Random Variables: Expectation

Random variables allow us to answer questions like "What is the number of heads in n coin tosses?" While we understand exactly what distribution answers that question, not all distributions are so simple to describe. We can instead ask: "What value does a random variable X *typically* take?" also makes sense — it is asking the average value of X is if we sample it a large number of times. This average value is called the expectation of X, and is one of the most useful summary (also called *statistics*) of an experiment.

1 Expectation

The distribution of a r.v. contains *all* the probabilistic information about the r.v. In most applications, however, the complete distribution of a r.v. is very hard to calculate. For example, consider the homework permutation example from Note 15 with $n = 20$. In principle, we would have to enumerate $20! \approx 2.4 \times 10^{18}$ sample points, compute the value of X at each one, and count the number of points at which X takes on each of its possible values (though in practice we could streamline this calculation a bit)! Moreover, even when we can compute the complete distribution of a r.v., it is often not very informative.

For these reasons, we seek to *summarize* the distribution into a more compact, convenient form that is also easier to compute. The most widely used such form is the *expectation* (or *mean* or *average*) of the r.v.

Definition 15.1 (Expectation)**.** *The expectation of a discrete random variable X is defined as*

$$\mathbb{E}[X] = \sum_{a \in \mathscr{A}} a \times \mathbb{P}[X = a], \tag{1}$$

where the sum is over all possible values taken by the r.v.

(**Technical Note.** Expectation is well defined provided that the sum on the right hand side of (1) is absolutely convergent, i.e., $\sum_{a \in \mathscr{A}} |a| \times \mathbb{P}[X = a] < \infty$. There are random variables for which expectations do not exist.)

For our simpler permutation example with only 3 students, the expectation is

$$\mathbb{E}[X] = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{2}\right) + \left(3 \times \frac{1}{6}\right) = 0 + \frac{1}{2} + \frac{1}{2} = 1.$$

That is, the expected number of fixed points in a permutation of three items is exactly 1.

The expectation can be seen in some sense as a "typical" value of the r.v. (though note that $\mathbb{E}[X]$ may not actually be a value that X can take). The question of how typical the expectation is for a given r.v. is a very important one that we shall return to in a later lecture.

Here is a physical interpretation of the expectation of a random variable: imagine carving out a wooden cutout figure of the probability distribution as in Figure 1. Then the expected value of the distribution is the balance point (directly below the center of gravity) of this object.

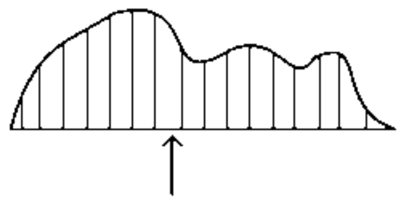

Figure 1: Physical interpretation of expected value as the balance point.

1.1 Examples

1. **Single die.** Throw a fair die once and let X be the number that comes up. Then X takes on values $1,2,\ldots,6$ each with probability $\frac{1}{6}$, so

$$\mathbb{E}[X] = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = \frac{7}{2}.$$

   Note that X never actually takes on its expected value $\frac{7}{2}$.

2. **Two dice.** Throw two fair dice and let X be the sum of their scores. Then the distribution of X is

a	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}[X=a]$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

   The expectation is therefore

$$\mathbb{E}[X] = \left(2\times\frac{1}{36}\right) + \left(3\times\frac{1}{18}\right) + \left(4\times\frac{1}{12}\right) + \cdots + \left(12\times\frac{1}{36}\right) = 7.$$

3. **Roulette.** A roulette wheel is spun (recall that a roulette wheel has 38 slots: the numbers $1,2,\ldots,36$, half of which are red and half black, plus 0 and 00, which are green). You bet \$1 on Black. If a black number comes up, you receive your stake plus \$1; otherwise you lose your stake. Let X be your net winnings in one game. Then X can take on the values $+1$ and -1, and $\mathbb{P}[X=1] = \frac{18}{38}$, $\mathbb{P}[X=-1] = \frac{20}{38}$. Thus,

$$\mathbb{E}[X] = \left(1\times\frac{18}{38}\right) + \left(-1\times\frac{20}{38}\right) = -\frac{1}{19};$$

   i.e., you expect to lose about a nickel per game. Notice how the zeros tip the balance in favor of the casino!

4. **Function of a RV.** Let X be a random variable with values in $\mathbb{R}$, and let f be a function from $\mathbb{R}$ to $\mathbb{R}$. Let $\mathscr{A}$ be the set of all values taken by X. Recall that $f(X)$ is also a random variable, that take values $f(a)$ with probability $\mathbb{P}[X=a]$, for all $a\in\mathscr{A}$. Thus, applying the definition of expectation:

$$\mathbb{E}[f(X)] = \sum_{a\in\mathscr{A}} f(a)\times\mathbb{P}[X=a]$$

1.2 Expectation of a Poisson Random Variable

We calculate the expectation of a Poisson random variable.

Theorem 15.1. *For a Poisson random variable $X \sim \text{Poisson}(\lambda)$, we have $\mathbb{E}[X] = \lambda$.*

Proof. We can calculate $\mathbb{E}[X]$ directly from the definition of expectation:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \times \mathbb{P}[X = i] \\
&= \sum_{i=1}^{\infty} i \, \frac{\lambda^i}{i!} \mathrm{e}^{-\lambda} && \text{(the } i = 0 \text{ term is equal to 0 so we omit it)} \\
&= \lambda \mathrm{e}^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda \mathrm{e}^{-\lambda} \mathrm{e}^{\lambda} && \text{(since } \mathrm{e}^{\lambda} = \textstyle\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \text{ with } j = i-1) \\
&= \lambda.
\end{aligned}$$

□

2 Linearity of Expectation

So far, we have computed expectations by brute force: i.e., we have written down the whole distribution and then added up the contributions for all possible values of the r.v. The real power of expectations is that in many real-life examples they can be computed much more easily using a simple shortcut. The shortcut is the following:

Theorem 15.2. *For any two random variables X and Y on the same probability space, we have*

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Also, for any constant c, we have

$$\mathbb{E}[cX] = c\,\mathbb{E}[X].$$

Proof. We first rewrite the definition of expectation in a more convenient form. Recall from Definition 15.1 that

$$\mathbb{E}[X] = \sum_{a \in \mathscr{A}} a \times \mathbb{P}[X = a].$$

Consider a particular term $a \times \mathbb{P}[X = a]$ in the above sum. Notice that $\mathbb{P}[X = a]$, by definition, is the sum of $\mathbb{P}[\omega]$ over those sample points ω for which $X(\omega) = a$. Furthermore, we know that every sample point $\omega \in \Omega$ is in exactly one of these events $X = a$. This means we can write out the above definition in a more long-winded form as

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \times \mathbb{P}[\omega]. \tag{2}$$

This equivalent definition of expectation will make the present proof much easier (though it is usually less convenient for actual calculations). Applying (2) to $\mathbb{E}[X + Y]$ gives:

$$\begin{aligned}
\mathbb{E}[X + Y] &= \textstyle\sum_{\omega \in \Omega} (X + Y)(\omega) \times \mathbb{P}[\omega] \\
&= \textstyle\sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \times \mathbb{P}[\omega] \\
&= \textstyle\sum_{\omega \in \Omega} \Big(X(\omega) \times \mathbb{P}[\omega] \Big) + \textstyle\sum_{\omega \in \Omega} \Big(Y(\omega) \times \mathbb{P}[\omega] \Big) \\
&= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}$$

In the last step, we used (2) twice.

This completes the proof of the first equality. The proof of the second equality is much simpler and is left as an exercise. □

Theorem 15.2 is very powerful: it says that the expectation of a sum of r.v.'s is the sum of their expectations, with no assumptions about the r.v.'s. We can use Theorem 15.2 to conclude things like $\mathbb{E}[3X-5Y] = 3\mathbb{E}[X]-5\mathbb{E}[Y]$, regardless of whether or not X and Y are independent. This important property is known as linearity of expectation.

Important caveat: Theorem 15.2 does **not** *say that* $\mathbb{E}[XY]=\mathbb{E}[X]\,\mathbb{E}[Y]$*, or that* $\mathbb{E}\left[\frac{1}{X}\right]=\frac{1}{\mathbb{E}[X]}$*, etc. These claims are not true in general. It is only* sums *and* differences *and* constant multiples *of random variables that behave so nicely.*

2.1 Applications of Linearity of Expectation

Now let us see some examples of Theorem 15.2 in action.

1. **Two dice again.** Here is a much less painful way of computing $\mathbb{E}[X]$, where X is the sum of the scores of the two dice. Note that $X = Y_1 + Y_2$, where Y_i is the score on die i. We know from example 1 in Section 1.1 that $\mathbb{E}[Y_1] = \mathbb{E}[Y_2] = \frac{7}{2}$. So, by Theorem 15.2, we have $\mathbb{E}[X] = \mathbb{E}[Y_1]+\mathbb{E}[Y_2] = 7$.

2. **More roulette.** Suppose we play the roulette game mentioned in Section 1.1 $n \geq 1$ times. Let X_n be our expected net winnings. Then $X_n = Y_1 + Y_2 + \cdots + Y_n$, where Y_i is our net winnings in the i-th play. We know from earlier that $\mathbb{E}[Y_i] = -\frac{1}{19}$ for each i. Therefore, by Theorem 15.2, $\mathbb{E}[X_n] = \mathbb{E}[Y_1]+\mathbb{E}[Y_2]+\cdots+\mathbb{E}[Y_n] = -\frac{n}{19}$. For $n = 1000$, $\mathbb{E}[X_n] = -\frac{1000}{19} \approx -53$, so if you play 1000 games, you expect to lose about $53.

3. **Fixed points of permutations.** Let us return to the homework permutation example with an arbitrary number n of students. Let X_n denote the number of students who receive their own homework after shuffling (or equivalently, the number of fixed points). To take advantage of Theorem 15.2, we need to write X_n as a *sum* of simpler r.v.'s. Since X_n *counts* the number of times something happens, we can write it as a sum using the following useful trick:

$$X_n = I_1 + I_2 + \cdots + I_n, \qquad \text{where } I_i = \begin{cases} 1, & \text{if student } i \text{ gets their own homework,} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

[You should think about this equation for a moment. Remember that all the I_i's are random variables. What does an equation involving random variables mean? What we mean is that, *at every sample point* ω, we have $X_n(\omega) = I_1(\omega)+I_2(\omega)+\cdots+I_n(\omega)$. Why is this true?]

A Bernoulli random variable such as I_i is called an indicator random variable of the corresponding event (in this case, the event that student i gets their own homework). For indicator r.v.'s, the expectation is particularly easy to calculate. Specifically,

$$\mathbb{E}[I_i] = (0 \times \mathbb{P}[I_i = 0]) + (1 \times \mathbb{P}[I_i = 1]) = \mathbb{P}[I_i = 1].$$

In our case, we have

$$\mathbb{P}[I_i = 1] = \mathbb{P}[\text{student } i \text{ gets their own homework}] = \frac{1}{n}.$$

We can now apply Theorem 15.2 to (3), yielding

$$\mathbb{E}[X_n] = \mathbb{E}[I_1] + \mathbb{E}[I_2] + \cdots + \mathbb{E}[I_n] = n \times \frac{1}{n} = 1.$$

So, we see that the expected number of students who get their own homeworks in a class of size n is 1. That is, the expected number of fixed points in a random permutation of n items is always 1, regardless of n.

4. **Coin tosses.** Toss a fair coin $n \geq 1$ times. Let the r.v. X_n be the number of heads observed. As in the previous example, to take advantage of Theorem 15.2 we write

$$X_n = I_1 + I_2 + \cdots + I_n,$$

where I_i is the indicator r.v. of the event that the i-th toss is H. Since the coin is fair, we have

$$\mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \mathbb{P}[i\text{-th toss is } H] = \tfrac{1}{2}.$$

Using Theorem 15.2, we therefore get

$$\mathbb{E}[X_n] = \sum_{i=1}^{n} \frac{1}{2} = \frac{n}{2}.$$

In n tosses of a biased coin that shows H with probability p, $\mathbb{E}[X_n] = np$. (Check this.) So the expectation of a binomial r.v. $X \sim \text{Bin}(n, p)$ is equal to np. Note that it would have been harder to reach the same conclusion by computing this directly from the definition of expectation shown in (1).

5. **Balls and bins.** Throw m balls into n bins. Let the r.v. X be the number of balls that land in the first bin. Then X behaves exactly like the number of heads in m tosses of a biased coin with $\mathbb{P}[H] = \frac{1}{n}$ (why?). So, from the previous example, we get $\mathbb{E}[X] = \frac{m}{n}$. In the special case $m = n$, the expected number of balls in any bin is 1. If we wanted to compute this directly from the distribution of X, we would get into a messy calculation involving binomial coefficients.

Here is another example on the same sample space. Let the r.v. Y_n be the number of empty bins. The distribution of Y_n is horrible to contemplate: to get a feel for this, you might like to write it down for $m = n = 3$ (i.e., 3 balls, 3 bins). However, computing the expectation $\mathbb{E}[Y_n]$ is easy using Theorem 15.2. As in previous two examples, we write

$$Y_n = I_1 + I_2 + \cdots + I_n, \tag{4}$$

where I_i is the indicator r.v. of the event "bin i is empty". The expectation of I_i is easy to find:

$$\mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \mathbb{P}[\text{bin } i \text{ is empty}] = \left(1 - \frac{1}{n}\right)^m,$$

as discussed in Note 13. Applying Theorem 15.2 to (4), we therefore obtain

$$\mathbb{E}[Y_n] = \sum_{i=1}^{n} \mathbb{E}[I_i] = n\left(1 - \frac{1}{n}\right)^m,$$

a simple formula, quite easily derived. Let us see how it behaves in the special case $m = n$ (same number of balls as bins). In this case we get $\mathbb{E}[Y_n] = n\big(1 - \frac{1}{n}\big)^n$. Now the quantity $\big(1 - \frac{1}{n}\big)^n$ can be approximated (for large enough values of n) by the number $\frac{1}{e}$.[1] So we see that, for large n,

$$\mathbb{E}[Y_n] \approx \frac{n}{e} \approx 0.368n.$$

The bottom line is that, if we throw (say) 1000 balls into 1000 bins, the expected number of empty bins is about 368.

3 The Tail Sum Formula

Theorem 15.3 (Tail Sum Formula)**.** *Let X be a random variable that takes values in $\{0,1,2,\dots\}$. Then*

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}[X \geq i].$$

Proof. For notational convenience, let's write $p_i = \mathbb{P}[X = i]$, for $i = 0,1,2,\dots$. From the definition of expectation, we have

$$\begin{aligned}
\mathbb{E}[X] &= (0 \times p_0) + (1 \times p_1) + (2 \times p_2) + (3 \times p_3) + (4 \times p_4) + \cdots \\
&= p_1 + (p_2 + p_2) + (p_3 + p_3 + p_3) + (p_4 + p_4 + p_4 + p_4) + \cdots \\
&= (p_1 + p_2 + p_3 + p_4 + \cdots) + (p_2 + p_3 + p_4 + \cdots) + (p_3 + p_4 + \cdots) + (p_4 + \cdots) + \cdots \\
&= \mathbb{P}[X \geq 1] + \mathbb{P}[X \geq 2] + \mathbb{P}[X \geq 3] + \mathbb{P}[X \geq 4] + \cdots.
\end{aligned}$$

In the third line, we have regrouped the terms into convenient infinite sums, and each infinite sum is exactly the probability that $X \geq i$ for each i. You should check that you understand how the fourth line follows from the third.

Let us repeat the proof more formally, this time using more compact mathematical notation:

$$\mathbb{E}[X] = \sum_{j=1}^{\infty} j \times \mathbb{P}[X = j] = \sum_{j=1}^{\infty}\sum_{i=1}^{j} \mathbb{P}[X = j] = \sum_{i=1}^{\infty}\sum_{j=i}^{\infty} \mathbb{P}[X = j] = \sum_{i=1}^{\infty} \mathbb{P}[X \geq i],$$

where the third equality follows from interchanging the order of summations. □

3.1 Mean of a Geometric Random Variable

Let us now compute the expectation $\mathbb{E}[X]$. Applying the definition of expected value directly gives us:

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} i \times \mathbb{P}[X = i] = p \sum_{i=1}^{\infty} i(1-p)^{i-1}.$$

To compute this summation, we will use the tail sum bound.

We can now use Theorem 15.3 to compute $\mathbb{E}[X]$ more easily.

[1] More generally, it is a standard fact that for any constant c,

$$(1 + \tfrac{c}{n})^n \to e^c \quad \text{as } n \to \infty.$$

We just used this fact in the special case $c = -1$. The approximation is actually very good even for quite small values of n. (Try it yourself!) E.g., for $n = 20$ we already get $(1 - \frac{1}{n})^n \approx 0.358$, which is very close to $\frac{1}{e} \approx 0.368$. The approximation gets better and better for larger n.

Theorem 15.4. *For* $X \sim \text{Geometric}(p)$*, we have* $\mathbb{E}[X] = \frac{1}{p}$.

Proof. The key observation is that for a geometric random variable X,

$$\mathbb{P}[X \geq i] = (1-p)^{i-1} \text{ for } i = 1,2,\ldots. \tag{5}$$

We can obtain this simply by summing $\mathbb{P}[X = j]$ for $j \geq i$. Another way of seeing this is to note that the event "$X \geq i$" means at least i tosses are required. This is equivalent to saying that the first $i-1$ tosses are all Tails, and the probability of this event is precisely $(1-p)^{i-1}$. Now, plugging equation (5) into Theorem 15.3, we get

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}[X \geq i] = \sum_{i=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-(1-p)} = \frac{1}{p},$$

where we have used the formula for geometric series. □

So, the expected number of tosses of a biased coin until the first Head appears is $\frac{1}{p}$. Intuitively, if in each coin toss we expect to get p Heads, then we need to toss the coin $\frac{1}{p}$ times to get 1 Head. So for a fair coin, the expected number of tosses is 2, but remember that the actual number of coin tosses that we need can be any positive integers.

Remark: Another way of deriving $\mathbb{E}[X] = \frac{1}{p}$ is to use the interpretation of a geometric random variable X as the number of coin tosses until we get a Head. Consider what happens in the first coin toss: If the first toss comes up Heads, then $X = 1$. Otherwise, we have used one toss, and we repeat the coin tossing process again; the number of coin tosses after the first toss is again a geometric random variable with parameter p. Therefore, we can calculate:

$$\mathbb{E}[X] = \underbrace{p \cdot 1}_{\text{first toss is H}} + \underbrace{(1-p) \cdot (1 + \mathbb{E}[X])}_{\text{first toss is T, then toss again}}.$$

Solving for $\mathbb{E}[X]$ yields $\mathbb{E}[X] = \frac{1}{p}$, as claimed.