2.2 Expectation and Further Properties

In the previous section we looked at the distribution of discrete random variables. In this section we look at further properties of these distributions. Our first notion is that of expectation, the (weighted) mean value that the random variable takes. For example if we roll a fair six sided die a large number of times, and take an average of all the scores we would expect this value to be close to 3.5, the average of all the faces on the die. For a random variable X which is equal to outcome of the dice roll, the expected value of X is indeed 3.5. Thus the expectation is the average we would expect to see. We define the expectation as follows:

Definition 2.2.1. Suppose X is a discrete random variable, taking values in $\mathbb{N}_0$, then we define the expectation of X as:

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i \cdot f_X(i) = 0 \times \mathbb{P}[X=0] + 1 \times \mathbb{P}[X=1] + 2 \times \mathbb{P}[X=2] + 3 \times \mathbb{P}[X=3] + \dots.$$

Essentially the expectation describes a weighted average across all the possible outputs the random variable has. The more likely an output is, the more the expectation is skewed towards that output. It is worth noting that the expectation may not necessarily be a whole number, in general the expectation is usually not an output of the random variable. As we will see in the following example:

Example 2.2.1. Consider the random variable Z with distribution function:

$$f_Z(i) = \begin{cases} \frac{1}{4} & \text{if } i = 2; \\ \frac{1}{9} & \text{if } i = 3; \\ \frac{23}{36} & \text{if } i = 5; \\ 0 & \text{otherwise.} \end{cases}$$

What is the expectation of Z? From the definition we know that:

$$\mathbb{E}[Z] = \sum_{i=0}^{\infty} i \cdot f_Z(i).$$

As previously discussed if $i \notin \{2,3,5\}$ then we have that $f_Z(i) = 0$. Therefore by substituting in the appropriate values we have:

$$\begin{aligned} \mathbb{E}[Z] &= 2 \cdot f_Z(2) + 3 \cdot f_Z(3) + 5 \cdot f_Z(5) \\ &= 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{9} + 5 \cdot \frac{23}{36} \approx 4.799. \end{aligned}$$

From the above example we can see from the distribution function that Z is equal to 5 more than half the time on the average. This explains why the expectation is much closer to five, than any other value. The expectation is known as a measure of central tendency, it is one way of indicating where most of the probability distribution occurs.

A further property we need to consider, is how expectation behaves when taking a function of a random variable. In this course we will only be interested in finding $\mathbb{E}[X^2]$, referred to as the second moment of X. We define it as follows:

Definition 2.2.2. Let X be a discrete random variable, with support in $\mathbb{N}_0$, then we have that:

$$\mathbb{E}[X^2] = \sum_{i=0}^{\infty} i^2 \cdot f_X(i) = 1^2 \times \mathbb{P}[X = 1] + 2^2 \times \mathbb{P}[X = 2] + 3^2 \times \mathbb{P}[X = 3] + \dots$$

Example 2.2.2. Suppose we have a random X with the following distribution given by the table below:

x	1	2	3	4	5
$f_X(x)$	0.2	0.3	0.15	0	0.35

What is the value of $\mathbb{E}[X]$, and what is $\mathbb{E}[X^2]$?

For finding the expectation of X we appeal to the definition:

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i \cdot f_X(i) = 1 \times 0.1 + 2 \times 0.3 + 3 \times 0.15 + 5 \times 0.35 = 2.9.$$

For the second moment, $\mathbb{E}[X^2]$, we appeal to Definition 2.2.2. The calculation is similar to finding the expectation, instead we multiply by i^2. We complete the calculation below:

$$\mathbb{E}[X^2] = \sum_{i=0}^{\infty} i^2 \cdot f_X(i) = 1 \times 0.1 + 4 \times 0.3 + 9 \times 0.15 + 25 \times 0.35 = 11.4.$$

One remark on the previous example, it should be clear that in general, $\mathbb{E}[X^2]$ is usually not equal to $(\mathbb{E}[X])^2$.

2.3 Variance and Standard Deviation

In the previous section we looked at the expectation, which associates a form of a mean value to a random variable. Sometimes only knowing the mean value of the random variable is not sufficient information, sometimes we may need to know about how spread out the distribution is. Consider the following example:

Example 2.3.1. Suppose I make you two offers, say offer A and offer B. In offer A I will give you a guaranteed £100 pounds. While in the second offer we flip a fair coin, if the coin shows heads then I will give you £200, while if the coin shows tails you will receive nothing. Which offer should you choose?

We define two random variables A and B, each representing the amount you would earn from each of the offers. The distribution of A is straightforward to compute $\mathbb{P}(A = 100) = 1$, therefore it can be quickly checked that its expectation is also 100. For the random variable B we have that $\mathbb{P}(B = 0) = \mathbb{P}(B = 200) = 0.5$. Therefore we compute the expectation of B,

$$\mathbb{E}[B] = \sum_{i=0}^{\infty} i \cdot f_B(i) = 0 \cdot 0.5 + 200 \cdot 0.5 = 100.$$

So in expectation, both offers are worth the same amount; however something feels very different about betting on either receiving double or nothing, versus a guaranteed winning. While one offer is constant, the other feels subject to a form a variation, we quantify this behaviour with the notion of *variance*.

Definition 2.3.1. Suppose X is a random variable, we define the variance of X,

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

While this formula may look intimidating, we break it down term by term. Essentially the quantity $X - \mathbb{E}[X]$ tells us how far the random variable X is away from its mean value. We then take a square of this quantity, to ensure that all differences are positive. Now if we take $\mathbb{E}[(X - \mathbb{E}[X])^2]$, we are essentially looking at the mean squared difference that the random variable X has from its expectation. As the variance represents the sum of squared deviations, in order to contextulise our findings in terms of the original random variable, we need to take a square root. This in turn gives us the standard deviation of the random variable.

Definition 2.3.2. Suppose X is a random variable, we define the *standard deviation* of X as:

$$\sigma = \sqrt{\mathrm{Var}[X]}.$$

In essence the standard deviation tell us on average how far the random variable deviates from its expected value. When the standard deviation is high, this means that X is more likely to take values further away from its expectation. While if the standard deviation is low, then X typically takes values close to its expectation. In practice, to find the standard deviation we always need to find the variance first and then apply the definition. In practice Definition 2.3.1 is not the easiest for finding the variance, so we consider the following equivalent form:

Lemma 2.3.1. Suppose X is a random variable, then the variance can also be written as:

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

The proof of Lemma 2.3.1 follows from taking the definition of variance. Lemma 2.3.1 gives us a straight forward method for calculating the variance in terms of just the expectation, and the second moment.

Example 2.3.2. Consider Example 2.3.1, with random variables A and B as defined. We compute $\mathrm{Var}[A]$ and $\mathrm{Var}[B]$. We have already shown that $\mathbb{E}[A] = \mathbb{E}[B] = 100$, therefore we just need to compute the second moments. By directly applying the formula:

$$\mathbb{E}[A^2] = \sum_{i=0}^{\infty} i^2 \cdot f_A(i) = 100^2 \cdot 1 = 10000.$$

$$\mathbb{E}[B^2] = \sum_{i=0}^{\infty} i^2 \cdot f_B(i) = 0 \cdot 0.5 + 200^2 \cdot 0.5 = 20000.$$

Therefore by applying Lemma 2.3.1 we have that:

$$\mathrm{Var}[A] = 10000 - (100)^2 = 0.$$

While we also have that:

$$\mathrm{Var}[B] = 20000 - (100)^2 = 10000.$$

Similarly we also have that the standard deviation of $A = \sqrt{0} = 0$. While the standard deviation of $B = \sqrt{10000} = 100$.

Thus we have that there is a large variation in the random variable B which implies a more unpredictable outcome. Thus while both bets are worth the same on average, if you are feeling more lucky / risky you have more chance earning big on outcome B.