3.3 Some properties of the expected value

If X is a discrete random variable with Range$(X) = \{x_1, x_2, x_3, ...\}$, then its expected value is the weighted average:

$$\mathbb{E}[X] = \sum_{x \in \text{Range}(X)} x P(X = x). \tag{3.7}$$

In this section we itemize some of the important properties of the expected value

3.3.1 The mean may not exist or be finite

To begin, we make the important point that the expected value may not be finite or even exist. The issue arises when X can take on an infinite number of values, since then (3.7) is a *series* (an infinite sum), which may not always converge. We distinguish between four cases:

1. If
$$\sum_{x \in \text{Range}(X), x>0} x P(X = x) < \infty, \qquad \sum_{x \in \text{Range}(X), x<0} x P(X = x) > -\infty,$$
then $\mathbb{E}[X]$ exists and is finite.

2. If
$$\sum_{x \in \text{Range}(X), x>0} x P(X = x) < \infty, \qquad \sum_{x \in \text{Range}(X), x<0} x P(X = x) = -\infty,$$
then $\mathbb{E}[X]$ exists and equals $-\infty$.

3. If
$$\sum_{x \in \text{Range}(X), x>0} x P(X = x) = \infty, \qquad \sum_{x \in \text{Range}(X), x<0} x P(X = x) > -\infty,$$
then $\mathbb{E}[X]$ exists and equals $-\infty$

4. If
$$\sum_{x \in \text{Range}(X), x>0} x P(X = x) = \infty, \qquad \sum_{x \in \text{Range}(X), x<0} x P(X = x) = -\infty,$$
then $\mathbb{E}[X]$ does not exist.

In the fourth case, we get an indeterminate form: $\infty - \infty$ is not defined. The following examples clarify the possibilities.

Example 3.4. Let X be a random variable with Range$(X) = \{1, \sqrt{2}, \sqrt{3}, ...\}$ and

$$P(X = \sqrt{n}) = \frac{6}{\pi^2 n^2}, \quad n = 1, 2, ...$$

We see that this is a valid pmf:

$$\begin{aligned}
\sum_{n=1}^{\infty} P(X=n) &= \sum_{n=1}^{\infty} \frac{6}{\pi^2 n^2} \\
&= \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \qquad \text{(well known } p\text{-series)} \\
&= \frac{6}{\pi^2} \frac{\pi^2}{6} \\
&= 1.
\end{aligned}$$

So the series of probabilities is a convergent series (as it should be). But what about the series given by $\sum_x xP(X=x)$?

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{n=1}^{\infty} \sqrt{n} P(X=\sqrt{n}) \\
&= \sum_{n=1}^{\infty} \sqrt{n} \frac{6}{\pi^2 n^2} \\
&= \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^{1.5}} \qquad \text{(harmonic series)} \\
&< \infty.
\end{aligned}$$

This is a p-series, $\sum_{n=1}^{\infty} \frac{1}{n^p}$ with $p = 1.5$. The p-series test tells us the series converges provided $p > 1$.

Example 3.5. Let X be a random variable with $\text{Range}(X) = \{2, 4, 8, 16, \dots\}$ and

$$P(X = 2^n) = \frac{1}{2^n}, \quad n = 1,\ 2,\ \dots$$

Then

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{n=1}^{\infty} 2^n P(X=2^n) \\
&= \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} \\
&= \sum_{n=1}^{\infty} 1 \\
&= \infty.
\end{aligned}$$

The expectation is infinite, but we can still say that it *exists*.

Example 3.6. Now we consider a very similar example to the one above where, instead of the expected value existing and being infinite, we say that it simply does not exist. Let X be a random variable with

$$\text{Range}(X) = \{\dots,\ -16,\ -8,\ -4,\ 4,\ 8,\ 16,\ \dots\} = \{-2^k : k = 2,\ 3,\ \dots\} \cup \{2^k : k = 2,\ 3,\ \dots\}$$

and pmf

$$P\left(X = \pm 2^k\right) = \frac{1}{2^k}, \quad k = 2, 3, ...$$

Is this a valid pmf? Yes. First, observe that

$$\sum_{x \in \text{Range}(X)} P(X = x) = \sum_{k=2}^{\infty} P(X = 2^k) + \sum_{k=2}^{\infty} P(X = -2^k) = \sum_{k=2}^{\infty} \frac{1}{2^k} + \sum_{k=2}^{\infty} \frac{1}{2^k} = 1,$$

using the geometric series formula.

But note that $\sum_{k=2}^{\infty} 2^k P(X = 2^k) = \sum_{k=2}^{\infty} 1 = \infty$, and similarly, $\sum_{k=2}^{\infty}(-2^k)P(X = -2^k) = \sum_{k=2}^{\infty}(-1) = -\infty$. We can't add ∞ and $-\infty$, so the expectation does not exist. Interestingly, if we had tried to compute the expected value by enumerating Range(X) as $-4, 4, -8, 8, -16, 16, \dots$, then we would be led to believe that the series converges, since

$$\begin{aligned}
\sum_{k=2}^{\infty} \left[2^k P(X = 2^k) + (-2^k)P(X = -2^k)\right] &= \sum_{k=2}^{\infty} \left[2^k \frac{1}{2^k} - 2^k \frac{1}{2^k}\right] \\
&= \sum_{k=2}^{\infty} 0 \\
&= 0.
\end{aligned}$$

We do not want our definition of expectation to depend on the order in which we arrange the values that X can take! So the fact that the expectation seems to exist under a different ordering does not reassure us. We still say the expectation does not exist.

3.3.2 The expected value of a transformation

A random variable X is a real number, so you can apply transformations to it: X^2, $\ln(X)$, $aX + b$, etc. Because X is random, these transformations will also be random variables with their own distributions. Later on, we will learn how to compute the entire distribution of a transformation of a random variable. But for now, we restrict ourselves to a simpler task: given some function g, how do you compute $E[g(X)]$? Here's an example of doing it the long way:

Example 3.7. Start with a random variable X, where Range(X) $= \{-1, 0, 1\}$, $P(X = -1) = 2/10$, $P(X = 0) = 5/10$, and $P(X = 1) = 3/10$. Next, define a new random variable $Y = g(X) = X^2$. This new random variable has Range(Y) $= \{0, 1\}$, with $P(Y = 0) = P(X = 0) = 1/2$ and

$$P(Y = 1) = P(X \in \{-1, 1\}) = P(X = -1) + P(X = 1) = \frac{2}{10} + \frac{3}{10} = \frac{1}{2}.$$

So it turns out that $Y \sim \text{Bern}(1/2)$, and so $E(Y) = E[g(X)] = E(X^2) = 0.5$.

In order to compute the expected value of X^2 in this example, we first derived the entire distribution of X^2, and then applied the definition of the expected value to the transformed random variable. That was simple enough in this toy example, but in other contexts this could be hard, and if possible we would like to skip this intermediate step of deriving the whole distribution before computing one measly expected value. Theorem 3.1 provides the shortcut we crave.

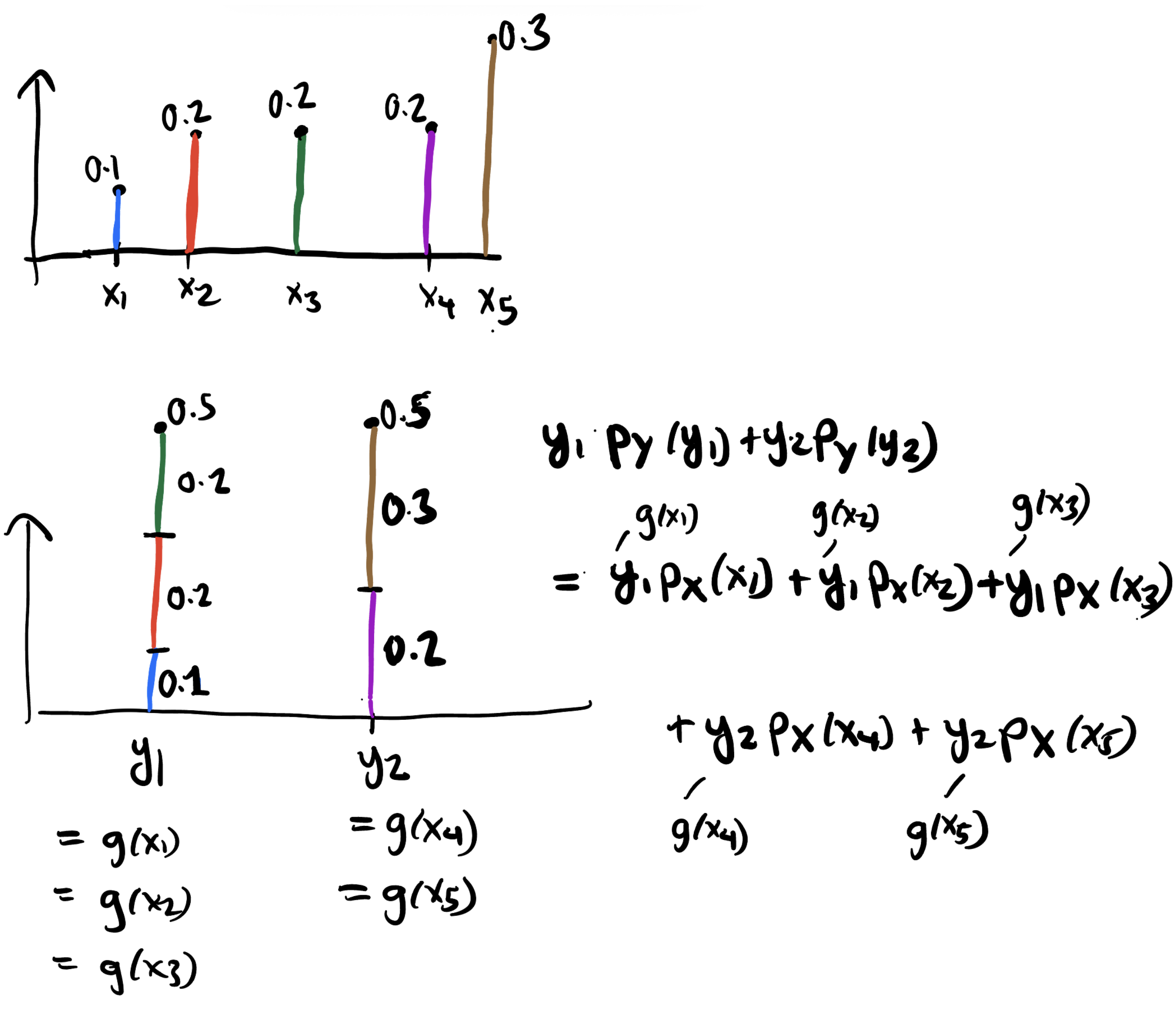

Figure 3.9

Theorem 3.1. If X is a discrete random variable and g : Range$(X) \to \mathbb{R}$ is a transformation, then

$$\mathbb{E}[g(X)] = \sum_{x \in \text{Range}(X)} g(x) \cdot P(X = x). \tag{3.8}$$

See Figure 3.9 for an example demonstrating why Theorem 3.1 is true. In the figure, we are using the key fact that $P(Y = y)$ is just the sum of all $P(X = x)$'s such that $g(x) = y$.

The upshot of Theorem 3.1 is that it allows you to compute $\mathbb{E}[g(X)]$ without having to first derive the range and distribution of the new random variable $g(X)$. The statement of $\mathbb{E}[g(X)]$ only refers to Range(X) and the pmf of X. This is very convenient.

Remark 3.3. Theorem 3.1 is called the **law of the unconscious statistician** (**LOTUS**) because equation 3.8 seems too good to be true. It seems like it's a mistake – the kind that an *unconscious statistician* might carelessly make in a bout of wishful thinking. Sort of like if you weren't paying close attention and blithely wrote that $(x + y)^2$ equaled $x^2 + y^2$. But unlike this, and fortunately for us, LOTUS is as true as the sky is blue.

Example 3.8. Consider the same X from Example 3.7. Theorem 3.1 now tells us that

$$\begin{aligned}\mathbb{E}[g(X)] = \mathbb{E}[X^2] &= (-1)^2\frac{2}{10} + 0^2\frac{5}{10} + 1^2\frac{3}{10}\\ &= \frac{2}{10} + \frac{3}{10}\\ &= \frac{1}{2}.\end{aligned}$$

So we get the same answer as before, but without the intermediate step of deriving the entire distribution of X^2. We only had to make use of the range and pmf of the original X.

Theorem 3.2. Let $a, b \in \mathbb{R}$ be constants, and consider any random variable X for which $\mathbb{E}[X]$ exists and is finite. Then

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b. \tag{3.9}$$

In other words, "the expected value of a linear transformation is the linear transformation of the expected value."

Proof. Let $\text{Range}(X) = \{x_1, x_2, ...\}$. We are interested in $E[g(X)]$ for $g(x) = ax + b$, so LOTUS tells us that

$$\begin{aligned}\mathbb{E}[aX + b] &= \sum_{i=1}^{\infty}(ax_i + b)P(X = x_i)\\ &= \sum_{i=1}^{\infty}[ax_iP(X = x_i) + bP(X = x_i)]\\ &= a\underbrace{\sum_{i=1}^{\infty}x_iP(X = x_i)}_{\text{Definition 3.5}} + b\underbrace{\sum_{i=1}^{\infty}P(X = x_i)}_{1}\\ &= a\mathbb{E}[X] + b.\end{aligned}$$

□

Remark 3.4. Theorem 3.2 establishes that for *linear* functions $g(x)$ of the form $g(x) = ax + b$, we have $\mathbb{E}[g(X)] = g(\mathbb{E}[X])$. This is *NOT* true for other functions g! Thinking $\mathbb{E}[g(X)] = g(\mathbb{E}[X])$ is also something an unconscious statistician might do (like LOTUS), but this time it does NOT work. So don't make this mistake.

Theorem 3.3. If $a_1, a_2, ..., a_n \in \mathbb{R}$ are constants and $X_1, X_2, ..., X_n$ are random variables for which $\mathbb{E}[X_i]$ all exist and are finite, then

$$\mathbb{E}\left[\sum_{i=1}^{n} a_iX_i\right] = \sum_{i=1}^{n} a_i\mathbb{E}[X_i]. \tag{3.10}$$

In other words, "the expected value of a linear combination is the linear combination of the expected values."

Once we have more tools, we will see a partial proof of this result in the case where $n = 2$. The general case then follows by induction.

Remark 3.5. For example, a special case of this result tells us that for two random variables X and Y, we have $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. To properly make sense of the lefthand side, we need to know the pmf of $X + Y$. This pmf depends on the *joint pmf*, which is the function $f_{X,Y}(x, y) = P(\{X = x\} \cap \{Y = y\})$. We will talk about joint pmfs later in the course. In any case, the theorem tells us that, whatever the pmf of $X + Y$ may be, we can compute the expected value of $X + Y$ just by adding $\mathbb{E}[X]$ and $\mathbb{E}[Y]$. So really, for the sake of computing an expectation only, we don't actually have to know the pmf of $X + Y$.

Remark 3.6. Theorem 3.3 is very general. It applies to any collection of random variables, regardless their type (discrete, continuous or neither) or their dependence structure. As long as each has finite mean, the theorem holds.

Example 3.9. If $X \sim \text{Binom}(n, p)$, we saw that $\mathbb{E}[X] = np$. But the calculation was ugly. We can use linearity to clean it up. A binomial random variable can be expressed as a sum of Bernoulli indicators. That is, for $i = 1, 2, ..., n$, let $I_i \sim \text{Bern}(p)$. Furthermore, assume these are all independent. Since the I_i all have the same p, we can summarize this situation by writing $I_i \overset{\text{iid}}{\sim} \text{Bern}(p)$, where iid stands for **independent and identically distributed**. Each I_i represents a 0/1 trial, they are all independent, and they have the same probability of success. As such, a new random variable X that counts the number of successes in these n trials is equivalent to $X = \sum_{i=1}^{n} I_i \sim \text{Binom}(n, p)$. Recalling that the expectation of a Bernoulli is $\mathbb{E}[I_i] = p$, we have by the linearity of expectation that

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} I_i\right] = \sum_{i=1}^{n} \mathbb{E}[I_i] = \sum_{i=1}^{n} p = np.$$

Much easier! Granted, we have yet to prove that the expected value is actually a linear operator. If doing so were a prerequisite for performing this particular calculation, perhaps you would not regard it as particularly easy. But taking linearity for granted, this is a far more elegant solution than when we proceeding from the definition.