Lecture 19: random variables and expectation

November 23, 2026

1 Multiplying probabilities

Recall the definition of conditional probability, and the corollary we got by rearranging it:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad \text{so} \qquad P(A \cap B) = P(A \mid B) \cdot P(B).$$

We can extend the second one to arbitrarily many events:

$$P(A \cap B \cap C \cap D) = P(A \mid B \cap C \cap D) \cdot P(B \mid C \cap D) \cdot P(C \mid D) \cdot P(D).$$

This is why we are allowed to "multiply probabilities" along a sequential procedure.

Example 1.1. I have 4 red, 3 blue, and 2 green balls in a bin. I pick 3 of them in a sequence. What is the probability that I pick red, then blue, then red?

Multiplying along the sequence,

$$\begin{aligned} P(\text{1st red} \cap \text{2nd blue} \cap \text{3rd red}) &= P(\text{1st red}) \cdot P(\text{2nd blue} \mid \text{1st red}) \\ &\quad \cdot P(\text{3rd red} \mid \text{1st red, 2nd blue}) \\ &= \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{3}{7}, \end{aligned}$$

since after a red is gone there are 8 balls left of which 3 are blue, and after a red and a blue are gone there are 7 balls left of which 3 are red.

When the events are independent, each conditional probability in the product is just the unconditional one: $P(A \mid \cdots) = P(A)$. This is what lets us multiply independent probabilities.

Example 1.2. I flip 4 coins. and let A, B, C, D be the events that the first, second, third, and fourth flip comes up heads.

- A is independent of $B \cap C \cap D$, so $P(A \mid B \cap C \cap D) = P(A)$.
- B is independent of $C \cap D$, so $P(B \mid C \cap D) = P(B)$
- C is independent of D, so $P(C \mid D) = P(C)$

Then

$$P(A \mid B \cap C \cap D) \cdot P(B \mid C \cap D) \cdot P(C \mid D) \cdot P(D) = P(A) \cdot P(B) \cdot P(C) \cdot P(D),$$

so

$$P(HHHH) = P(\text{1st } H) \cdot P(\text{2nd } H) \cdot P(\text{3rd } H) \cdot P(\text{4th } H) = \left(\tfrac{1}{2}\right)^4.$$

Example 1.3. I flip a biased coin that comes up heads 0.7 of the time. What are the odds that I get the sequence $HHHHHTTTTT$?

The flips are independent, so the product is $0.7^5 \cdot 0.3^5$.

Example 1.4. I flip a coin that comes up heads with probability p, n times. What is the probability that

(a) I get all heads?

(b) I get all tails?

(c) the first flip is tails and the rest are heads?

2 Random variables

Definition 2.1. *For a probability space (Ω, P) and some set S, an **S-valued random variable** X is a function $X : \Omega \to S$.*

Example 2.2. I roll 2 dice, so $\Omega = [6] \times [6]$ and $P(\omega) = 1/36$ for every outcome.

The sum of the two rolls is an integer-valued random variable:

$$X : [6] \times [6] \to \mathbb{Z} \quad \text{via} \quad X((a,b)) = a + b, \qquad \text{so} \qquad X((3,4)) = 7.$$

The max of the two rolls is also an integer-valued random variable:

$$Y : \Omega \to \mathbb{Z} \quad \text{via} \quad Y((a,b)) = \max(a,b), \qquad \text{so} \qquad Y((3,4)) = 4.$$

Whether I rolled a 6 is a $\{0,1\}$-valued random variable:

$$Z : \Omega \to \{0,1\} \quad \text{via} \quad Z((a,b)) = \begin{cases} 1 & a = 6 \text{ or } b = 6 \\ 0 & \text{otherwise,} \end{cases} \qquad \text{so} \qquad Z((3,4)) = 0.$$

A $\{0,1\}$-valued random variable is called an **indicator** random variable, because it indicates whether something happened.

Notation. We use random variables to define events, by overloading notation once more. For example, "$X = 7$" is the event that the sum is 7:

$$X = 7 \text{ is the event } \{w : X(w) = 7\} = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}.$$

Then we can write things like $P(X = 7)$, the probability that the sum is 7, or $P(Y \geq 3)$, the probability that the max roll is at least 3. This saves us from having to name a separate event for "the sum is 2," "the sum is 3," and so on — we can just refer to X.

2.1 Independent random variables

Definition 2.3. *We say that two random variables X and Y are independent iff for all a, b,*

$$P(X = a \text{ and } Y = b) = P(X = a) \cdot P(Y = b).$$

3 Expectation

The other nice thing about random variables is that they let us talk about *average outcomes*. When things are random, we want to know what happens on average, and that is hard to say in terms of events: if I flip 10 coins, what is the average outcome? $HHTTHTHTHT$, or something? But once the outcome is a number, like the number of heads, the average makes sense: on average, 5 heads.

Definition 3.1. *Let X be a real-valued random variable. The **expected value** of X is*

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$

Example 3.2. I roll a die, and my random variable is $X(\omega) = \omega$, the number on the die. What is $\mathbb{E}[X]$?

From the definition,

$$\mathbb{E}[X] = \sum_{\omega \in \{1,\ldots,6\}} X(\omega) \cdot P(\omega) = \sum_{\omega \in \{1,\ldots,6\}} \omega \cdot \tfrac{1}{6} = \tfrac{1}{6} \cdot (1+2+3+4+5+6) = \tfrac{21}{6} = 3.5.$$

Example 3.3. I flip 3 coins. Let Y be the number of heads. What is $\mathbb{E}[Y]$?

The space is uniform with 8 outcomes, so

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{\omega \in \Omega} Y(\omega) P(\omega) \\
&= \underbrace{0 \cdot \tfrac{1}{8}}_{TTT} + \underbrace{1 \cdot \tfrac{1}{8}}_{HTT} + \underbrace{1 \cdot \tfrac{1}{8}}_{THT} + \underbrace{1 \cdot \tfrac{1}{8}}_{TTH} + \underbrace{2 \cdot \tfrac{1}{8}}_{HHT} + \underbrace{2 \cdot \tfrac{1}{8}}_{HTH} + \underbrace{2 \cdot \tfrac{1}{8}}_{THH} + \underbrace{3 \cdot \tfrac{1}{8}}_{HHH} \\
&= \tfrac{12}{8} = 1.5.
\end{aligned}$$

Summing over outcomes gets tedious when there are many of them, and often several outcomes give X the same value. Grouping those together gives an equivalent definition.

Fact 3.4. $\mathbb{E}[X] = \sum_x x \cdot P(X = x)$*, where the sum is over the values x that X takes.*

Computing $\mathbb{E}[Y]$ that way instead:

$$\mathbb{E}[Y] = \sum_{x \in \{0,1,2,3\}} x \cdot P(Y = x) = 0 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{3}{8} + 3 \cdot \tfrac{1}{8} = \tfrac{12}{8} = 1.5,$$

the same answer.

Fact 3.5. *For an indicator random variable X, $\mathbb{E}[X] = P(X = 1)$.*

Proof. By Fact 3.4, since X only takes the values 0 and 1,

$$\mathbb{E}[X] = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = P(X = 1). \qquad \square$$

Example 3.6. I flip 4 coins. Let $X = 1$ if they all come up heads, and $X = 0$ otherwise. What is $\mathbb{E}[X]$?

It is an indicator, so $\mathbb{E}[X] = P(X = 1) = P(\text{all heads}) = 1/2^4$.

3.1 Duolingo streaks

I'm learning Spanish on Duolingo, and I want to maintain my streak. However, each day, I forget to do my lesson with probability p (and remember with probability $1-p$). Supposing each day is independent, how long can I expect my streak to last?

Let R mean remembered, and F mean forgot. The probability space is $\Omega = \{F, RF, RRF, RRRF, \cdots\}$. What is P?

$$P(n-1 \text{ days that I remember, then 1 day I forget}) = (1-p)^{n-1} \cdot p$$

Let X be the random variable corresponding to how many days my streak lasts (including the day on which we forgot, so $X(RRF) = 3$). What is the expectation of X?

We can try to compute this directly (either using the definition of expectation or Fact 3.4):

$$\mathbb{E}[X] = 1 \cdot p + 2p(1-p) + 3(1-p)^2 p + 4(1-p)^3 p \cdots$$

This is a bit annoying to compute. By calculus[1], this is finite, but what exactly is its value?

Let's expand out the definition of expectation and split into two groups:

$$\mathbb{E}[X] = \sum_{w \in \Omega} X(w)P(w) = \underbrace{1 \cdot p}_{\text{I forgot on the first day}} + \underbrace{\sum_{w=Rw'} X(w)P(w)}_{\text{I remembered at least once}}$$

Note two things:

- $X(Rw') = 1 + X(w')$: If our streak is R and then some further sequence w', the length of the streak is 1 more than the length of the streak w'.
- $P(Rw') = P(\text{first day R})P(w') = (1-p)P(w')$, since each day is independent.

We can plug these two in:

$$\mathbb{E}[X] = p + \sum_{w=Rw'} (1 + X(w'))((1-p)P(w')) \tag{1}$$

$$= p + (1-p)\left(\underbrace{\sum_{w' \in \Omega} P(w')}_{=1} + \underbrace{\sum_{w' \in \Omega} X(w')P(w')}_{\text{This is } \mathbb{E}[X]!} \right) \tag{2}$$

$$= p + (1-p)(1 + \mathbb{E}[X]) \tag{3}$$

Remark. This is a very interesting statement! The expected length of my streak has a probability p of being 1, and a probability $1-p$ of being 1 more than the expected length of my streak? Why does this make sense? This random variable is called "memoryless"; can you see why?

[1] Continuous math? heretical!

Since $\mathbb{E}[X]$ is finite (by calculus), we can rearrange:

$$\begin{aligned} & \mathbb{E}[X] = p + (1-p)(1+\mathbb{E}[X]) \\ \implies & \mathbb{E}[X] = p + 1 - p + (1-p)\mathbb{E}[X] \\ \implies & \mathbb{E}[X] = \frac{1}{p}. \end{aligned}$$

So the expected length of my streak is $1/p$. If I have a 10% chance of forgetting on any given day, my streak will on average last 10 days.

This random variable has a special name: it's called the *geometric* random variable. If I have a series of trials, each with probability p of failure, the number of trials before my first failure is a geometric random variable. In expectation, I get through $1/p$ trials before I get my first failure.

4 Linearity of expectation

We can combine random variables. When I roll two dice, let X_1 be the value of the first roll and X_2 the value of the second; then the sum is $Z = X_1 + X_2$. This is shorthand for the random variable defined by $Z(\omega) = X_1(\omega) + X_2(\omega)$.

Fact 4.1 (Linearity of expectation)**.** *For* any *random variables X_1 and X_2,*

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2].$$

More generally, for any random variables $X_1, \ldots, X_n$,

$$\mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$