Learning objectives:

- ☐ calculate the expectation (average) of a random variable,
- ☐ apply linearity of expectation to calculate the expectation of the sum of several random variables.

In the previous lectures, we used a deterministic approach for computing probabilities of events. These events were a set of outcomes from the sample space, such as the event you pick the door with a car in the Monty Hall problem. We might have questions about numerical values associated with a particular experiment, such as the number of times we get heads when flipping a fair coin, or how much money we will lose when playing roulette at the casino. You might also think about how many times you need to conduct an experiment to achieve some desired result. For example, when I was younger, I always wondered what toy I would get in my Happy Meals. The techniques in this lecture will allow us to analyze this problem (jump to the example). First, we need to introduce random variables, which map the outcomes in the sample space to some number. Let's warm up with an exercise about flipping coins.

But there's nothing random about running a computer code?

That might be true, but we can use computers to generate *pseudo*-random numbers. They depend on some kind of algorithm and seed to generate random numbers. There are some devices that can create truly random numbers, using thermal or atmospheric noise instead of deterministic methods. See this article for more information.

Example 1:

Suppose you toss four fair coins ("fair" means the probability of getting heads or tails is 50% when flipping any given coin). How many outcomes are there in the sample space? Also, define a function $f(x)$ which takes an outcome x in the sample space (x would be any quadruplet obtained after flipping all four coins, such as $x = (H, T, H, T)$), and returns the number of heads that occur. Compute $f((T, H, H, H))$ and $f((T, T, H, T))$. What is the average number of heads in four coin flips?

Solution:

The sample space has $2^4 = 16$ possible outcomes. Since f returns the number of heads in each outcome, then $f((T, H, H, H)) = 3$ and $f((T, T, H, T)) = 1$. Assuming the coins are fair, this means that *on average* we should get heads half the time with a particular coin. With four coins, this gives $0.5 \times 4 = 2$.

1 Random variables and expectation

As mentioned earlier, *random variables* map outcomes to some numerical value. You really should think of them as functions in which the *domain* is the entire sample space. The *codomain* is $\mathbb{R}$.

Random variables are functions!

The term "random variable" can be confusing because they are neither random nor variables. They are actually **functions** that take, as input, the sample space, and output some number.

Definition 1. *Given a sample space S, a **random variable** X is a function $X\colon S \to \mathbb{R}$.*

There are various ways to define the function that maps an outcome to a value, the simplest is by assigning a 0 or a 1, depending on the case. For example, we might want to assign a win (1) or a loss (0) or heads (1) or tails (0). These types of random variables are called **indicator random variables**.

Other types?

Yes! There are other types of random variables, such as uniform random variables and binomial random variables, however, we won't cover these in this course. Depending on the type of random variable, we might be interested in plotting the probability of a given value of a random variable. This is known as the *probability distribution function* (or PDF) for short).

Definition 2. *An **indicator random variable** X is a random variable such that $X\colon S \to \{0,1\}$.*

Indicator random variables are also called Bernoulli random variables or characteristic random variables.

The **expected value** (or **average**) of a random variable X is denoted by $E[X]$ and is the weighted sum of the probabilities of all possible values of the random variable.

Definition 3. *Let X be a real-valued random variable defined on a sample space S. The **expectation** (*expected value*) of X is*

$$E[X] = \sum_{s \in S} X(s)p(s). \tag{1}$$

That is, the expectation is the linear combination of the value of X for a given outcome s, weighted by the probability of that outcome, $p(s)$.

Note that for an indicator random variable, $X(s)$ is either 0 or 1. If we consider an event $A \subseteq S$ and assign the indicator random variable $X(s)$ to be 1 for $s \in A$ (0 otherwise), then the expectation of X is

$$E[X] = \sum_{s \in A} \underbrace{X(s)}_{=1} p(s) + \sum_{s \notin A} \underbrace{X(s)}_{=0} p(s) = \sum_{s \in A} p(s) = p(A).$$

Example 2:
Calculate the expected value of the number of heads when flipping four coins using Equation 1.

Solution:
Let us define a random variable $X(s)$ for the coin toss problem as follows:

$$X(s) = \begin{cases} 0 & 0 \text{ heads}, \\ 1 & 1 \text{ heads}, \\ 2 & 2 \text{ heads}, \\ 3 & 3 \text{ heads}, \\ 4 & 4 \text{ heads}. \end{cases}$$

The probability of any particular outcome is $\frac{1}{16}$ since there are 16 possible outcomes in the sample space. The expected value can be calculated as

$$E[X] = \sum_{s \in S} X(s)p(s) = \sum_{s \in S:\ X(s)=0} p(s) \cdot 0 + \sum_{s \in S:\ X(s)=1} p(s) \cdot 1 + \sum_{s \in S:\ X(s)=2} p(s) \cdot 2 + \sum_{s \in S:\ X(s)=3} p(s) \cdot 3 + \sum_{s \in S:\ X(s)=4} p(s) \cdot 4$$

There is only 1 outcome in which there are 0 heads, 4 in which there is 1 head, 6 in which there are 2 heads, 4 in which there are 3 heads and 1 in which there are 4 heads. The expectation is then $E[X] = 0 \cdot 1 \cdot \frac{1}{16} + 1 \cdot 4 \cdot \frac{1}{16} + 2 \cdot 6 \cdot \frac{1}{16} + 3 \cdot 4 \cdot \frac{1}{16} + 4 \cdot 1 \cdot \frac{1}{16} = 2$.

2 Linearity of expectation

Sometimes, you need to calculate the average of a random variable, which you can break up as a linear combination of other random variables, of which you know how to calculate their expected values. Well, we can calculate the expected value of our random variable of interest by using an important property called *linearity of expectation*. Ultimately, this theorem states that the expectation of the sum of two (or more) random variables is the sum of their expectations.

Theorem 1. *Let X_1 and X_2 be random variables on a sample space S. The expectation of $X_1 + X_2$ is*

$$E[X_1 + X_2] = E[X_1] + E[X_2]. \tag{2}$$

Furthermore, let a, b be real numbers. The expectation of $aX + b$ is

$$E[aX + b] = aE[X] + b \tag{3}$$

Proof. We can verify this theorem directly by substituting the definition of expectation (Equation 1) into Equation 2, which gives

$$E[X_1 + X_2] = \sum_{s \in S} p(s)(X_1(s) + X_2(s)) = \sum_{s \in S} p(s)X_1(s) + \sum_{s \in S} p(s)X_2(s) = E[X_1] + E[X_2].$$

Equation 3 can be proved in a similar fashion:

$$E[aX + b] = \sum_{s \in S} p(s)(aX(s) + b) = a \sum_{s \in S} X(s) + b \underbrace{\sum_{s \in S} p(s)}_{=1} = aE[X] + b.$$

□

Let us now apply linearity of expectation to some problems.

Example 3:
Suppose you roll 2 fair 6-sided dice. Let X_1 be the outcome of the first die and X_2 the outcome of the second. Show that the expectation of the sum of both dice is 7.

Solution:
For a single die, there are 6 outcomes, each with probability $\frac{1}{6}$ (since it's fair). The random variable X_1 is the value of the first die. The expectation of X_1 is $E[X_1] = \sum_{s \in S} p(s)X(s) = \sum_{i=1}^{6} \frac{1}{6} \cdot i = \frac{7}{2}$. The expectation of X_2 is the same: $E[X_2] = \frac{7}{2}$. The expected value of the sum of X_1 and X_2 is then $E[X_1 + X_2] = \frac{7}{2} + \frac{7}{2} = 7$.

A general strategy

1. Identify your sample space, outcomes, events and random variables X_i.
2. Express random variable X as a weighted sum of indicator random variables.
3. Apply linearity of expectation.

Example 4:
What is the average number of heads in 4 coin flips?

Solution:
Let's use the general strategy suggested on the right.

1. The sample space consists of all possible combinations obtained from 4 coin flips (there are 2^4). We can create an indicator random variable for each coin flip: X_1, X_2, X_3, X_4, which is 1 if the coin is heads and 0 if it is tails.

2. When a coin comes up heads, it increases the *total* number of heads, $X(s)$, by 1, so the total number of heads $X(s)$ can be expressed as the weighted sum of all the indicator random variables: $X(s) = 1 \cdot X_1 + 1 \cdot X_2 + 1 \cdot X_3 + 1 \cdot X_4$.

3. For a fair coin (probability of the event of getting heads is 50%), the expectation of each coin flip is $E[X_i] = \frac{1}{2}$, $i = 1, 2, 3, 4$. By linearity of expectation, the expectation of X is $E[X_1] + E[X_2] + E[X_3] + E[X_4] = 4 \cdot \frac{1}{2} = 2$.

Example 5:
Consider a website with k servers that needs to manage n incoming requests. Suppose we randomly pick one of the k servers to process a request. Use linearity of expectation to show that for n requests ($n > k$), the expected load on any server is n/k.

Solution:

1. The sample space consists of all ways we can assign n requests to k servers. There are k options for request 1, k options for request 2, ... and k options for request n, giving k^n outcomes.

2. Define a random variable $X_i(s)$ to represent the number of requests handled by the ith server. Also define an indicator random variable $z_j(s)$ to represent whether a request j was assigned to processor i:

$$z_j(s) = \begin{cases} 1 & \text{if request } j \text{ assigned to server } i, \\ 0 & \text{otherwise.} \end{cases}$$

   We can then write $X_i(s)$ as: $X_i(s) = \sum_{j=1}^{n} z_j(s)$.

3. Now we can use linearity of expectation to determine the expected number of requests assigned to server i.

$$E[X_i(s)] = E\left[\sum_{j=1}^{n} z_j(s)\right] = \sum_{j=1}^{n} E[z_j(s)] = \sum_{j=1}^{n} z_j(s) \cdot p(j^{\text{th}} \text{ request assigned to server } i) = \sum_{j=1}^{n} \frac{1}{k} = \frac{n}{k}.$$

Although the expected value of a random variable is useful, it doesn't tell us *how much* that random variable deviates from the mean. We can characterize this deviation by measuring the *distance from the mean*, which is called the variance.

Definition 4. *Let X be a random variable on a sample space S. The* variance *of X, denoted by $V(X)$ is*

$$V(X) = \sum_{s \in S} (X(s) - E[X])^2 \, p(s). \tag{4}$$

In other words, the variance is the weighted *average of the square of the deviation of X.*

You may have also heard of the *standard deviation*, denoted by $\sigma(X)$ and defined as $\sigma(X) = \sqrt{V(X)}$. There are a few ways to compute the variance of a random variable:

1. Compute $E[X]$ and then plug it into Equation 4.

2. Compute $V(X) = E[X^2] - E[X]^2$. In other words, compute the expected value of X and X^2 and then subtract them.

3. Compute $E[(X-\mu)^2]$ where $\mu = E[X]$. In other words, compute the expected value of the deviation from the mean squared.

Methods (2) and (3) can be proved (using a direct proof) using the definition of variance, as well as linearity of expection. Let's do some examples.

Example 6:
What is the variance of the random variable X, where X is the number that comes up when a fair die is rolled?

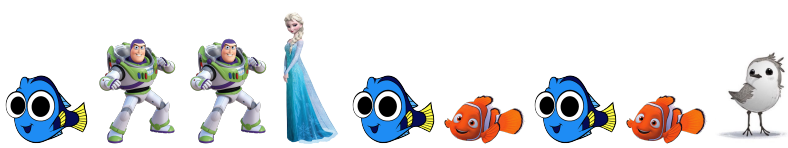
Solution:
We already computed the expected value as $E[X] = 7/2$ (see Example 2). Let's compute the expected value of the random variable X^2. Note that X^2 can have values of 1, 4, 9, 16, 25, 36. Therefore, the expected value of X^2 is

$$E[X^2] = \frac{1}{6}\sum_{i=1}^{n} i^2 = \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6}. \qquad (5)$$

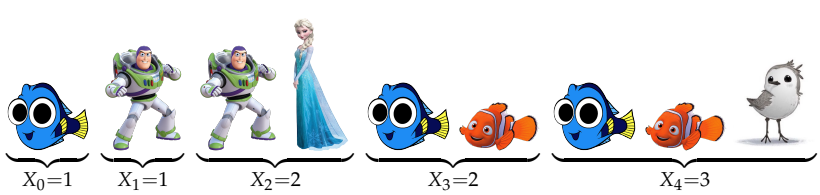
The variance is $V(X) = E[X^2] - E[X] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$.

3 Happy meals and toys

I don't know about you, but I used to wonder which toy I would get in my next Happy Meal. This is also known as the Coupon Collector's Problem, which we can analyze using linearity of expectation. Suppose there are five different toys we can get in a Happy Meal: Buzz, Dory, Nemo, Elsa and Piper. On average, how many Happy Meals do we need to buy to collect all five toys? Consider this possible sequence:



To solve this problem, we're going to break up our sequence into *segments* with the rule that a new segment begins if the toy we just got was different than the toys we already received. For the sequence above, we would then have five segments:



Let X_i be the length of segment i and note that, by our definition of the segments, we will have five segments. The total number of happy meals, M, we purchased in order to get all n (here, $n = 5$) toys is then

$$M = X_0 + X_1 + X_2 + X_3 + X_4.$$

It'll be easier to compute the average number of happy meals we need to buy (to get a *new* toy) within a single segment X_i and then apply linearity of expectation to compute the overall average number of happy meals we need to get all five toys.

At the beginning of segment i, we have i different toys (because of the rule we used to define segments). When we already have i toys, the probability of getting a *new* toy is $p_i = (n-i)/n$. To see this, note that we have already obtained i toys, so there are $n - i$ to pick from, out of a possible n, therefore, each of the $n - i$ toys has an equal probability of $(n-i)/n$.

Okay, well what is the average number of happy meals until we get this new toy, $E[X_i]$? Since the probability of getting a new toy is $p_i = (n-i)/n$, then X_i obeys what is called a *geometric distribution* (see the note on the right). The expected value for this kind of random variable is $E[X_i] = \frac{1}{p_i}$, so the expected length of segment i (X_i) is $n/(n-i)$. By linearity of expectation, the expected number of (total) meals until we get n toys is:

$$\begin{aligned}
E[M] &= E[X_0 + X_1 + X_2 + X_3 + X_4] \\
&= E[X_0] + E[X_1] + E[X_2] + E[X_3] + E[X_4] \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \frac{n}{n-2} + \frac{n}{n-3} + \frac{n}{n-4} \\
&= 5\left(\frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2} + 1\right)
\end{aligned}$$

More generally, $E[M] = n \sum_{i=1}^{n} \frac{1}{i}$. This means the expected number of happy meals is $E[M] \approx 11.42$, so you should go to McDonald's about 12 times to make sure you get all five toys.

As you continue in computer science, you will encounter a wide variety of applications of probability theory. In CS201, you might think about how to apply probability theory to determine the probability of a collision in a *hashing function*, which is a function that maps your keys to a particular storage location (used in dictionaries). You might also use probability theory to analyze the *average* complexity of searching (linear, binary) or sorting (insertion, merge-sort) algorithms.

Geometric distribution?

Don't worry, we're a little beyond the scope of the course at this point, but think back to a geometric series:

$$\sum_{i=0}^{n} r^i = \frac{1 - r^{n+1}}{1 - r}.$$

When $|r| < 1$ and $n \to \infty$, this becomes:

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1 - r}.$$

With a geometric distribution [1], like in our example, $r = 1 - p_i$, so the expected value for the length of segment i is

$$E[X_i] = \frac{1}{1 - (1 - p_i)} = \frac{1}{p_i} = \frac{n}{n - i}.$$

References

[1] E. Lehman, F. T. Leighton, and A. Meyer. *Mathematics for Computer Science*. 2018 (cit. on p. 7).