3 Random variables and probability distributions - I

3.1 Random variables

A random variable is a mapping of a sample space to the number line. This means that each outcome in the sample space will be associated with a point on the number line, i.e. with a real number. In some cases this mapping is straight forward. For example, we can think of the sample space as the number of economics students turning up for a tutorial. The sample space will contain outcomes ranging from 0 to 24. We can easily associate each outcome with a number on the real line that has the same value as the outcome. So, the random variable X in this case can take values 0, 1, 2, .., 24, with the probability of its taking each of those values being the same as the probability of the respective outcome.

Now let us think of the example of a coin toss. The sample space is $\{H, T\}$. There is no natural way of mapping these outcomes to a real number. We can arbitrarily map H to 1 and T to 0. In this case, the random variable X can take values 0 and 1, with probability of X being 1, i.e. $P(X = 1)$ being equal to the probability of getting H. We can think of this random variable as representing a coin toss resulting in a head. We will use this kind of mapping for most binary outcomes. Eg. whether or not it rains tomorrow, can be mapped as $X = 1$ representing rain and $X = 0$ representing no rain.

The same sample space can be mapped as different random variables depending on what we are looking for. For example, the sample space for the results of two coin tosses is $\{HH, HT, TH, TT\}$. We can think of one random variable X representing the number of heads turning up. X can take the values 0, 1 and 2, with $P(X = 0) = P(TT)$, $P(X = 1) = P(HT \text{ or } TH)$, $P(X = 2) = P(HH)$. We can think of another random variable Y representing both coins

turning up different sides. Hence, $P(Y = 0) = P(HH \text{ or } TT)$, and $P(Y = 1) = P(HT \text{ or } TH)$.

There are two types of random variables: discrete and continuous. Random variables that can only take a finite, or countably infinite, number of values and are hence represented by points on the number line are called discrete. All the examples that we have seen so far are discrete. Random variables that can take infinite number of values and are represented by continuous sections of the number line are called continuous random variables. An example would be the length of an economics lecture, or the average height of students in class.

3.2 Probability distribution of a discrete random variable

The probability associated with different values of the random variable is called the probability distribution and as we saw earlier, depends of the probability of the outcomes in the sample space. For a discrete random variable, this distribution can be expressed in the form of a function that associates each value of the random variable to its respective probability. This is called the **probability mass function** or *pmf*. For the example of a coin toss, the *pmf* $f_X(x)$ is given by

$$\begin{aligned} f_X(1) &= P(X = 1) = 0.5, \\ f_X(0) &= P(X = 0) = 0.5 \\ f_X(x) &= 0, \text{ for all } x \neq 0, 1 \end{aligned}$$

In general, we can write the *pmf* as

$$f_X(x) = P(X = x)$$

Note that this is a function *of* x, i.e. the various values that X can take, and it gives out the probability of X taking the value. Therefore, if X takes the values x_i, $i = 1, 2, ..., n$, then $\sum_{i=1}^{n} f_X(x_i) = 1$.

Exercise: Let the random variable X represent the number of heads that appear when three coins are tossed together. Write down the *pmf* and draw its graph.

Since all values that the random variable takes are on the number line, hence they are ordered. Therefore, we can define another way of representing the probability distribution, called the **cumulative distribution function** or *cdf*. The *cdf* gives the probability of the value of the random variable being less than a particular number. Hence for a coin toss, the *cdf* $F_X(x)$ will be

$$F_X(x) = \begin{cases} 0 & for\ x < 0, \\ 0.5 & for\ 0 \leq x < 1, \\ 1 & for\ x \geq 1 \end{cases}$$

In general, the *cdf* of a random variable X can be written as

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$$

Exercise: Let the random variable X represent the number of heads that appear when three coins are tossed together. Write down the *cdf* and draw its graph.

We can see that by definition, a *cdf* will satisfy some properties. As x approach $-\infty$, the *cdf* will approach zero as the probability of the random variable taking a value below an extremely low value of x will become zero. Similarly, As x approach ∞, the *cdf* will approach 1 as the probability of the random variable taking a value below an extremely high value of x will become 1. Also, as x increases, the *cdf* will either increase or stay the same, it will never decrease. This is because we are adding up probabilities, which can be zero or positive, but never negative. We can write down these properties in notation.

- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to \infty} F_X(x) = 1$
- For any $y > x$, $F_X(y) \geq F_X(x)$

We will see that these properties will also be valid for continuous random variables.

3.3 Preview of estimation

Before we move ahead, let us see a preview of how we are going to use these random variables. One of the most commonly used probability distributions of binary random variables, i.e. random variables that take values 0 and 1, is called the Bernoulli distribution. It is used to model a process with two outcomes where the probabilities of getting these outcomes are fixed. Let the probability of getting $X = 1$ be p, then the *pmf* of a random variable following the Bernoulli distribution is given by.

$$f_X(x) = \begin{cases} p & for\ x = 1, \\ 1 - p & for\ x = 0, \\ 0 & for\ x \neq 0, 1 \end{cases}$$

It is sometimes written in a more compact form as follows.

$$f_X(x) = \begin{cases} p^x(1-p)^{1-x} & for\ x = 0, 1 \\ 0 & for\ x \neq 0, 1 \end{cases}$$

Now suppose we know that a process follows the Bernoulli distribution, but we don't know the probability p. In such a case, p is known as an *unknown parameter* of the distribution. Think of a coin that is not fair. We know that tossing the coin has two outcomes and the probability of getting heads is fixed, but we don't know what that probability is. The idea is to use data to find this unknown parameter p. To do this we will use something called the **expected value** of a random variable. The expected value of a random variable is the mean of that random variable. To understand this let us consider the sample mean. Think of a dataset where the observations for a variable y are

Obs.	y
1	0
2	0
3	0
4	1
5	1
6	2

The sample mean for y is given by

$$\begin{aligned}\bar{y} &= \frac{0+0+0+1+1+2}{6} \\ &= \left(\frac{3}{6}\right)0 + \left(\frac{2}{6}\right)1 + \left(\frac{1}{6}\right)2\end{aligned}$$

We multiply the fraction of times a value occurs to the value itself, and then add all such products up. Similarly, in the case of a discrete random variable, we multiply each possible value of the random variable with the probability of that value occurring, and then add all such products up. The expected values of a discrete random variable X is given by

$$E[X] = \sum_{i=1}^{n} x_i f_X(x_i)$$

For a random variable following a Bernoulli distribution with a probability parameter p, the expected value will be

$$E[X] = p.1 + (1-p)0 = p$$

Now to find the value of the parameters we will use something called the Central Limit Theorem (CLT), which we will discuss in detail later, that basically says that if we take a large enough sample from a distribution, then the sample mean will be very close to the expected value of the distribution. Think of the population as infinite number of repeated trials of the underlying process that is defined by the distribution. Think of the samples as a collection of some of the

realisations of such trials that we observe. In the case of the Bernoulli distribution, the outcomes are signified by two values of the random variable, 0 and 1. Hence the sample would be a collection of 0s and 1s. Think of the biased coin again. If we toss it a number of times and record heads as 1 and tails as 0, then this would constitute a sample taken from the Bernoulli distribution. The CLT tells us that for large enough sample, the sample mean, which in this case would be the fraction of heads turning up, would be very close to the expected value, which we showed was the parameter p. This is a simple illustration of one of the methods that we use data to estimate the unknown parameters of a distribution.

Question. Do you see that expected value is the same as population mean?

3.4 Probability distribution of continuous random variables

A continuous random variable can take an infinite number of values. Hence, we cannot assign a probability to it taking a particular value. Think of the random variable representing the fraction of income spent on consumption. This variable could take any value from 0 to 1. If all values are equally likely, then what would be the probability of this fraction being 0.5? It would be zero.

What we can define is that probability of the random variable taking a value between two point. In the example above, if all values are equally likely, then the probability that the fraction is between 0 and 0.5 is 0.5. Hence, for continuous random variables we define a function such that the area under the graph of the function between any two point on the number line, gives the probability of the random variable taking a value between those two points. This is called the **probability density function** or *pdf*. Formally we can express the *pdf* $f_X(x)$ as follows.

$$\int_a^b f_X(x)dx = P(a \leq x \leq b)$$

Although the convention is to express *pmf* and *pdf* with similar notation, remember that they are not exactly analogous. While the value of a *pmf* at any point is an actual probability, a *pdf* only gives probability density, that has to be integrated to obtain the probability. This implies that while *pmf* will always be less than or equal to 1, as it is expressing actual probability, a *pdf* can take any non-negative value.

Since total probability is 1, therefore for any *pdf* $f_X(x)$,

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

The *cdf* of a continuous random variable is given as follows.

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)dy$$

Note that $\frac{dF_X(x)}{dx} = f_X(x)$.