

# 1 Expected Value and Variance

Let  $X$  be a continuous (potentially very complicated) random variable with PDF  $f$ . Suppose we were asked to pick *one real number*  $\mu$ , subject to a certain constraint, that best represents the random variable  $X$ . Of course, note that  $X$  is a random variable and might be very different than just one single (deterministic) real number. That being said, we want to try our best to pick said real number that satisfies this constraint.

## 1.1 Expected Value

One natural thing to do is to pick a  $\mu$  so that best represents the random variable (i.e., is closest to the random variable). So, we can consider the *squared distance*. In particular, we want to find a  $\mu$  such that

$$(X - \mu)^2$$

is likely to be as close to 0 as possible. Note that  $(X - \mu)^2$  is itself a random variable. Put it another way, we want to find a  $\mu$  that minimize the random distance  $(X - \mu)^2$ . So, we want to minimize

$$\int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

By linearity, we have

$$\begin{aligned} \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx &= \int_{-\infty}^{\infty} (x^2 - 2x\mu + \mu^2) f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx - 2\mu \int_{-\infty}^{\infty} x f(x) dx + \mu^2 \int_{-\infty}^{\infty} f(x) dx. \end{aligned}$$

Since  $f(x)$  is a PDF, the third integral is just 1. Hence, we want to minimize

$$\int_{-\infty}^{\infty} x^2 f(x) dx - 2\mu \int_{-\infty}^{\infty} x f(x) dx + \mu^2.$$

Differentiating with respect to  $\mu$ , we get

$$-2 \int_{-\infty}^{\infty} x f(x) dx + 2\mu.$$

If we set this equal to 0 and then solve for  $\mu$ , we find the minimizer is

$$\mu = \int_{-\infty}^{\infty} x f(x) dx.$$

We note that  $\mu$  is a **weighted average**. We integrate over all possible values of  $x$ , and then weigh them by their corresponding density  $f(x)$ .

### Definition 1.1: Expected Value

Suppose that  $X$  is a continuous random variable with PDF  $f$ . Then, the **expected value** (or **mean**) of  $X$  is

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

#### Remarks:

- For some random variables, this expected value (integral) does not exist. We won't need to worry about this here, though.
- If  $X$  is a **discrete** random variable with PMF  $p(x) = \mathbb{P}(X = x)$ , then  $\mu = \mathbb{E}(X) = \sum_x xp(x)$ .

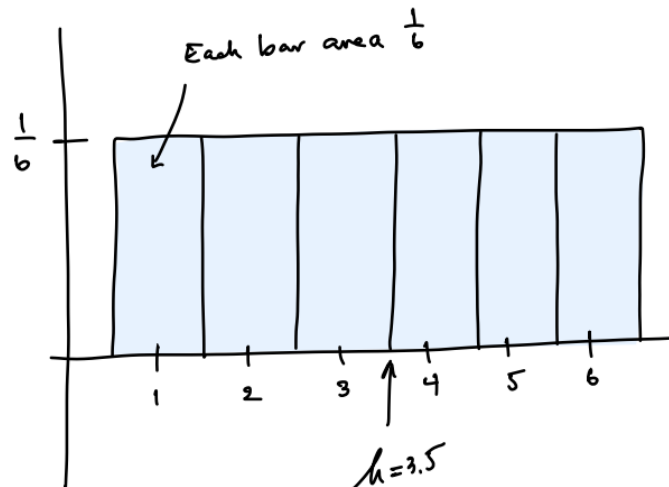
The **mean**  $\mu$  gives us valuable information about the “center” of a distribution. This, along with **standard deviation**  $\sigma$  (which tells us about the “spread” of a distribution), tells us the fundamental quantities in regards to the “shape” of a distribution.

(Example.) Suppose we roll a fair die, which takes values between 1 and 6, each with equal probability  $\frac{1}{6}$ . Then,

$$\mu = \sum_{k=1}^6 k \cdot \frac{1}{6} = \frac{1}{6} \sum_{k=1}^6 k = \frac{7}{2}.$$

Here, we're considering all possible values that the die can take (from  $k = 1$  to 6). Each value has a probability of  $\frac{1}{6}$ .

While  $\frac{7}{2}$  is not a value that the random variable can actually take (note that the random variable can only take values 1, 2, 3, 4, 5, and 6), but in some sense it is the expected value, or the average value. To see why this is the case, consider the following histogram:



We can think of the 3.5 mark as the “center of mass.”

### Theorem 1.1: Law of Unconscious Statistics (LotUS)

Suppose that  $X$  is a random variable and  $\phi : \mathbb{R} \mapsto \mathbb{R}$  is a function. Then,

1. If  $X$  is continuous with PDF  $f$ , then

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(x)f(x)dx.$$

2. If  $X$  is discrete with PMF  $p$ , then

$$\mathbb{E}[\phi(X)] = \sum_x \phi(x)p(x).$$

#### Remarks:

- It is enough *just* to know the PDF/PMF of  $X$  to find the expected value of  $Y = \phi(X)$ .
- This shows us that expectation is linear. In particular, the **Linearity of Expectation** (LoE) says that if we let  $X$  be a random variable, then we can suppose that  $a, b \in \mathbb{R}$ . Then,  $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ .

#### Warning:

- LoE does not imply that  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .
- Note that this is true if  $X$  and  $Y$  are independent, but not true otherwise.

## 1.2 Variance

### Definition 1.2: Variance

The **variance** of a random variable  $X$ , denoted by  $\sigma^2 = \text{Var}(X)$ , is the expected squared distance of  $X$  from its mean  $\mu$ . That is,  $\sigma^2 = \mathbb{E}[(X - \mu)^2]$ .

#### Remarks:

- This is telling us about the *spread* of the distribution; that is, how far away should we expect the random variable to be away from its mean.
- Recall that the mean  $\mu$  is the real number that minimizes  $\mathbb{E}[(X - a)^2]$  over all real numbers  $a$ . Whatever this quantity happens to be at the minimizer  $a = \mu$  is called the variance  $\sigma^2$ .

### Theorem 1.2

Let  $X$  be a random variable and let  $a, b \in \mathbb{R}$ . Then,

$$\text{Var}(aX + b) = a^2 \text{Var}(X) = a^2 \mathbb{E}[(X - \mu)^2].$$

**Remark:** Note that the variance of a constant is 0, since the distribution of a constant random variable has no “spread” at all.

**Warning:** While expected value is linear, variance is not. In other words, if  $X$  and  $Y$  are *independent*, then  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ . However, this is not true in general.

Note that the formula  $\sigma^2 = \mathbb{E}[(X - \mu)^2]$  is useful conceptually, but not so convenient for computations. Instead, we can use the Linearity of Expectation to obtain the following, more convenient, formula:

$$\sigma^2 = \mathbb{E}(X^2 - 2\mu X + \mu^2) = \mathbb{E}(X^2) - 2\mu\mu + \mu^2 = \mathbb{E}(X^2) - \mu^2.$$

Therefore, the **computational variance formula** is given by

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

The quantity  $\mathbb{E}(X^2)$  is called the **second moment** of  $X$ , and can be computed using LotUS. That is, if  $X$  is continuous, then

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx.$$

If  $X$  is discrete, then

$$\mathbb{E}(X^2) = \sum_x x^2 p(x).$$

## 1.3 Standard Deviation

### Definition 1.3: Standard Deviation

The square root of the variance,  $\sigma = \sqrt{\text{Var}(X)}$ , is called the **standard deviation** of  $X$ .

The reason for the name is that, for “most reasonable” distributions, the “bulk” of the mass/density deviates by at most  $\sigma$  from the center  $\mu$ . That is, usually, “most” of the mass/density is on the values of

$$x \in [\mu - \sigma, \mu + \sigma].$$

**Theorem 1.3: Chebyshev's Inequality**

Let  $X$  be a random variable with *both* mean  $\mu$  and standard deviation  $\sigma$ . Then, for any real number  $a > 0$ , we have that

$$\mathbb{P}(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}.$$

**Remarks:**

- In other words, for any distribution with a mean and variance, the probability that  $X$  is at least  $a = 2$  standard deviations from the center  $\mu$  of the distribution is at most  $\frac{1}{4}$ .
- For many distributions, the actual probability is *much* smaller. However, this is an useful upper bound that works for all distributions.

This result follows by

**Theorem 1.4: Markov's Inequality**

Let  $X$  be a non-negative random variable, i.e.  $\mathbb{P}(X \geq 0) = 1$ , with mean  $\mu$  and  $b > 0$  a positive number. Then,  $\mathbb{P}(X \geq b) \leq \frac{\mu}{b}$ .

*Proof.* (Continuous Case.) Since  $X$  is non-negative, we know that

$$\mu = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} xf(x)dx.$$

Splitting the integral, we have

$$\int_0^b xf(x)dx + \int_b^{\infty} xf(x)dx \geq \int_b^{\infty} xf(x)dx \geq b \int_b^{\infty} f(x)dx = b\mathbb{P}(X \geq b).$$

Hence,  $\mathbb{P}(X \geq b) \leq \frac{\mu}{b}$ , as claimed. □