# Lecture 2

## 1 A comment about "with high probability."

Last time, I somewhat arbitrarily used $1/n$ as the probability of "failure." Specifically, I found a value $k$ for which the maximum load was more than $k$ (this is "failure") with probability at most $1/n$.

Why did I choose $1/n$? For example, I could have chosen 1% or .1% or some other number. In general, the notion is that we would like the probability of failure to be small. One notion of small is that it goes to 0. For theorists, we tend to think of $n$ as getting large so $1/n$ tends to go to zero. Other expressions go to zero as well, e.g., $1/\log n$, but somehow $1/n$ is convenient. Moreover, in this case, with respect to the bound on the maximum load $k$, choosing $1/n$ rather than $1/\log n$ only affects the value of $k$ by a constant factor (actually by a small order term if one were more careful in the analysis.)

I am totally happy to have such discussions in class. As you see, I can expound happily on such topics. Still, I believe it better to have a conversation about these topics.

## 2 Wrapping up previous lecture.

Recall that last time we were discussing the strategy of assigning balls to bins where the balls sequentially choose two bins at random and picked the most lightly load bin. We set it up by choosing $n/8$ balls and $n$ bins and defined a random graph with a node for each bin, and an edge for each ball between its two random choices. We showed that the largest connected component was of size $O(\log n)$ with high probability, and that any subset had average induced degree at most 6 with high probability. We introduced the iterative process to remove edges from the random graph where in each iteration we remove all nodes of degree at most 12 and their incident edges. This removes all nodes in $O(\log \log n)$ time since the size of any connected component drops by at least a factor of two size the average degree 6 condition implies that at least half the nodes have degree at most 12. We will review some of these calculations again, below.

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 height of the bin in which it is placed at the time it was placed into the bin. What is the maximum height it could obtain. We make the following claim.

Claim 1
*The maximum height is at most 12 times the iteration number in which its corresponding edge is removed in the process above.*

The claim allows us to conclude that the maximum load is $O(\log \log n)$ with high probability!! An exponentially smaller load!

The claim is really an inductive claim. If an edge is removed in iteration 0, 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.

## 3 Probabiliy Review

We will discuss briefly some concepts in probability that we use often in our calculations. We start with the basics of discrete probability. A probability space is a set $\Omega$ (the sample space) and a function $Pr : \Omega \to \mathbb{R}$, such that $\sum_{\omega \in \Omega} Pr(w) = 1$. For example, flipping a coin gives us the probability space $\Omega = \{H, T\}$ with $Pr[H] = Pr[T] = 1/2$. Other examples; flipping $k$ coins which has a sample space of all possible $k$ length strings over the alphabet $\{H, T\}$ and where each is equally likely, i.e., $1/2^k$, choosing a five card poker hand whose sample space is the set of all five card poker hands, and the probability function is equally likely, i.e., 1/(525). We generally consider uniform probability spaces, ones where each sample point is equally likely.

Events in a probability space are just subsets of the sample points. For example, the event that we have at least 3 heads in $k$ coin tosses consists of the sample points with 3 or more heads. The probability of an event is the sum of the probabilities of the sample points that compose it. We can calculate probabilities of events using this method. For example, let $A$ be the event that we have at least 3 heads in $k$ coin tosses, we have

$$Pr[A] = \sum_{\omega \text{ with 3 heads}} \frac{1}{2^k} = \sum_{i \geq 3} \binom{n}{3} \frac{1}{2^k}.$$

The sum is over all the sample points with 3 heads, of the probability of each such sample point. The second expression counts the number of such sample points and multiplies by the (uniform) probability of the sample points.

We are often only concerned with approximately bounding probabilities. One thing we could do is to write an expression for the probability and then mathematically bound it. We often use other methods to simplify (or make possible) our calculations. We use the product rule for independent events, $A$ and $B$, i.e.,

$$Pr[A \cap B] = Pr[A] * Pr[B].$$

We also use the union bound, which states that for any two events $A$ and $B$

$$Pr[A \cup B] \leq Pr[A] + Pr[B].$$

Thus, we could write a bound on the probability that we have three heads as follows.

$$Pr[\text{ we have three heads }] \leq \sum_{i \neq j \neq k} Pr[\text{coin } i \text{ is heads}] * Pr[\text{coin } j \text{ is heads}] * Pr[\text{coin } k \text{ is heads}] \leq \binom{n}{3} 1/8.$$

In this case, we do not get a very good bound. (It is usually bigger than 1.) But, we use this "naive" method quite effectively in the last lecture.

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

CLAIM 2
*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 observed that

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

since any component with $k$ or more nodes contains a connected component with $k$ nodes.
We then observed that

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

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 $S$. 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 rule, we can upper bound our probability by

$$\binom{n}{k}\binom{n/8}{k-1}(\left(\frac{k}{n}\right)^{2(k-1)}.$$

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

**Question 2:** Write out more carefully than in the first lecture, the proof of claim 2 of lecture 1, that average induced degree of every subset of nodes in the random graph is smaller than 6, with high probability. Basically, explain (or correct if necessary) the calculation there.

## 4 Expectation: Linearity

We consider a random variable $X$ in some probability space, (which using the above development is defined as a function on a sample space.)

The expectation of the random variable is defined to be

$$E[X] = \sum_{a_i} a_i Pr[X = a_i].$$

For example, in a situation where we throw a ball into one of $n$ bins. Let $X$ be the $0-1$ valued random variable that indicates whether the ball fell into bin 1. It is easy to see that $E[X] = 1/n$ from the definition of expectation. On other hand, using this definition for the situation of throwing $n$ balls into $n$ bins, let's define $X$ as the load on bin 1. What is its expectation? Well,

$$E[X] = \sum_i i * Pr[X = i] = \sum_i i * \binom{n}{i} (1/n)^i (1 - 1/n)^{n-i}.$$

Yuck. We know it is 1. But, the formal reason is *linearity of expectation*, which states that for a random variable $X = Y + Z$,

$$E[X] = E[Y] + E[Z].$$

**Question 2:** Prove this inequality using the definition of expectation.

Now, we can define our random variable $X$ which is the number of balls in bin 1, as the $X = \sum_i X_i$, where $X_i$ is the $0 - 1$ valued random variable that indicates whether ball 1 chose bin 1. Now,

$$E[X] = \sum_i E[X_i] = \sum_i 1/n = 1.$$

So, here are a couple of fun diversions using linearity of expectation. First, if you have $n$ buses come at random throughout the day. What is your expected time wait time? Well the time between buses ought to be $24/n$..so $24/2n$?

First let's examine the expected time beteen busses. Well, we can look at this as placing $n$ points on a length 24 circle. We have a random variable for the length $X$ along the circle between busses.

Now, we arbitrary label the intervals between the points $1, 2, 3..n$. And consider a random variable $X_1, ..X_n$ for the length of the interval. We know the following

$$\text{Length of the circle} = \sum_i X_i.$$

Using linearity of expectation, we have

$$\text{E[Length of the circle]} = \sum_i E[X_i].$$

Since, each $X_i$ has the same distribution, we get that $E[X] = 24/n$ as we thought. But, what about our wait time? Well, our arrival is just another random point, so it is $24/(n+1)$.

Another example, is the following. Lets draw an infinite number of parallel lines unit distance apart in the plane. Lets drop a unit length needle at random on the plane. What is the expected number of times the needle crosses a line in the plane?

Here, lets break the needle into very small pieces of length $\epsilon$. By linearity of expection, $E[X] = 1/\epsilon E[X_\epsilon]$. Now, lets put together these very small needle pieces into a circle of diameter 1. How many of these pieces do we need? Well, $2\pi/\epsilon$ of them. Now, let's throw done the circle at random on the plane with a random spin. We have

$$E[\text{ number of times the circle crosses a line }] = 2\pi/\epsilon E[X_\epsilon].$$

Well, the circle always crosses a line twice. So, we have

$$E[X_\epsilon] = \epsilon/\pi,$$

which allows us to conclude that the expected number of times a unit length needle crosses the lines is $1/\pi$.