4.5 Variance

In this chapter, one of our goals is to characterise discrete random variables. To do this, we need to determine our criteria which will differentiate between different random variables. The first criteria is the expectation which we defined in the previous section. However, by itself, expectation is a quite weak criteria. It is quite easy to find two very different random variables whose expectation coincide.

To overcome this issue, we should introduce another notion, called Variance. But first, let's discuss the origins.

Suppose two random variables, say X and Y have the same expectation, that is, $\mathbb{E}(X) = \mathbb{E}(Y)$. What is next best expression that gives some information about our random variables? If we can find an expression to measure how spread the data about its expectation is, that would be a good starting point. Together with expectation, the measure of how spread the data is can give quite some information.

How can we measure how spread the data is? An immediate way is to check $X - \mu_x$ which is the distance between X and its mean μ_x. Note that this is still a random variable. So we can check the expectation of this difference.

$$\mathbb{E}(X - \mu_x) = \mathbb{E}(X) - \mu_X = \mu_X - \mu_X = 0.$$

Note that there is nothing specific about X in the above calculation. Hence one may similarly calculate $\mathbb{E}(Y - \mu_Y)$ which again will result in zero. But that's not what we are seeking. Our goal is to determine a criteria which shows difference between different random variables.

Why is that expectation equals zero? Basically, we were looking the "weighted average" of $X - \mu_X$. Naturally, some of the data is located on the left of the expectation μ_x and some are on the right of μ_X. Hence they cancel each other on average. So the problem arises due to those negative terms. How can we handle these terms? Why not do we take the absolute value? $|X - \mu_X|$ would be what we want. However, absolute values are not the best objects to deal with. So, can we replace absolute value with a better operation.

Recall the definition of the absolute value:

$$|x| := \sqrt{x^2}.$$

This leads to the idea that we can consider $\sqrt{(X - \mu_X)^2}$ instead of $|X - \mu_X|$. As you may guess, roots are not the easiest operators, either. So how can we simplify our expression without loosing the "difference form the mean".

For this purpose, we'll drop the square root for now and consider $(X - \mu_X)^2$ only. This is a relatively easy operation. We can take its expectation without any trouble. You may wonder how to fix the square root. To fix it, first we will take expectation of $(X - \mu_X)^2$ and call it the variance of X

$$Var(X) = \mathbb{E}((X - \mu_X)^2)$$

and then we'll take the root of this real number. That root will be called standard deviation.

Definition 4.5.1 If X is a random variable with mean (expectation) μ_X then the variance of X is defined by

$$Var(X) = \mathbb{E}((X - \mu_X)^2).$$

R There is a useful representation of variance. To see this, let's expand the square and use the linearity of the expectation.

$$\begin{aligned}
Var(X) &= \mathbb{E}((X-\mu_X)^2) \\
&= \mathbb{E}(X^2 - 2\mu_X X + \mu_X^2) \\
&= \sum_x (x^2 - 2\mu_X x + \mu_X^2) p(x) \\
&= \sum_x x^2 p(x) - 2\mu_X \sum_x x p(x) + \mu_X^2 \sum_x p(x) \\
&= \mathbb{E}(X^2) - 2\mu_X^2 + \mu_X^2 \\
&= \mathbb{E}(X^2) - \mu_X^2
\end{aligned}$$

Proposition 4.5.1

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Notation 4.2. *From now on, we'll write* $\mathbb{E}^n(X)$ *instead of* $(\mathbb{E}(X))^n$. *So* $\mathbb{E}^2(X)$ *means* $(\mathbb{E}(X))^2$ *and*

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

■ **Example 4.19** What is the expectation and variance of a flipped fair coin?

$$X = \{\text{the number of heads}\}$$

Hence the PMF is

$$p(x) = \begin{cases} 0.5 & \text{if } x = 0 \\ 0.5 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}.$$

The expectation

$$\mathbb{E}(X) = 0 \cdot p(0) + 1 \cdot p(1) = 0.5$$

and $\mathbb{E}(X^2)$

$$\mathbb{E}(X^2) = 0^2 \cdot p(0) + 1^2 \cdot p(1) = 0.5.$$

So the variance is

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = 0.5 - (0.5)^2 = 0.25.$$

■

■ **Example 4.20** Compute the expectation and variance of X that shows the number on a fair dice. PMF is

$$p(x) = \begin{cases} 1/6 & \text{if } x = 1,2,3,4,5,6 \\ 0 & \text{otherwise} \end{cases}.$$

The expectation

$$\mathbb{E}(X) = 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) + 4 \cdot p(4) + 5 \cdot p(5) + 6 \cdot p(6) = 7/2$$

and $\mathbb{E}(X^2)$

$$\mathbb{E}(X^2) = 1^2 \cdot p(1) + 2^2 \cdot p(2) + 3^2 \cdot p(3) + 4^2 \cdot p(4) + 5^2 \cdot p(5) + 6^2 \cdot p(6) = 91/6.$$

So the variance is

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = 91/6 - (7/2)^2 = 35/12.$$

■

Theorem 4.5.2 If a and b are two real numbers then

$$Var(aX+b) = a^2 Var(X).$$

Proof. By definition,

$$\begin{aligned}
Var(aX+b) &= \mathbb{E}\left[(aX+b-\mathbb{E}(aX+b))^2\right] \\
&= \mathbb{E}\left[(aX+b-a\mathbb{E}(X)-b)^2\right] \\
&= \mathbb{E}\left[a^2\left(X-\mathbb{E}(X)\right)^2\right] \\
&= a^2\mathbb{E}\left[(X-\mathbb{E}(X))^2\right] \\
&= a^2 Var(X).
\end{aligned}$$

■

Definition 4.5.2 For a random variable X the standard deviation of X, σ_X or $SD(X)$, is defined by

$$\sigma_X = \sqrt{Var(X)}.$$

R Note that variance is always non negative. It is easy to see once we consider the definition $\mathbb{E}\left[(X-\mathbb{E}(X))^2\right]$. Hence standard deviation is well-defined, and it is non-negative as well.