Lecture 9: Expectation of functions of X, the variance and higher moments

1 Expectation of functions of X

Using the accounting trick, we can now derive the formula for computing the expectation of a function of X.

Formula. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a given function. If X is a discrete random variable with pmf $p_X(x)$, then

$$\mathbb{E}[\psi(X)] = \sum_{x \in S_X} \psi(x) p_X(x), \tag{1.1}$$

provided that the series on the right hand side is absolutely convergent. If X is a continuous random variable with pdf $f_X(x)$, then

$$\mathbb{E}[\psi(X)] = \int_{-\infty}^{\infty} \psi(x) f_X(x) dx, \tag{1.2}$$

provided that the integral on the right hand side is absolutely convergent.

Proof. We only consider the discrete case for simplicity. As seen before, we make use of the following partition:

$$\Omega = \cup_{x \in S_X} \{X = x\}.$$

According to the accounting trick, we have

$$\begin{aligned}
\mathbb{E}[\psi(X)] &= \sum_{\omega \in \Omega} \psi(X(\omega)) \mathbb{P}(\{\omega\}) = \sum_{x \in S_X} \sum_{\omega \in \{X = x\}} \psi(X(\omega)) \mathbb{P}(\{\omega\}) \\
&= \sum_{x \in S_X} \psi(x) \cdot \sum_{\omega \in \{X = x\}} \mathbb{P}(\{\omega\}) = \sum_{x \in S_X} \psi(x) \mathbb{P}(X = x) \\
&= \sum_{x \in S_X} \psi(x) p_X(x).
\end{aligned}$$

□

Example 1.1. Toss a fair die. Let X be the resulting value. Find $\mathbb{E}[X^2]$.

Solution. By using the formula (1.1), we have

$$\mathbb{E}[X^2] = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \cdots + 6^2 \times \frac{1}{6} = \frac{91}{6}.$$

On the other hand, we have seen in previous lecture that $\mathbb{E}[X] = \frac{7}{2}$. This tells us that in general $\mathbb{E}[X^2] \neq \mathbb{E}[X]^2$.

Example 1.2. Let X be a continuous random variable with pdf

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1; \\ 0, & \text{otherwhse.} \end{cases}$$

Compute $\mathbb{E}[X]$ and $\mathbb{E}[\frac{1}{X}]$.

Solution. By using the formula (1.2), we have

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \cdot 2x dx = \frac{2}{3},$$

and

$$\mathbb{E}[\frac{1}{X}] = \int_{-\infty}^{\infty} \frac{1}{x} f_X(x) dx = \int_0^1 \frac{1}{x} \cdot 2x dx = 2.$$

This example tells us that $\mathbb{E}[\frac{1}{X}] \neq \frac{1}{\mathbb{E}[X]}$ in general.

The above two examples suggest that, in general

$$\mathbb{E}[\psi(X)] \neq \psi(\mathbb{E}[X]).$$

There is one important exception, which is the case when ψ is a *linear* function, namely when $\psi(x) = ax + b$ where a, b are given real numbers.

Linearity property of the expectation. For any $a, b \in \mathbb{R}$, we have

$$\mathbb{E}[aX + b] = a \cdot \mathbb{E}[X] + b.$$

Proof. Let us prove the property for the continuous case:

$$\begin{aligned} \mathbb{E}[aX + b] &= \int_{-\infty}^{\infty} (ax + b) f_X(x) dx \qquad \text{(by the formula (1.2))} \\ &= a \cdot \int_{-\infty}^{\infty} x f_X dx + b \cdot \int_{-\infty}^{\infty} f_X(x) dx \qquad \text{(linearity of integration)} \\ &= a \cdot \mathbb{E}[X] + b \qquad (\text{since } \int_{-\infty}^{\infty} f_X(x) dx = 1). \end{aligned}$$

□

2 The variance of a random variable

We have seen that the expectation $\mathbb{E}[X]$ gives us the average value of a random variable. This can be used for instance as an important feature to decide whether an investment is worthy or not, according to whether the expected return is positive or negative. However, the expectation only provides first level information about the distribution that is often not sufficient for making decisions. For instance, we may also want to know the risk of the investment rather than just knowing the average return. One natural way of quantifying the risk is to introduce another quantity capturing

how much in magnitude that the random variable X deviates from its mean $\mathbb{E}[X]$?

If the average deviation is large, that means the investment is risky, while if the deviation is small, the investment is safer as the (random) return seems to be more stable (around its mean).

A simple way to capture the deviation magnitude is to look at the quantity $(X - \mathbb{E}[X])^2$ which is always non-negative. By considering the average of this quantity, we are led to another important feature for the distribution: the variance.

Definition 2.1. Let X be a random variable. The *variance* of X, denoted as $V[X]$ (or $\mathrm{Var}[X]$), is defined by

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

As we have mentioned, the variance measures the average degree to which the random variable deviates from its mean $\mathbb{E}[X]$. The smaller $V[X]$ is, the more often that X stays near $\mathbb{E}[X]$. The larger $V[X]$ is, the more likely that X spreads far away from its mean.

Remark 2.1. An alternative quantity for capturing the magnitude of deviation from $\mathbb{E}[X]$ is the absolute value $|X - \mathbb{E}[X]|$. One reason why we consider the variance instead of $\mathbb{E}[|X - \mathbb{E}[X]|]$ is because the latter is much harder to compute algebraically due to the presence of the absolute value function inside the expectation.

Example 2.1. Consider the batting performance of two cricketers. The first player scores either 100 or 0 with equal probability. The second player always scores 50. Let X_i be the score of the i-th cricketer ($i = 1, 2$). Then the two players have the same average score:

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = 50.$$

However, we have

$$V[X_1] = \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] = (100-50)^2 \times \frac{1}{2} + (0-50)^2 \times \frac{1}{2} = 2500,$$

while

$$V[X_2] = (50-50)^2 \times 1 = 0.$$

As a result, the second player is considered to be playing more consistently than the first one, although both have the same average performance.

Notation. We often use μ_X (or μ) to denote the expectation of X, and use σ_X^2 (or σ^2) to denote the variance. The notation σ_X (or σ) refers to the *standard deviation* of X, which is defined by $\sqrt{\mathrm{Var}[X]}$.

There are several important properties of the variance which we now discuss. Let $\mu_X = \mathbb{E}[X]$.

Property 1. The variance is always non-negative.

Proof. Recall that $V[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$ and the random variable $(X - \mathbb{E}[X])^2$ is always non-negative. As a result, its average must also be non-negative. $\square$

Remark 2.2. We are in fact using a general property that

$$Y \geqslant 0 \implies \mathbb{E}[Y] \geqslant 0.$$

Let us show this in the discrete case. Recall that

$$\mathbb{E}[Y] = \sum_{y \in S_Y} y \cdot \mathbb{P}(Y = y).$$

Since $Y \geqslant 0$, the possible values of Y are all non-negative. The above formula clearly yields a non-negative number in this case. A less obvious property is that, for a *non-negative* random variable Y, we have

$$\mathbb{E}[Y] = 0 \implies \mathbb{P}(Y = 0) = 1.$$

To see this, we again consider the discrete case. Suppose on the contrary that, there was a $y > 0$ with positive probability mass (i.e. $\mathbb{P}(Y = y) > 0$). Then the quantity $y \cdot \mathbb{P}(Y = y)$ gives a positive contribution to $\mathbb{E}[Y]$ and $\mathbb{E}[Y]$ cannot be zero in this case. Therefore, if $\mathbb{E}[Y] = 0$, no positive values can have non-zero masses. In other words, $\mathbb{P}(Y = 0) = 1$.

Property 2. We have $V[X] = 0$ if and only if $\mathbb{P}(X = \mu_X) = 1$. This is essentially saying that the random variable X is deterministic, as it equals a constant (its mean) with probability one.

Proof. Sufficiency. Suppose that $\mathbb{P}(X = \mu_X) = 1$. Then

$$V[X] = \mathbb{E}[(X - \mu_X)^2] = (\mu_X - \mu_X)^2 \times 1 = 0.$$

Necessity. Suppose that $V[X] = 0$. Then $\mathbb{E}[(X - \mu_X)^2] = 0$. Since $Y = (X - \mu_X)^2$ is a non-negative random variable, from the previous remark we conclude that $\mathbb{P}(Y = 0) = 1$, or equivalently, $\mathbb{P}(X = \mu_X) = 1$. □

Property 3. We can compute the variance via the following useful formula:

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

Proof. By definition and the linearity property of the expectation, we have

$$\begin{aligned}
V[X] &= \mathbb{E}[(X - \mu_X)^2] \\
&= \mathbb{E}[X^2 - 2\mu_X \cdot X + \mu_X^2] \\
&= \mathbb{E}[X^2] - 2\mu_X \cdot \mathbb{E}[X] + \mu_X^2 \cdot \mathbb{E}[1] \\
&= \mathbb{E}[X^2] - 2\mu_X^2 + \mu_X^2 \\
&= \mathbb{E}[X^2] - \mu_X^2.
\end{aligned}$$

□

Property 4. For any $a, b \in \mathbb{R}$, we have

$$V[aX + b] = a^2 V[X].$$

Proof. By definition,

$$\begin{aligned}
V[aX + b] &= \mathbb{E}\big[\big((aX + b) - \mathbb{E}[aX + b]\big)^2\big] \\
&= \mathbb{E}\big[\big(aX + b - a\mathbb{E}[X] - b\big)^2\big] \quad \text{(by linearity of expectation)} \\
&= \mathbb{E}[a^2(X - \mathbb{E}[X])^2] \\
&= a^2 V[X].
\end{aligned}$$

□

A useful corollary of Property 4 is the following. Suppose that X has mean μ and variance σ^2. Then the random variable $X_s = \frac{X-\mu}{\sigma}$ has mean 0 and variance 1. This new random variable is often known as the *standardisation* (or *normalisation*) of X.

Example 2.2. Toss a fair die. Let X be the resulting value. By using Property 3 and the results in Example 1.1, we have

$$V[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

3 Higher moments of a random variable

We can think of the mean μ_X and the variance σ_X^2 as the first two levels of information about the distribution. This quantities are far from being sufficient to encode all information about the distribution. In order to extract finer information, it is natural to consider "higher moments" of a random variable.

Definition 3.1. Let X be a random variable. For each $n \geqslant 1$, the *n-th moment* of X is defined by

$$\mu_n = \mathbb{E}[X^n].$$

Respectively, the *n-th central moment* of X is defined by

$$\nu_n = \mathbb{E}[(X - \mu_X)^n].$$

For instance, the expectation is the first moment and the variance is the second central moment. One important reason for introduce higher moments is that, in most of the interesting situations (but not always!), the sequence of moments

$$\{\mu_1, \mu_2, \mu_3, \cdots\}$$

uniquely determines the distribution of the random variable.

Computing moments via tail probabilities

There is a powerful way of computing the n-th moment of a *non-negative* random variable in terms of the tail probabilities. This is the content of the following result.

Formula. Let X be a random variable such that $\mathbb{P}(X \geqslant 0) = 1$. Then for each $n \geqslant 1$, we have

$$\mathbb{E}[X^n] = n \cdot \int_0^\infty x^{n-1}(1 - F_X(x))dx = n \cdot \int_0^\infty x^{n-1}\mathbb{P}(X > x)dx,$$

where $F_X(x)$ is the cdf of X. In particular, by taking $n = 1$ we have

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)dx.$$