5.1 Expectation

Unit 5: Averages

Adapted from Blitzstein-Hwang Chapters 4, 5, and 10.

Often it is useful to have one number summarizing the ''average" value of a random variable. There are several senses in which the word ''average" is used, but by far the most commonly used is the *mean* of an r.v., also known as its *expected value*. In addition, much of statistics is about understanding *variability* in the world, so it is often important to know how "spread out" the distribution is; we will formalize this with the concepts of *variance* and *standard deviation*. As we'll see, variance and standard deviation are defined in terms of expected values, so the uses of expected values go far beyond just computing averages.

Given a list of numbers $x_1, x_2, \ldots, x_n$, the familiar way to average them is to add them up and divide by n. This is called the *arithmetic mean*, and is defined by

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j.$$

More generally, we can define a *weighted mean* of $x_1, \ldots, x_n$ as

$$\text{weighted-mean}(x) = \sum_{j=1}^{n} x_j p_j,$$

where the weights $p_1, \ldots, p_n$ are pre-specified nonnegative numbers that add up to 1 (so the unweighted mean $\bar{x}$ is obtained when $p_j = 1/n$ for all j).

The definition of expectation for a discrete r.v. is inspired by the weighted mean of a list of numbers, with weights given by probabilities.

Definition 5.1.1 (Expectation of a discrete r.v.).

The *expected value* (also called the *expectation* or *mean*) of a discrete r.v. X whose distinct possible values are $x_1, x_2, \ldots$ is

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j).$$

If the support is finite, then this is replaced by a finite sum. We can also write

$$E(X) = \sum_{x} \underbrace{x}_{\text{value}} \underbrace{P(X = x)}_{\text{PMF at } x},$$

where the sum is over the support of X (in any case, $xP(X = x)$ is 0 for any x not in the support). The expectation is undefined if $\sum_{j=1}^{\infty} |x_j| P(X = x_j)$ diverges, since then the series for $E(X)$ diverges or its value depends on the order in which the x_j are listed.

In words, the expected value of X is a weighted average of the possible values that X can take on, weighted by their probabilities. Let's check that the definition makes sense in two simple examples:

1. Let X be the result of rolling a fair 6-sided die, so X takes on the values $1, 2, 3, 4, 5, 6$, with equal probabilities. Intuitively, we should be able to get the average by adding up these values and dividing by 6. Using the definition, the expected value is

   $$E(X) = \frac{1}{6}(1 + 2 + \cdots + 6) = 3.5,$$

   as we expected. Note though that X *never* equals its mean in this example. This is similar to the fact that the average number of children per household in some country could be 1.8, but that doesn't mean that a typical household has 1.8 children!

2. Let $X \sim \text{Bern}(p)$ and $q = 1 - p$. Then

   $$E(X) = 1p + 0q = p,$$

   which makes sense intuitively since it is between the two possible values of X, compromising between 0 and 1 based on how likely each is. This is illustrated in Figure 5.1.2 for a case with $p < 1/2$: two pebbles are being balanced on a seesaw. For the seesaw to balance, the fulcrum (shown as a triangle) must be at p, which in physics terms is the *center of mass*.

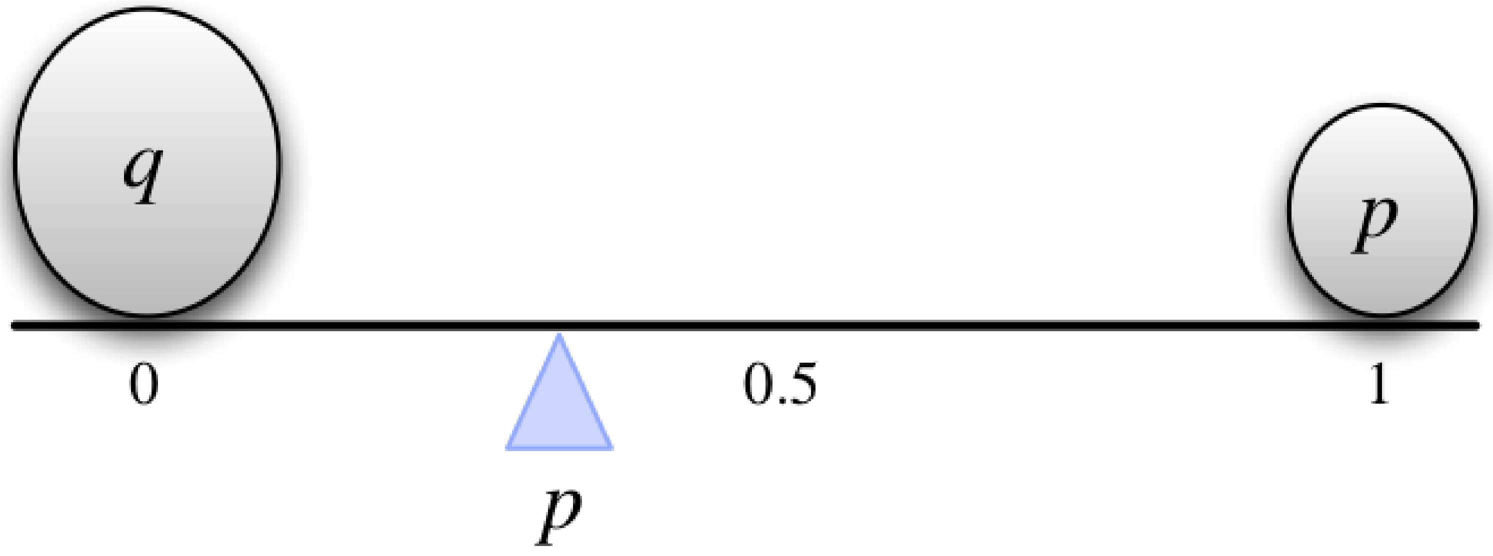

Figure 5.1.2: Center of mass of two pebbles, depicting that $E(X) = p$ for $X \sim \mathrm{Bern}(p)$. Here q and p denote the masses of the two pebbles.

View Larger Image
Image Description

☣ WARNING 5.1.3 (REPLACING AN R.V. BY ITS EXPECTATION).

For any discrete r.v. X, the expected value $E(X)$ is a *number* (if it exists). A common mistake is to replace an r.v. by its expectation without justification, which is wrong both mathematically (X is a function, $E(X)$ is a constant) and statistically (it ignores the variability of X), except in the degenerate case where X is a constant.

Notation 5.1.4.

We often abbreviate $E(X)$ to EX. Similarly, we often abbreviate $E(X^2)$ to EX^2. Thus EX^2 is the expectation of the random variable X^2, *not* the square of the number EX. In general, unless the parentheses explicitly indicate otherwise, the expectation is to be taken at the very end. For example, $E(X-3)^2$ is $E\left((X-3)^2\right)$, not $(E(X-3))^2$. As we will see, the order of operations here is very important!

Learn About Verified Certificates