, when $m = n$, this approach would not work for an important reason: a random graph with n random edges with high probability has a connected component of size $\sim n^{2/3}$ vertices with high probability (as opposed to $O(\log n)$ in Lemma 3) and that would completely break the arguments used in this proof. The bound on the size of this connected component is related to the notion of "giant component" in random graphs that is a fascinating topic in random graph theory but is beyond the scope of this course[a].

[a] Just for some context: when $m = (1+\varepsilon) \cdot n$ for a constant $\varepsilon > 0$, then with high probability there is a single connected component of size $\Omega(n)$—called the giant component—with all other components being of size only $O(\log n)$!

This concludes our study of power of two choices in balls and bins experiments and the introduction/background part of this course.

References

[KLMadH92] Richard M. Karp, Michael Luby, and Friedhelm Meyer auf der Heide. Efficient PRAM simulation on a distributed memory machine. In S. Rao Kosaraju, Mike Fellows, Avi Wigderson, and John A. Ellis, editors, *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 318–326. ACM, 1992. 2

[Mit01] Michael Mitzenmacher. The power of two choices in randomized load balancing. *IEEE Trans. Parallel Distributed Syst.*, 12(10):1094–1104, 2001. 1