Random Variables

Suppose we flip a fair coin twice. What is the sample space Ω?

$$\Omega = \{HH, HT, TH, TT\}$$

In many cases, we don't really care whether we got $\{HT\}$ or $\{TH\}$ – sometimes all we care about is that we got exactly one head. To do this (in this case and many others), we define a **random variable** X as a numeric function of the outcome. That is, we look at an outcome in the sample space, and we only care about a single representative number for each one. Let's make this clearer by continuing our example.

Let X be the number of heads in 2 coin flips, $X: \Omega \to \Omega_X$. We only bother including values that X can actually take on in Ω_X, which we will call the **support** of a random variable. In this case, $X: \Omega \to \{0,1,2\}$. Remember a random variable is a <u>function</u> of the outcomes – that means every outcome ω in Ω will be mapped to a number.

$$X(\{HH\}) = 2$$
$$X(\{HT\}) = 1$$
$$X(\{TH\}) = 1$$
$$X(\{TT\}) = 0$$

To describe the probability of each of X's possible values, we use a **probability mass function (pmf)**. This simply describes the probability of each value the random variable can take on. From the example, it's clear that

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

This is an acceptable form for a probability mass function if asked for one. However, sometimes this notation can get a bit tedious, so we define a shorthand for $P(X = k)$ as $p_X(k)$ or simply $p(k)$. Using this notation, we can rewrite the above as

$$p_X(k) = \begin{cases} 1/4, & k = 0 \\ 1/2, & k = 1 \\ 1/4, & k = 2 \\ 0, & otherwise \end{cases}$$

k is a dummy variable (any value that X can take on)– we could have easily used x or any other letter instead. For example,

$$p_X(x) = \begin{cases} 1/4, & x = 0 \\ 1/2, & x = 1 \\ 1/4, & x = 2 \\ 0, & otherwise \end{cases}$$

We can also omit the subscript X if it is clear that the pmf we are talking about is for X. For example,

$$p(x) = \begin{cases} 1/4, & x = 0 \\ 1/2, & x = 1 \\ 1/4, & x = 2 \\ 0, & otherwise \end{cases}$$

Another way to define the support is all the values x that X takes on such that $p_X(x) > 0$ (in other words, values that actually have nonzero probability). The support of X in this case is $\Omega_X = \{0,1,2\}$.

Sometimes, we are in situations where there are multiple random variables in a question. Say X and Y are random variables, and we write $p(2)$. What does this mean? $P(X = 2)$ or $P(Y = 2)$? To be clear, we instead write $p_X(2) = P(X = 2)$ or $p_Y(2) = P(Y = 2)$. It is highly recommended (but not required yet) that you always subscript your probability mass functions. You will have to write subscripts later in this course, so you may as well get used to it now.

To summarize, at the beginning we had a full enumeration of the sample space

$$\Omega = \{HH, HT, TH, TT\}$$

with $P(\{HH\}) = P(\{HT\}) = P(\{TH\}) = P(\{TT\}) = \frac{1}{4}$.

Now, since we only care about how many heads we get, we can instead define the probabilities of the outcomes as

$$p_X(k) = \begin{cases} 1/4, & k = 0 \\ 1/2, & k = 1 \\ 1/4, & k = 2 \\ 0, & otherwise \end{cases}$$

The point of random variables is to summarize the outcomes of the sample space in a way that is useful. Note that $\sum_k p_X(k) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$, as it should.

Functions of Random Variables

Consider this contrived example: suppose instead of X, we were interested in X^3. X^3 is also a random variable since X was a random variable. What does this mean? It literally translates to the cube of X, or the cubed number of heads.

$$X^3(\{HH\}) = 8$$
$$X^3(\{HT\}) = 1$$
$$X^3(\{TH\}) = 1$$
$$X^3(\{TT\}) = 0$$

So we associate $\{HH\}$ with $2^3 = 8$ since it has 2 heads. Notice the other values don't change since $1^3 = 1$ and $0^3 = 0$. Now how do we describe the probability mass function for X^3?

$$p_{X^3}(k) = \begin{cases} 1/4, & k = 0 \\ 1/2, & k = 1 \\ 1/4, & k = 8 \\ 0, & otherwise \end{cases}$$

Notice we don't touch the probabilities at all – just the values. Notice that the probabilities do not change so they still sum to 1. Now, the X^3 cannot be 2 since there's no way a cubed number of heads can be 2. Instead, we reassign the probability to the value $2^3 = 8$. Notice the subscript is especially important in this case as we can distinguish between p_X and p_{X^3}. The support is also different - $\Omega_{X^3} = \{0,1,8\} \neq \{0,1,2\} = \Omega_X$.

What if two or more values ended up the same? For example, suppose Y represents the location of a frog starting at the origin who moves one unit either left or right, or not at all. Suppose he has probability of $1/6$ of going left or right, and $2/3$ of being lazy and not moving. Then Y has the following pmf:

$$p_Y(k) = \begin{cases} 1/6, & k = -1 \\ 2/3, & k = 0 \\ 1/6, & k = 1 \\ 0, & otherwise \end{cases}$$

Suppose we are interested only in how far the frog ends up, rather than the exact location. That is, we are interested in $W = |Y|$, a new random variable.

Then, a trick to find the pmf for W is to just directly apply the absolute value function to the values of the pmf. For example:

$$p_W(k) = \begin{cases} 1/6, & k = |-1| \\ 2/3, & k = |0| \\ 1/6, & k = |1| \\ 0, & otherwise \end{cases}$$

Notice that, now $|-1| = |1| = 1$. In this case, we simply add the probabilities since they were disjoint to get the final pmf:

$$p_W(k) = \begin{cases} 2/3, & k = 0 \\ 1/3, & k = 1 \\ 0, & otherwise \end{cases}$$

This should make sense – think about it. The frog moved one unit away with probability $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$, so $P(W = 1) = p_W(1) = 1/3$.

Other examples of random variables:
- The number of heads in n independent flips of a fair coin
- The number of people who are born every minute
- The number of flips of a fair coin until I flip a head
- The amount of time (in minutes) I wait until the next bus
- The weight (in pounds) of your backpack

Do you notice something different about the last two examples? They are still random variables, but they can take on an uncountable number of values. I can wait 4.833312 minutes for the next bus, or π minutes, or any arbitrary real number. Your backpack could weight 15.3235353 pounds, or e^5 pounds, or any real number. The first two take on finite values, and the third can take on a countably infinite number of values (the size of the natural numbers). We call a random variable **discrete** if it takes on a countable number of values. We call a random variable **continuous** if it takes on an uncountable number of values.

The reason we usually write $p_X(k)$ instead of $p_X(x)$ is because k is <u>usually</u> interpreted as an integer, whereas we think of x as being any real number. However, either is acceptable.

Expectation

Suppose we play the following game: you roll a 10 sided die and I pay you based on what you roll. If you roll a 1 or 3, I will pay you $\$9$. If you roll an even number, I will pay you $\$4$. If you roll anything else (5,7,9), I will pay you $\$6$. Let X be the (discrete or continuous?) random variable denoting how much you win from one play of a game. How much do you <u>expect</u> to win in a game?

Note that the probability mass function of X is

$$p_X(k) = \begin{cases} 2/10, & k = 9 \\ 5/10, & k = 4 \\ 3/10, & k = 6 \\ 0, & otherwise \end{cases}$$

And note that $\sum_k p_X(k) = \frac{2}{10} + \frac{5}{10} + \frac{3}{10} = 1$.

Maybe it's too hard to think of how much you expect to win in one game. Since there are 10 outcomes, let's think of how much you expect to win in 10 games, and then divide that by 10 (why does this make sense?).

Let's play 10 times. Then you expect to win $\$9$ exactly twice (when you roll a 1 or 3), $\$4$ 5 times (when you roll an even), and $\$6$ 3 times (when you roll 3,5,7,9). Then, your expected winning over 10 games is $9 * 2 + 4 * 5 + 6 * 3 = 56$. Therefore, your expected winning in 1 game is exactly $\$5.60$.

If we had divided by 10 earlier, we could write

$$\frac{1}{10}(9 * 2 + 4 * 5 + 6 * 3) = 9 * \frac{2}{10} + 4 * \frac{5}{10} + 6 * \frac{3}{10} = 5.6$$

This is exactly $\sum_k k p_X(k)$. Notice how we multiplied each possible value of the random variable by its probability. This is the definition of **expectation** – it is a weighted sum of the outcomes of the random variable, weighted by their probabilities. We define $E[X] = \sum_k k p_X(k)$.

Let X be how much money I win in my first play of this game, and Y be how much I win in my second play. Then if $Z = X + Y$, Z would represent the amount of money won in two games. Since X and Y have the same pmf (why?), we expect to win $\$5.60$ in each game – that is, $E[X] = E[Y] = 5.6$. Note that we would <u>not</u> say $X = Y$, since they could take on different values, even though they have the same pmf (and therefore expectation).

Something interesting to think about – let $Z = X + Y$ be the amount of money I win in two plays of this game. Would it make sense that I expect to win $2 * \$5.60 = \$11.20$? That is,

$$E[Z] = E[X + Y] = E[X] + E[Y] = 5.6 + 5.6 = 11.2$$

Yes! This property is called **linearity of expectation** and will be covered soon.