

## Lecture 22: Expectation

Readings: Section 19.4, 19.5.

### 1 Recap

So far, we saw

- random variables: (total) functions from the sample space  $S$  to another set (typically, non-negative real numbers).
- indicator random variables: functions that map to  $\{0, 1\}$ .
- random var  $R$  to events  $[R = x] := \{\omega : R(\omega) = x\}$  for any  $x \in \text{range}(R)$ .
- an event  $E$  to an indicator random var  $I_E(\omega) = 1$  if  $\omega \in E$  and 0 otherwise.
- independence of random variables: given a sample space  $S$ , for all  $x_1 \in \text{range}(R_1)$  and  $x_2 \in \text{range}(R_2)$ , the events  $[R_1 = x_1]$  and  $[R_2 = x_2]$  are independent.
- probability mass functions and cumulative distribution functions.

### 2 Expectation

This lecture is about the notion of expectation of a random variable, and a really cool property called linearity of expectation which makes computing expected values so much easier.

**Definition 1** (Expected value of a random variable, or the average, or the mean).

$$\text{Ex}[R] := \sum_{\omega \in S} R(\omega) \cdot \text{Pr}[\omega]$$

Example 1: A single throw of a die, sample space  $S = \{1, 2, 3, 4, 5, 6\}$ .  $R$  is the number that comes up, so  $R(\omega) = \omega$ .

$$\text{Ex}[R] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

The expected value of the dice is NOT necessarily the value you ever expect to see! Chalk it up to the weirdness of the language of mathematics.

The expectation of a random variable with a uniform distribution on a set  $\{a_1, \dots, a_n\}$  is simply the average of the  $a_i$ 's.

Example 2: The expected value of an indicator random variables is the probability that it takes on the value 1. If  $I_A$  is the indicator random variable for an event  $A$ ,

$$\begin{aligned} \text{Ex}[I_A] &= \sum_{\omega \in S} I_A(\omega) \cdot \Pr[\omega] \\ &= \sum_{\omega \in \bar{A}} 0 \cdot \Pr[\omega] + \sum_{\omega \in A} 1 \cdot \Pr[\omega] \\ &= 0 \cdot \Pr[I_A = 0] + 1 \cdot \Pr[I_A = 1] \\ &= \Pr[I_A = 1] = \Pr[A] \end{aligned}$$

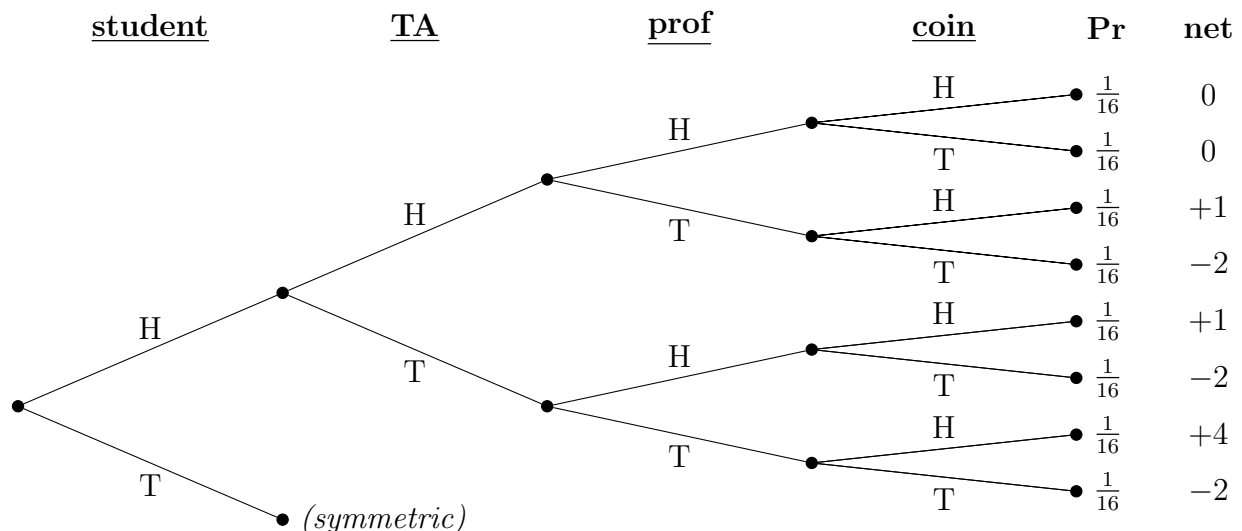
where the latter equality is by the definition of  $I_A$ .

### 3 How to win the lottery

Example 3: Each of three people (student, TA, prof) place an ante (two candies each) guess heads/tails. A coin is tossed. If all three agree or disagree with the coin, all three get their candies back. If two agree with the coin, the pot of 6 candies is split between the two (so, one person loses two candies and the other two gain one each). If one agrees with the coin, she gets the entire pot (so, she gains 4 candies and the other two lose two each.)

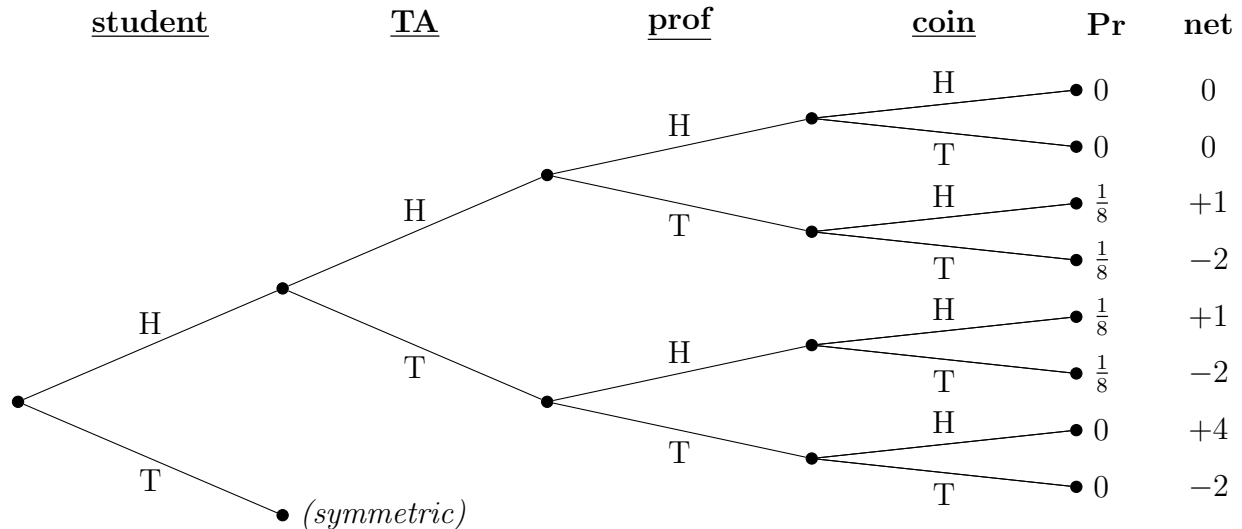
(play game)

Is the game fair? Let's see. Here is the tree diagram:



Let  $P$  be the random variable denoting the student's payoff.  $\text{Ex}[P] = 0$ .

What if the TA and prof decide beforehand that they are going to collude, i.e. choose opposite guesses always. This changes the probability distribution and consequently, the expected payoff of the student.



Now,  $\text{Ex}[P] = -1/2$ .

What just happened? Intuitively, the effect of the prof and TA always choosing opposite guesses is that the student never has a chance to win alone and pocket the sizeable sum of 4, and that shows up in her diminished payoff of  $-1/2$ .

(Professor Hermann Chernoff's story: see textbook)

Back to math...

## 4 Alternate ways to compute the expectation

**Theorem 1.**

$$\text{Ex}[R] = \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x]$$

This formula is more efficient when there are lots of outcomes, but fewer values that  $R$  can take.

The proof is similar to what we saw for indicator variables earlier; full details in the lecture notes if you're interested.

*Proof details; skipped during lecture.*

$$\begin{aligned}
 \text{Ex}[R] &= \sum_{\omega \in S} R(\omega) \cdot \Pr[\omega] \\
 &= \sum_{x \in \text{range}(R)} \sum_{\omega: R(\omega)=x} R(\omega) \cdot \Pr[\omega] \\
 &= \sum_{x \in \text{range}(R)} \sum_{\omega: R(\omega)=x} x \cdot \Pr[\omega] \\
 &= \sum_{x \in \text{range}(R)} x \cdot \sum_{\omega: R(\omega)=x} \Pr[\omega] \\
 &= \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x]
 \end{aligned}$$

which is what we wanted. QED. □

**Example:** the expected number of hearts in a 3-card hand. Definition would be a sum with  $\binom{52}{3}$  terms, since that's the number of outcomes. Or, a sum with just 4 terms (really 3):

$$\begin{aligned}
 &0 \cdot \Pr[H = 0] + 1 \cdot \Pr[H = 1] + 2 \cdot \Pr[H = 2] + 3 \cdot \Pr[H = 3] \\
 &= 1 \cdot \frac{\binom{13}{1} \binom{39}{2}}{\binom{52}{3}} + 2 \cdot \frac{\binom{13}{2} \binom{39}{1}}{\binom{52}{3}} + 3 \cdot \frac{\binom{13}{3}}{\binom{52}{3}}.
 \end{aligned}$$

Note: this example becomes even easier using linearity of expectation (introduced later this lecture). Try it out yourself! Hint: the answer is much simpler than the formula above might indicate.

If the range of  $R$  is the natural numbers, we get an infinite sum, which can be rewritten in a useful way:

**Theorem 2.** If  $\text{range}(R) \subseteq \mathbb{N}$ ,

$$\text{Ex}[R] = \sum_{i=0}^{\infty} i \cdot \Pr[R = i] = \sum_{i=0}^{\infty} \Pr[R > i]$$

Why is this last equality true?

*Proof.*

$$\begin{aligned}
 \sum_{i=0}^{\infty} \Pr[R > i] &= \Pr[R > 0] + \Pr[R > 1] + \Pr[R > 2] + \dots \\
 &= (\Pr[R = 1] + \Pr[R = 2] + \Pr[R = 3] + \dots) + \\
 &\quad (\Pr[R = 2] + \Pr[R = 3] + \dots) + \dots
 \end{aligned}$$

$\Pr[R = i]$  appears  $i$  times, so this is exactly computing the expectation. □

## 5 Mean Time to Failure

Manufacture parts, each part *independently* defective with probability  $p$ . What is the expected number of parts manufactured by the time you see the first defective one? (counting the defective one)

Let  $R$  be the random variable that denotes the number of parts manufactured by the time you see the first defective part.

$$\text{Ex}[R] = \sum_{i=0}^{\infty} \Pr[R > i] = \sum_{i=0}^{\infty} (1-p)^i = 1/(1 - (1-p)) = 1/p$$

## 6 Caveat about similar looking formulas

We've seen four formulas that look very similar:

$$\text{Ex}[R] = \sum_{\omega \in S} R(\omega) \cdot \Pr[\omega] \quad \text{Definition of } \text{Ex}[R]. \text{ Sum over outcomes } \omega$$

$$\text{Ex}[R] = \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x] \quad \text{Alternate formula for } \text{Ex}[R]. \text{ Sum over RV values } x$$

$$\text{Ex}[R] = \sum_{i=0}^{\infty} \Pr[R > i] \quad \text{Only valid when } \text{range}(R) \subseteq \mathbb{N}. \text{ Sum over natural numbers } i.$$

$$\text{Ex}[R_1 + R_2] = \text{Ex}[R_1] + \text{Ex}[R_2] \quad \text{Linearity of expectation. Sum over multiple RVs } R_i.$$

These formulas are often confused for each other or incorrectly merged together, e.g., summing over  $i \cdot \Pr[R > i]$  (which incorrectly combines the second and third formulas). Please be sure you know which of these (correct) formulas you are using, so you don't mix them up!

## 7 Infinite expectations

Suppose we want to estimate the expected latency on a communication channel. Strategy: try to measure 100 times and take the average (the so-called "empirical mean"). how well does the empirical mean approximate the real mean / expected value?

Say, the latency is  $i$  ms with probability  $1/i$ . the expected latency is then

$$\sum_{i=1}^{\infty} 1/i$$

which diverges. The expectation is unbounded while the empirical mean is finite.

## 8 Linearity of Expectation

**Theorem.** For random variables  $R_1$  and  $R_2$ ,

$$\text{Ex}[R_1 + R_2] = \text{Ex}[R_1] + \text{Ex}[R_2]$$

Compare: For events  $A$  and  $B$ ,  $\Pr(A \wedge B) = \Pr(A) \cdot \Pr(B)$ . Only true when  $A$  and  $B$  are **independent**. And  $\Pr[A \cup B] = \Pr[A] + \Pr[B]$ , but only when disjoint.

In contrast, linearity of expectation applies to any random variables. No need for independence! Which is what makes it a formidable tool in our arsenal.

*Proof details; skipped during lecture.*

$$\begin{aligned} \text{Ex}[R_1 + R_2] &= \sum_{\omega \in S} (R_1 + R_2)(\omega) \cdot \Pr[\omega] \\ &= \sum_{\omega \in S} (R_1(\omega) + R_2(\omega)) \cdot \Pr[\omega] \\ &= \sum_{\omega \in S} R_1(\omega) \cdot \Pr[\omega] + \sum_{\omega \in S} R_2(\omega) \cdot \Pr[\omega] \\ &= \text{Ex}[R_1] + \text{Ex}[R_2] \end{aligned}$$

□

Note: Also works for more than two random variables. But need a fixed number of random variables.

Powerful technique to compute  $\text{Ex}[R]$  via linearity of expectation: *Write  $R$  as a sum of two simpler random variables.*

**Example:**  $R$  is the sum of two fair dice.  $R = R_1 + R_2$  where  $R_1$  is the result of the first die and  $R_2$  the result of the second die. Then,

$$\text{Ex}[R] = \text{Ex}[R_1 + R_2] = \text{Ex}[R_1] + \text{Ex}[R_2] = \frac{7}{2} + \frac{7}{2} = 7$$

The dice did not have to be independent! The two dice could be perfectly correlated, i.e., always have exactly the same rolls, and the calculation we did above applies just as well!

## 9 Cellphone check problem

Before entering the final exam room, you check your cellphone by putting into a bag (together with everyone else's phones). When leaving, you pick a random one of the phones from the bag. What is the expected number of people who get their cellphone back?

Let  $R$  be the number of people who got their phones back.

$$\text{Ex}[R] = \sum_{k=0}^n k \cdot \Pr[R = k]$$

But what is  $\Pr[R = k]$ ?? How do we compute it?

Fortunately, linearity of expectation comes to the rescue. Let  $I_i$  be indicator random variables where

$$I_i = \begin{cases} 1 & \text{if the } i\text{th person gets their phone back} \\ 0 & \text{otherwise} \end{cases}$$

Now, and it's important you make sure you understand this:

$$R = I_1 + I_2 + \dots + I_n$$

So,

$$\text{Ex}[R] = \sum_{i=1}^n \text{Ex}[I_i] = \sum_{i=1}^n \Pr[I_i = 1]$$

where the second equality is because the expected value of an *indicator* random variable is just the probability that it is 1.

What is  $\Pr[I_i = 1]$ ? It's the probability that the  $i$ th person gets their cellphone back. This is  $1/n$ . (Why is this the case? Imagine the cellphones being permuted randomly and the  $i$ th person gets the  $i$ th cellphone in the permuted sequence. Then, the chance that anyone gets their phone back is  $1/n$ .)

So,

$$\text{Ex}[R] = n \cdot 1/n = 1$$

which is our answer. QED.

Incidentally, it turns out that

$$\Pr[R = k] = \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}$$

which is a nasty expression. Using this, we can compute the expectation, but it's a rather unpleasant calculation. Linearity of expectation makes it all so much easier. Plus it's much more general in that it deals with random variables which are correlated in arbitrary ways.

**Rotating table version.** Now, let's change the game a bit.  $n$  of us go to the restaurant [Mu Lan](#), place our cellphones on the lazy suzan in the middle of the table and give it an energetic spin. What's the expected number of us that get their phone back in front of us? Either we all get our phones back (which happens with probability  $1/n$ ), or none of us do

(which happens with probability  $1 - 1/n$ ). So, if  $R$  is the random variable as above, that counts the number of us that got our phones back in front of us, then

$$\text{Ex}[R] = n \cdot \frac{1}{n} + 0 \cdot \left(1 - \frac{1}{n}\right) = n \cdot \frac{1}{n} = 1.$$

That was simple enough, but the point is that the exact same calculation that we did above would've given us the answer here as well, even though the random variables  $I_i$  are *very* not independent. (Note that in the cellphones in a bag game as well, the  $I_i$  were not independent, but "less so" than the lazy suzan game.)



MIT OpenCourseWare  
<https://ocw.mit.edu>

6.1200J Mathematics for Computer Science  
Spring 2024

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>