Discussion 9B Recap

Tyler Zhu

April 3, 2020

1 Recall

A **random variable** is a quantity that we take to vary randomly. Formally, given some probability space $(\Omega, \mathbb{P})$, we assign quantities to each sample $\omega \in \Omega$ so that these quantities can have probabilities.

For example, suppose I have a random variable $D =$ the outcome of a standard 6-sided die roll. Then the probability space is $\{(1, \frac{1}{6}), (2, \frac{1}{6}), \ldots, (6, \frac{1}{6})\}$. Then

$$\mathbb{P}(D = 1) = \mathbb{P}(D = 2) = \cdots = \mathbb{P}(D = 6) = \frac{1}{6}.$$

In this case, since every sample event has the same probability, we would call this a *uniform* space.

The **expected value** of a random variable X is just the "average" of its values weighted by their probabilities. We denote it as such:

$$\mathbb{E}[X] = \sum_{x \in \mathcal{A}} \mathbb{P}(X = x)x$$

where $\mathcal{A}$ is the set of values that X takes. Using the same example as above, the average die roll that you'd expect is

$$\mathbb{E}[D] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \cdots + \frac{1}{6} \cdot 6 = 3.5,$$

which is just in the middle. You'll notice that D never actually takes on this value, and that's fine (and most often the case).

2 Computing Expectations

Expectation is useful to know, since it gives us a good easy summary of the values a random variable takes on (often looked at in combination with the *variance*, which we'll see later). Here's a surprising fact about expectation that makes computing them *much* easier.

Theorem 1 (Linearity of Expectation). For any constant c and any two random variables X, Y, *not necessarily independent*, we have $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ and $\mathbb{E}[cX] = c\mathbb{E}[X]$.

It's hard to say more about why this is true other than the proof is a double counting proof; it turns out we can forget about whether X and Y are dependent by computing the expected value as a sum over the sample points $\omega \in \Omega$ rather than over the values that X and Y take as in our definition above.

This tool unlocks one of the most powerful techniques in probability for computing expectations; that of **indicator random variables**. Simply put, an indicator r.v. I takes on either 1 or 0, and usually indicates whether some sub-event is true or not. From our definition of expected value,

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

so we only have to worry about computing probabilities. The key idea is that most quantities can be broken up as sum of indicators, which combined with linearity of expectation, can make quick work of expectations. Here's the general idea.

1. Have some quantity X that we want the expectation of.
2. Break X into the sum of indicators, $X = X_1 + X_2 + \cdots + X_n$.
3. By Linearity of Expectation, $\mathbb{E}[X] = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]$.
4. Since $\mathbb{E}[X_1] = \mathbb{P}(X_1 = 1)$, we just compute these probabilities of small events and sum.

Example 2.1. There are n kids gathered in a room. Everyone takes their hats off and puts it in a pile in the middle. Then the hats are randomly distributed back to the kids. What is the expected number of kids that get their own hat back?

Solution. Let X be a r.v. counting the number of kids that get their own hat back. We can break it down as $X = X_1 + \cdots + X_n$, where X_i is 1 if kid i gets their hat back, and 0 otherwise. Then $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i]$ as we saw above, so we just need to find $\mathbb{P}(X_i = 1)$. This happens only if kid i gets their own hat back, which happens with probability $\frac{1}{n}$. So

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{1}{n} = \boxed{1}.$$

□