Lecture 2

1 Power of two choices...

Today we will analyze the strategy of assigning n balls to n bins by sequentially choosing the least loaded of two random bins?

What about strategy 2 or choosing the least loaded of two random bins? What is the max-load? n ..uh..no $n/2$!! Oh, shut up.

Ok, any guesses? Better by a constant factor? Maybe as good as $O(\sqrt{\log n})$? Hmmm...

Let's try to analyze this.

1.1 A random graph and "removal process".

First, we consider throwing only $n/8$ balls into n bins. We define a graph on n nodes, where an edge is defined for each ball according to its two choices of bins to look at. That is, if the first ball examines bins 1 and 9, the graph contains an edge $(1, 9)$.

We consider the following process. We iteratively removing all nodes (and incident edges) with remaining degree less than 12. How many iterations does this take? (We note that the number of iterations will give us an upper bound on the max load of the balls in bin strategy. The idea is that low degree bins can't have high load. The proof is a bit subtle and we defer it to later.)

Let's consider some properties of the graph.

What is the average degree of this graph? (1/4.)

What is the maximum degree of this graph, with high probability? (Certainly, $O(\log n)$.)

What is the maximum size of any connected component?

Recall, we added $n/8$ edges to a graph with n nodes, by choosing the endpoints for each edge uniformly at random. We proved the following claim.

Claim 1
The maximum component size is $O(\log n)$ with high probability.

Proof:
We wish to bound $Pr[$ there is a connected component of size $\geq k]$.

We first observe that

$$Pr[\text{ there is a connected component of size } \geq k] = Pr[\text{there is a connected component of size } k],$$

since any component with k or more nodes contains a connected component with k nodes.

We then observe that

$$Pr[\text{there is a connected component of size } k] \leq Pr[\text{there is a subset of } k \text{ nodes with} \geq k-1 \text{ internal edges}],$$

since any connected graph has to have at least $k-1$ edges. (This follows from the argument that adding an edge to a graph can only decrease the number of components by at most one, and noting that a k node graph with no edges has k components.)

Notice that the latter event is the union of events of the form "this subset S of k nodes has this set E of $k-1$ internal edges".

For any particular subset, S, and E, the probability that that this event holds is at most

$$\left(\frac{k}{n}\right)^{2(k-1)},$$

since each of the $k-1$ edges must choose both of its endpoints to be in the set S and the probability of a randomly chosen vertex being in S is k/n. Here, we applied the product rule since the choices are independent.

The union of $\binom{n}{k}\binom{n/8}{k-1}$ such events comprises the event that "there is a subset of k nodes with $\geq k-1$ internal edges."

Now, applying the union bound, we can upper bound our probability by

$$\binom{n}{k}\binom{n/8}{k-1}\left(\frac{k}{n}\right)^{2(k-1)} \leq \left(\frac{ne}{k}\right)^k \left(\frac{ne}{8(k-1)}\right)\left(\frac{k}{n}\right)^{2(k-1)} \leq n\frac{(ek)^{2(k-1)}}{8^{k-1}(k-1)^{2k-1}} \approx n\left(\frac{e^2}{8}\right)^{k-1}.$$

Using the inequality that $\binom{n}{k} \leq (ne/k)^k$, yields an upper bound of $n(e^2/8)^{k-1}$. When $k \geq 2\log_{e^2/8} n$, this upper bound is less than $1/n$.

$\square$

Now, we examine a more subtle concept (or at least a generalization of what we did above.) The **internal average degree of a subset** S is the average degree on the induced graph on S. For example, a triangle has average internal degree of 2 regardless of how many edges are incident to it from nodes outside of the triangle.

What is the maximum internal average degree of any subset S of nodes?

Let us consider k sized subsets.

The probability that the internal degree is greater than 6 is at most

$$\binom{n}{k}\binom{n/8}{3k}\left(\frac{k}{n}\right)^{6k},$$

where $\binom{n}{k}$ is the number of subsets of size k, $\binom{n/8}{3k}$ is the number of sets of edges of size $3k$ and $(k/n)^{6k}$ is the probability that both endpoints of a set of $3k$ edges are internal to a set of k nodes. Here, we used the union bound twice.

The expression above can be bounded by $(e^2/24)n^2$ for any k. Using the union bound over choices of k, we get that the "failure" probability is at $O(1/n)$. Thus, we have the following claim.

Claim 2

With probability greater than $1-O(1/n)$, the average internal degree of every subset is at most 6.

Now we can bound the number of iterations that our removal process requires.

Consider each connected component separately. At least half the nodes have degree less than 12, by the claim above (and thinking about a statement of the form that "everyone can't be better than average"). Thus, each iteration halves the size of every connected component. Since the maximum size of any connected component is $O(\log n)$, all the nodes and edges are removed in $O(\log \log n)$ iterations. That is, we have that.

Claim 3
The removal process terminates in $O(\log \log n)$ iterations.

2 And max-load is...

But what is the maximum load in strategy 2? Well, let's consider a ball (which corresponds to an edge in the graph.) We define its height in the solution, as the number of balls already in the bin in which it is placed at the time it was placed into the bin. What is an upper bound on the height of the ball? Note that we are defining the height of the ball, not the bin. We make the following claim.

Claim 4
The height of a ball is at most 12 times the iteration number in which its corresponding edge is removed in the graph removal process above.

Clearly, since the max height of any ball is also the max load, this claim along with claim 3 allows us to conclude that the maximum load is $O(\log \log n)$ with high probability!! An exponentially smaller load than what is obtained by choosing a single bin.

The claim is really an inductive claim. If an edge is removed in the first iteration, the corresponding ball could have been placed in a bin that was only chosen by 12 balls, thus its maximum height is 12.

Question 1. Flesh out this inductive proof sketch. You may wish to define a bit of notation.