Prereq Review #3: Indicator Random Variables, Linearity of Expectation

Ian Tullis, CS161 Summer 2022

Expectations

When analyzing some algorithms – particularly randomized algorithms – we will be using the concept of an **expectation** of a random variable. What is a random variable? Think of it as being something like the value that comes up on a roll of a (not necessarily fair) die. In the case of a fair six-sided die, that random variable would be 1 with probability $\frac{1}{6}$, 2 with probability $\frac{1}{6}$, and so on. That is, a random variable is governed by some probability distribution, which in this case happens to be uniform across all possible values of the die. (If X is the number that comes up on the die, then $P(X = i) = \frac{1}{6}$ for each i in $[1, 2, 3, 4, 5, 6]$, and 0 otherwise.)

But random variables can be governed by other, non-uniform distributions as well. For example, if X is a random variable for the total number of heads that you get in 2 fair coin flips, then it turns out that $X = 0$ with probability $\frac{1}{4}$, $X = 1$ with probability $\frac{1}{2}$, and $X = 2$ with probability $\frac{1}{4}$.[1]

The expectation $\mathbf{E}[X]$ of a random variable X is defined as the sum, over all values X can take on, of each value times the probability of that value. That is, the expectation is a *weighted average*. So, for example, the expectation of our variable X above is:

$$0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

In CS161, when we talk about the average-case running time t_R of a randomized algorithm R, we are taking an expectation $\mathbf{E}[t_R]$. Now, because R is a randomized algorithm, t_R is a random variable. What is the probability distribution associated with this random variable? It depends on how R works. For example, suppose that R has a 90% chance of finishing in 1 second, but takes 11 seconds the rest of the time. Then $\mathbf{E}[t_R] = 0.9 \cdot 1 + 0.1 \cdot 11 = 2$.

One important thing to keep in mind in this class: unless stated otherwise, the expectation of the running time of a randomized algorithm is taken with respect to the randomness in how the algorithm works, *not* with respect to the different possible inputs to the algorithm.

[1] The idea is that the four possible (and equally likely) outcomes are HH, HT, TH, TT.

Indicator Random Variables

Consider an event that either happens (with some probability p) or doesn't (with probability $1-p$). An indicator random variable I is a special kind of random variable based on such an event, and it equals 1 if the event happens and 0 otherwise.

What is the expectation of an indicator random variable? Let's calculate it directly:

$$\mathbf{E}[I] = 0 \cdot P(I=0) + 1 \cdot P(I=1) = P(I=1) = p$$

So the expectation of an indicator random variable is just the probability p of the underlying event! That's convenient!

Let's think about a concrete example. On the last page, we considered a random variable X for the number of heads in 2 fair coin flips, and calculated that $\mathbf{E}[X] = 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$. Is there another way we could have done that?

Let I_1 and I_2 be indicator random variables for whether the first and second coins, respectively, come up heads. Then convince yourself that the total number of heads is given by

$$X = I_1 + I_2$$

That is, suppose the first coin comes up heads and the second one doesn't. Then in this case we have $X = 1 + 0 = 1$.

What if we take the expectation of both sides of the above equation? Then we get:

$$\mathbf{E}[X] = \mathbf{E}[I_1 + I_2]$$

Linearity of Expectation

There is a property of expectation that might initially seem bland, but can let us do some amazing things. It turns out that **the expectation of any sum of random variables (not just indicator random variables) is the sum of the individual expectations of those variables.**[2] So in the above case, we can write

$$\mathbf{E}[X] = \mathbf{E}[I_1 + I_2] = \mathbf{E}[I_1] + \mathbf{E}[I_2]$$

Now, as we argued above, the expectation of an indicator random variable is just the probability p of the associated event. In this case, for each of I_1 and I_2, this p is $\frac{1}{2}$, since the coins

[2] We won't prove this here, but it's surprisingly easy to prove. It follows directly from the definition of expectation.

are fair. So we get $\mathbf{E}[I_1] + \mathbf{E}[I_2] = \frac{1}{2} + \frac{1}{2} = 1$.

This is the same answer we got before for $E[X]$. But notice that in this case, we never needed to know that the probability of getting 0 heads out of 2 was $\frac{1}{4}$.

Let's drive home the point. What if we wanted to know the expected number of heads X when flipping 10000 coins? Now it would be cumbersome to calculate this using the definition of expectation, because we would have to find each of $P(X = 0), P(X = 1), ..., P(X = 10000)$. But using linearity of expectation, it's easy. Let $I_1, ..., I_{10000}$ be the indicator random variables for whether the first, ..., ten thousandth coins come up heads:

$$\mathbf{E}[X] = \mathbf{E}[I_1 + ... + I_{10000}] = \mathbf{E}[I_1] + ... + \mathbf{E}[I_{10000}] = p + ... + p \text{ (10000 times) } = 10000 \cdot \frac{1}{2} = 5000$$

But That's Not All!

Maybe the previous result seemed pretty intuitively clear even without going through all that rigamarole with indicator random variables. After all, if we flip 10000 coins, and on average they should be half heads, what else should the weighted average be but half of 10000?

That was easy to reason about because the 10000 coins are independent. (Informally, knowing the result of any subset of them does not tell us anything about the results of any subset of the others.) But what if we want to take the expectation of a sum of variables that are not independent? It turns out that **linearity of expectation doesn't care**. It still holds.[3]

I can't overstate how powerful this is. Let's consider an example. Suppose that we have a list of the integers from 1 to n, in order. We then randomly permute the list, such that any of the possible orderings is equally likely. What is the expected number of values that remain in their original positions?

Let's approach this using indicator random variables: I_1 is for whether element 1 ends up in original spot, I_2 is for whether element 2 ends up in its original spot, and so on. Now there is a critical difference from our coin-based examples: **these variables are not independent**. To see this, suppose that we know that I_1 through I_{n-1} are all 1, i.e., that elements 1 through $n - 1$ are all in their original spots after the permutation. But then it has to be the case that $I_n = 1$ as well, because there is nowhere else for element n to go. Knowing something about a subset of the variables told us something about another of them, so they are not independent.[4]

[3] This also turns out to follow, in a pretty simple way, from the definition of expectation.

[4] They are not even pairwise independent. Knowing that element 57 is in its original spot actually makes us just a little more confident that element 16 is in its original spot as well, since we now know that 16 is not in 57's original spot.

So if we tried to figure out the exact joint probability distribution of I_1 through I_n, it would be a nightmare. But we don't have to! Even despite the non-independence, we can say

$$\mathbf{E}[I_1 + ... + I_n] = \mathbf{E}[I_1] + ... + \mathbf{E}[I_n]$$

Now, just arguing by symmetry, all of those terms should be the same, since there is nothing different about the element 1 versus the element 16 versus the element n. And what is $\mathbf{E}[I_1]$, for example? The probability of element 1 randomly ending up back in its correct spot (out of the n possible spots) is $\frac{1}{n}$. So the overall answer is $n \cdot \frac{1}{n} = 1$. Notice that it (perhaps surprisingly) doesn't even depend on n...

As another example, suppose you have a very long string of English uppercase letters in which each letter is uniformly randomly chosen. How many instances of a string like `ROAR` show up, in expectation? Solving this problem without using linearity of expectation would be a nightmare – whether a `ROAR` starts at position 1 of the string is clearly not independent of whether one starts at position 2, or 3, or 4. (There can't be one starting at positions 2 or 3, but it is *more* likely to have another overlapping one start at position 4, forming `ROAROAR`!) But if we use linearity of expectation, this problem becomes very simple – the answer ends up being $n - 3$ (the number of possible starting points for a `ROAR`) times the probability that any one contiguous block of four is `ROAR`, which is $(\frac{1}{26})^4$.