Chapter 3. Discrete Random Variables

3.1: Discrete Random Variables Basics

Slides (Google Drive) Alex Tsun Video (YouTube)

3.1.1 Introduction to Discrete Random Variables

Suppose you flip a fair coin twice. Then the sample space is:

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

Sometimes, though, we don't care about the order (HT vs TH), but just the fact that we got one heads and one tail. So we can define a **random variable** as a numeric function of the outcome.

For example, we can define X to be the number of heads in the two independent flips of a fair coin. Then X is a function, $X : \Omega \to \mathbb{R}$ which takes outcomes $\omega \in \Omega$ and maps them to a number. For example, for the outcome HH, we have $X(HH) = 2$ since there are two heads. See the rest below!

$$\begin{aligned} X(HH) &= 2 \\ X(HT) &= 1 \\ X(TH) &= 1 \\ X(TT) &= 0 \end{aligned}$$

X is an example of a random variable, which brings us to the following definition:

Definition 3.1.1: Random Variable

Suppose we conduct an experiment with sample space Ω. A **random variable (rv)** is a numeric function of the outcome, $X : \Omega \to \mathbb{R}$. That is it maps outcomes $\omega \in \Omega$ to numbers, $\omega \to X(\omega)$.
The set of possible values X can take on is its **range/support**, denoted Ω_X.
If Ω_X is finite or countable infinite (typically integers or a subset), X is a **discrete random variable (drv)**. Else if Ω_X is uncountably large (the size of real numbers), X is a **continuous random variable**.

Example(s)

Below are some descriptions of random variables. Find their ranges and classify them as a discrete random variable (DRV) or continuous random variable (CRV). The first row is filled out for you as an example!

RV Description	Range	DRV or CRV?
X, the # of heads in n flips of a fair coin	$\{0, 1, \ldots, n\}$	DRV
N, the # of people born this year.	TODO	TODO
F, the # of flips of a fair coin up to and including my first head.	TODO	TODO
B, the amount of time I wait for the next bus in seconds.	TODO	TODO
C, the temperature in Celsius of liquid water	TODO	TODO

Solution Here is the solution in a table, with explanations below.

RV Description	Range	DRV or CRV?
X, the # of heads in n flips of a fair coin	$\{0, 1, \ldots, n\}$	DRV
N, the # of people born this year.	$\{0, 1, 2, \ldots\}$	DRV
F, the # of flips of a fair coin up to and including my first head.	$\{1, 2, \ldots, \}$	DRV
B, the amount of time I wait for the next bus in seconds.	$[0, \infty)$	CRV
C, the temperature in Celsius of liquid water	$(0, 100)$	CRV

- The range of X is $\Omega_X = \{0, 1, \ldots, n\}$ because there could be any where from 0 to n heads flipped. It is a discrete random variable because there are finite $n + 1$ values that it takes on.
- The range of N is $\Omega_N = \{0, 1, 2 \ldots\}$ because there is no upper bound on the number of people that can be born. This is countably infinite as it is a subset of all the integers, so it is a discrete random variable.
- The range of F is $\Omega_F = \{1, 2, \ldots\}$ because it will take at least 1 flip to flip a head or it could always be tails and never flip a head (although the chance is low). This is still countable as a subset of all the integers, so it is a discrete random variable.
- The range of B is $\Omega_B = [0, \infty)$, as there could be partial seconds waited, and it could be anywhere from 0 seconds to a bus never coming. This is a continuous random variable because there are uncountably many values in this range.
- The range of C is $\Omega_C = (0, 100)$ because the temperature can be any real number in this range. It cannot be 0 or below because that would be frozen (ice), nor can it be 100 or above because this would be boiling (steam). This is a continuous random variable.

□

3.1.2 Probability Mass Functions

Let's return to X which we defined to be the number of heads in the flip of two fair coins. We already determined that $\Omega = \Omega = \{HH, HT, TH, TT\}$ and $X(HH) = 2, X(HT) = 1, X(TH) = 1$ and $X(TT) = 0$. The range, Ω_X, is $\{0, 1, 2\}$.

We can define the **probability mass function (pmf)** of X, as $p_X : \Omega_X \to [0, 1]$:

$$p_X(k) = \mathbb{P}(X = k)$$

to calculate the probabilities that X takes on each of these values.

In this case we have the following:

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

this is because the number of outcomes for $X = 0$ is 1 of the 4, the number of outcomes for $X = 1$ is 2 of the 4, and the number of outcomes for $X = 2$ is 1 of the 4.

This brings us to the formal definition of a probability mass function:

Definition 3.1.2: Probability Mass Function (pmf)

The **probability mass function (pmf)** of a discrete random variable X assigns probabilities to the possible values of the random variable. That is $p_X : \Omega_X \to [0, 1]$ where:

$$p_X(k) = \mathbb{P}(X = k)$$

Note that $\{X = a\}$ for $a \in \Omega$ form a partition of Ω, since each outcome $a \in \Omega$ is mapped to exactly one number. Hence,

$$\sum_{z \in \Omega_X} p_X(z) = 1$$

Notice here the only thing consistent is p_X, as it's the PMF of X. The value inside is a dummy variable - just like we can write $f(x) = x^2$ or $f(t) = t^2$. To reinforce this, I will constantly use different letters for dummy variables.

3.1.3 Expectation

We have this idea of a random variable, which is actually neither random nor a variable (it's a deterministic function $X : \Omega \to \Omega_X$.) However, the way I like to think about it is: it a random quantity which we do not know the value of yet. You might want to know what you might expect it to equal on average. For example, X could be the random variable which represents the number of babies born in Seattle per day. On average, X might be equal to 250, and we would write that its average/mean/expectation/expected value is $\mathbb{E}[X] = 250$.

Let's go back to the coin example though to define expectation. Your intuition might tell you that the expected number of heads in 2 flips of a fair coin would be 1 (you would be correct).

Since X was the random variable defined to be the number of heads in 2 flips of a fair coin, we denote this $\mathbb{E}[X]$. Think of this as the average value of X.

More specifically, imagine if we repeated the two coin flip experiment 4 times. Then we would "expect" to get HH, HT, TH, and TT each once. Then, we can divide by the number of times (4) to get 1.

$$\frac{2 + 1 + 1 + 0}{4} = 2 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} = 1$$

Notice that:

$$\begin{aligned} 2 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} &= X(HH)\mathbb{P}(HH) + X(HT)\mathbb{P}(HT) + X(TH)\mathbb{P}(TH) + X(TT)\mathbb{P}(TT) \\ &= \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega) \end{aligned}$$

This is the the sum of the random variable's value for each outcome multiplied by the probability of that outcome (a weighted average).

Another way of writing this is by multiplying every value that X takes on (in its range) with the probability of that value occurring (the pmf). Notice that below is the same exact sum, but groups the common values together (since $X(HT) = X(TH) = 1$). That is:

$$2 \cdot \frac{1}{4} + 1 \cdot \left(\frac{1}{4} + \frac{1}{4}\right) + 0 \cdot \frac{1}{4} = 2 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 0 \cdot \frac{1}{4} = \sum_{k \in \Omega_X} k \cdot p_X(k)$$

This brings us to the definition of expectation.

Definition 3.1.3: Expectation

The **expectation/expected value/average** of a discrete random variable X is:

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

or equivalently,

$$\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot p_X(k)$$

The interpretation is that we take an average of the possible values, but weighted by their probabilities.

3.1.4 Exercises

1. Let X be the value of single roll of a fair six-sided dice. What is the range Ω_X, the PMF $p_X(k)$, and the expectation $\mathbb{E}[X]$?

   Solution: The range is $\Omega_X = \{1, 2, 3, 4, 5, 6\}$. The pmf is

   $$p_X(k) = \frac{1}{6}, k \in \Omega_X$$

   The expectation is

   $$\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot p_X(k) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{1}{6}(1 + 2 + \cdots + 6) = 3.5$$

   This kind of makes sense right? You expect the "middle number" between 1 and 6, which is 3.5.

2. Suppose at time $t = 0$, a frog starts on a 1-dimensional number line at the origin 0. At each step, the frog moves *independently*: left with probability 1/10, and right (with probability 9/10). Let X be the

position of the frog after 2 time steps. What is the range Ω_X, the PMF $p_X(k)$, and the expectation $\mathbb{E}[X]$?

Solution: The range is $\Omega_X = \{-2, 0, 2\}$. To find the pmf, we find the probabilities of being each of those three values.

(a) For X to equal -2, we have to move left both times, which happens with probability $\frac{1}{10} \cdot \frac{1}{10}$ by independence of the moves.

(b) For X to equal 2, we have to move right both times, which happens with probability $\frac{9}{10} \cdot \frac{9}{10}$ by independence of the moves.

(c) Finally, for X to equal 0, we have to take opposite moves. So either LR or RL, which happens with probability $2 \cdot \frac{1}{10} \cdot \frac{9}{10} = \frac{18}{100}$. Alternatively, the easier way is to note that these three values sum to 1, so $\mathbb{P}(X=0) = 1 - \mathbb{P}(X=2) - \mathbb{P}(X=-2) = 1 - \frac{81}{100} - \frac{1}{100} = \frac{18}{100}$

So our PMF is:

$$p_X(k) = \begin{cases} 1/100 & k = -2 \\ 18/100 & k = 0 \\ 81/100 & k = 2 \end{cases}$$

The expectation is

$$\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot p_X(k) = -2 \cdot \frac{1}{100} + 0 \cdot \frac{18}{100} + 2 \cdot \frac{81}{100} = 1.6$$

You **might** have been able to guess this, but how? At each time step you "expect" to move to the right by $\frac{9}{10} - \frac{1}{10}$ which is 0.8. So after two steps, you would expect to be at 1.6. We'll formalize this approach more in the next chapter!

3. Let X be the number of independent coin flips up to and including our first head, where $\mathbb{P}(head) = p$. What is the range Ω_X, the PMF $p_X(k)$, and the expectation $\mathbb{E}[X]$?

Solution: The range is $\Omega_X = \{1, 2, 3, ...\}$, since it could theoretically take any number of flips. The pmf is

$$p_X(k) = (1-p)^{k-1}p, k \in \Omega_X$$

Why? We can start slowly.

(a) $\mathbb{P}(X = 1)$ is the probability we get heads (for the first time) on our first try, which is just p.

(b) $\mathbb{P}(X = 2)$ is the probability we get heads (for the first time) on our second try, which is $(1-p)p$ since we had to get a tails first.

(c) $\mathbb{P}(X = k)$ is the probability we get heads (for the first time) on our k^{th} try, which is $(1-p)^{k-1}p$, since we had to get all tails on the first $k-1$ tries (otherwise, our first head would have been earlier).

The expectation is pretty complicated and uses a calculus trick, so don't worry about it too much. Just understand the first two lines, which are the setup! But before that, what do you think it should be? For example, if $p = 1/10$, how many flips do you think it would take until our first head? Possibly 10? And if $p = 1/7$, maybe 7? So seems like our guess will be $\mathbb{E}[X] = \frac{1}{p}$. It turns out this intuition is actually correct!

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k \in \Omega_X} k \cdot p_X(k) && \text{[def of expectation]} \\
&= \sum_{k=1}^{\infty} k(1-p)^{k-1} p \\
&= p \sum_{k=1}^{\infty} k(1-p)^{k-1} && [p \text{ is a constant with respect to } k] \\
&= p \sum_{k=1}^{\infty} \frac{d}{dp}(-(1-p)^k) && \left[\frac{d}{dy} y^k = k y^{k-1}\right] \\
&= -p \left(\frac{d}{dp} \sum_{k=1}^{\infty} (1-p)^{k-1} \right) && \text{[swap sum and integral]} \\
&= -p \left(\frac{d}{dp} \frac{1}{1-(1-p)} \right) && \left[\text{geometric series formula: } \sum_{i=0}^{\infty} r^i = \frac{1}{1-r}\right] \\
&= -p \left(\frac{d}{dp} \frac{1}{p} \right) \\
&= -p \left(-\frac{1}{p^2} \right) \\
&= \frac{1}{p}
\end{aligned}$$