The *expectation* or *expected value* of a random variable is a single number that reveals a lot about the behavior of the variable. The expectation of a random variable is also known as its *mean* or *average*. For example, the first thing you typically want to know when you see your grade on an exam is the average score of the class. This average score turns out to be precisely the expectation of the random variable equal to the score of a random student.

More precisely, the expectation of a random variable is its "average" value when each value is weighted according to its probability. Formally, the expected value of a random variable is defined as follows:

**Definition 18.4.1.** If $R$ is a random variable defined on a sample space $\mathcal{S}$, then the expectation of $R$ is

$$\text{Ex}[R] ::= \sum_{\omega \in \mathcal{S}} R(\omega) \Pr[\omega]. \tag{18.2}$$

Let's work through some examples.

### 18.4.1 The Expected Value of a Uniform Random Variable

Rolling a 6-sided die provides an example of a uniform random variable. Let $R$ be the value that comes up when you roll a fair 6-sided die. Then by (18.2), the expected value of $R$ is

$$\text{Ex}[R] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

This calculation shows that the name "expected" value is a little misleading; the random variable might *never* actually take on that value. No one expects to roll a $3\frac{1}{2}$ on an ordinary die!

In general, if $R_n$ is a random variable with a uniform distribution on $\{a_1, a_2, \ldots, a_n\}$, then the expectation of $R_n$ is simply the average of the $a_i$'s:

$$\text{Ex}[R_n] = \frac{a_1 + a_2 + \cdots + a_n}{n}.$$

### 18.4.2 The Expected Value of a Reciprocal Random Variable

Define a random variable $S$ to be the reciprocal of the value that comes up when you roll a fair 6-sided die. That is, $S = 1/R$ where $R$ is the value that you roll. Now,

$$\text{Ex}[S] = \text{Ex}\left[\frac{1}{R}\right] = \frac{1}{1} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{49}{120}.$$

Notice that

$$\text{Ex}\left[1/R\right] \neq 1/\text{Ex}[R].$$

Assuming that these two quantities are equal is a common mistake.

### 18.4.3 The Expected Value of an Indicator Random Variable

The expected value of an indicator random variable for an event is just the probability of that event.

**Lemma 18.4.2.** *If $I_A$ is the indicator random variable for event $A$, then*

$$\text{Ex}[I_A] = \Pr[A].$$

*Proof.*

$$\begin{aligned}\text{Ex}[I_A] &= 1 \cdot \Pr[I_A = 1] + 0 \cdot \Pr[I_A = 0] = \Pr[I_A = 1] \\ &= \Pr[A]. \qquad \text{(def of } I_A\text{)}\end{aligned}$$

For example, if $A$ is the event that a coin with bias $p$ comes up heads, then $\text{Ex}[I_A] = \Pr[I_A = 1] = p$.

### 18.4.4 Alternate Definition of Expectation

There is another standard way to define expectation.

**Theorem 18.4.3.** *For any random variable $R$,*

$$\text{Ex}[R] = \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x]. \tag{18.3}$$

The proof of Theorem 18.4.3, like many of the elementary proofs about expectation in this chapter, follows by regrouping of terms in equation (18.2):

*Proof.* Suppose $R$ is defined on a sample space $\mathcal{S}$. Then,

$$\begin{aligned}
\text{Ex}[R] &::= \sum_{\omega \in \mathcal{S}} R(\omega) \Pr[\omega] \\
&= \sum_{x \in \text{range}(R)} \sum_{\omega \in [R=x]} R(\omega) \Pr[\omega] \\
&= \sum_{x \in \text{range}(R)} \sum_{\omega \in [R=x]} x \Pr[\omega] && \text{(def of the event } [R = x]) \\
&= \sum_{x \in \text{range}(R)} x \left( \sum_{\omega \in [R=x]} \Pr[\omega] \right) && \text{(factoring } x \text{ from the inner sum)} \\
&= \sum_{x \in \text{range}(R)} x \cdot \Pr[R = x]. && \text{(def of } \Pr[R = x])
\end{aligned}$$

The first equality follows because the events $[R = x]$ for $x \in \text{range}(R)$ partition the sample space $\mathcal{S}$, so summing over the outcomes in $[R = x]$ for $x \in \text{range}(R)$ is the same as summing over $\mathcal{S}$. ■

In general, equation (18.3) is more useful than the defining equation (18.2) for calculating expected values. It also has the advantage that it does not depend on the sample space, but only on the density function of the random variable. On the other hand, summing over all outcomes as in equation (18.2) sometimes yields easier proofs about general properties of expectation.

### 18.4.5 Conditional Expectation

Just like event probabilities, expectations can be conditioned on some event. Given a random variable $R$, the expected value of $R$ conditioned on an event $A$ is the probability-weighted average value of $R$ over outcomes in $A$. More formally:

**Definition 18.4.4.** The *conditional expectation* $\text{Ex}[R \mid A]$ of a random variable $R$ given event $A$ is:

$$\text{Ex}[R \mid A] ::= \sum_{r \in \text{range}(R)} r \cdot \Pr\left[R = r \mid A\right]. \tag{18.4}$$

For example, we can compute the expected value of a roll of a fair die, given that the number rolled is at least 4. We do this by letting $R$ be the outcome of a roll of the die. Then by equation (18.4),

$$\text{Ex}[R \mid R \geq 4] = \sum_{i=1}^{6} i \cdot \Pr\left[R = i \mid R \geq 4\right] = 1 \cdot 0 + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot \tfrac{1}{3} + 5 \cdot \tfrac{1}{3} + 6 \cdot \tfrac{1}{3} = 5.$$

Conditional expectation is useful in dividing complicated expectation calculations into simpler cases. We can find a desired expectation by calculating the conditional expectation in each simple case and averaging them, weighing each case by its probability.

For example, suppose that 49.6% of the people in the world are male and the rest female—which is more or less true. Also suppose the expected height of a randomly chosen male is $5'\,11''$, while the expected height of a randomly chosen female is $5'\,5.''$ What is the expected height of a randomly chosen person? We can calculate this by averaging the heights of men and women. Namely, let $H$ be the height (in feet) of a randomly chosen person, and let $M$ be the event that the person is male and $F$ the event that the person is female. Then

$$\begin{aligned}
\text{Ex}[H] &= \text{Ex}[H \mid M] \Pr[M] + \text{Ex}[H \mid F] \Pr[F] \\
&= (5 + 11/12) \cdot 0.496 + (5 + 5/12) \cdot (1 - 0.496) \\
&= 5.6646 \ldots .
\end{aligned}$$

which is a little less than 5’ 8.”

This method is justified by:

**Theorem 18.4.5** (Law of Total Expectation)**.** *Let $R$ be a random variable on a sample space $\mathcal{S}$, and suppose that $A_1, A_2, \ldots,$ is a partition of $\mathcal{S}$. Then*

$$\text{Ex}[R] = \sum_i \text{Ex}[R \mid A_i] \Pr[A_i].$$

*Proof.*

$$\begin{aligned}
\text{Ex}[R] &= \sum_{r \in \text{range}(R)} r \cdot \Pr[R = r] && \text{(by 18.3)} \\
&= \sum_r r \cdot \sum_i \Pr\left[R = r \mid A_i\right] \Pr[A_i] && \text{(Law of Total Probability)} \\
&= \sum_r \sum_i r \cdot \Pr\left[R = r \mid A_i\right] \Pr[A_i] && \text{(distribute constant } r\text{)} \\
&= \sum_i \sum_r r \cdot \Pr\left[R = r \mid A_i\right] \Pr[A_i] && \text{(exchange order of summation)} \\
&= \sum_i \Pr[A_i] \sum_r r \cdot \Pr\left[R = r \mid A_i\right] && \text{(factor constant } \Pr[A_i]\text{)} \\
&= \sum_i \Pr[A_i] \, \text{Ex}[R \mid A_i]. && \text{(Def 18.4.4 of cond. expectation)}
\end{aligned}$$

■

### 18.4.6 Mean Time to Failure

A computer program crashes at the end of each hour of use with probability $p$, if it has not crashed already. What is the expected time until the program crashes? This will be easy to figure out using the Law of Total Expectation, Theorem 18.4.5. Specifically, we want to find $\text{Ex}[C]$ where $C$ is the number of hours until the first crash. We'll do this by conditioning on whether or not the crash occurs in the first hour.

So define $A$ to be the event that the system fails on the first step and $\overline{A}$ to be the complementary event that the system does not fail on the first step. Then the mean time to failure $\text{Ex}[C]$ is

$$\text{Ex}[C] = \text{Ex}[C \mid A] \Pr[A] + \text{Ex}[C \mid \overline{A}] \Pr[\overline{A}]. \tag{18.5}$$

Since $A$ is the condition that the system crashes on the first step, we know that

$$\text{Ex}[C \mid A] = 1. \tag{18.6}$$

Since $\overline{A}$ is the condition that the system does *not* crash on the first step, conditioning on $\overline{A}$ is equivalent to taking a first step without failure and then starting over without conditioning. Hence,

$$\text{Ex}[C \mid \overline{A}] = 1 + \text{Ex}[C]. \tag{18.7}$$

Plugging (18.6) and (18.7) into (18.5):

$$\begin{aligned}\text{Ex}[C] &= 1 \cdot p + (1 + \text{Ex}[C])(1 - p) \\ &= p + 1 - p + (1 - p)\,\text{Ex}[C] \\ &= 1 + (1 - p)\,\text{Ex}[C].\end{aligned}$$

Then, rearranging terms gives

$$1 = \text{Ex}[C] - (1 - p)\,\text{Ex}[C] = p\,\text{Ex}[C],$$

and thus

$$\text{Ex}[C] = 1/p.$$

The general principle here is well-worth remembering.

## Mean Time to Failure

If a system independently fails at each time step with probability $p$, then the expected number of steps up to the first failure is $1/p$.

So, for example, if there is a 1% chance that the program crashes at the end of each hour, then the expected time until the program crashes is $1/0.01 = 100$ hours.

As a further example, suppose a couple insists on having children until they get a boy, then how many baby girls should they expect before their first boy? Assume for simplicity that there is a 50% chance that a child will be a boy and that the genders of siblings are mutually independent.

This is really a variant of the previous problem. The question, "How many hours until the program crashes?" is mathematically the same as the question, "How many children must the couple have until they get a boy?" In this case, a crash corresponds to having a boy, so we should set $p = 1/2$. By the preceding analysis, the couple should expect a baby boy after having $1/p = 2$ children. Since the last of these will be a boy, they should expect just one girl. So even in societies where couples pursue this commitment to boys, the expected population will divide evenly between boys and girls.

There is a simple intuitive argument that explains the mean time to failure formula (18.8). Suppose the system is restarted after each failure. This makes the mean time to failure the same as the mean time between successive repeated failures. Now if the probability of failure at a given step is $p$, then after $n$ steps we expect to have $pn$ failures. Now, by definition, the average number of steps between failures is equal to $np/p$, namely, $1/p$.

For the record, we'll state a formal version of this result. A random variable like $C$ that counts steps to first failure is said to have a *geometric distribution* with parameter $p$.

**Definition 18.4.6.** A random variable, $C$, has a *geometric distribution* with parameter $p$ iff codomain$(C) = \mathbb{Z}^+$ and

$$\Pr[C = i] = (1 - p)^{i-1} p.$$

**Lemma 18.4.7.** *If a random variable $C$ has a geometric distribution with parameter $p$, then*

$$\text{Ex}[C] = \frac{1}{p}. \tag{18.8}$$

### 18.4.7 Expected Returns in Gambling Games

Some of the most interesting examples of expectation can be explained in terms of gambling games. For straightforward games where you win $w$ dollars with probability $p$ and you lose $x$ dollars with probability $1 - p$, it is easy to compute your *expected return* or *winnings*. It is simply

$$pw - (1 - p)x \text{ dollars.}$$

For example, if you are flipping a fair coin and you win \$1 for heads and you lose \$1 for tails, then your expected winnings are

$$\frac{1}{2} \cdot 1 - \left(1 - \frac{1}{2}\right) \cdot 1 = 0.$$

In such cases, the game is said to be *fair* since your expected return is zero.

#### Splitting the Pot

We'll now look at a different game which is fair—but only on first analysis.

It's late on a Friday night in your neighborhood hangout when two new biker dudes, Eric and Nick, stroll over and propose a simple wager. Each player will put \$2 on the bar and secretly write "heads" or "tails" on their napkin. Then you will flip a fair coin. The \$6 on the bar will then be "split"—that is, be divided equally—among the players who correctly predicted the outcome of the coin toss. Pot splitting like this is a familiar feature in poker games, betting pools, and lotteries.

This sounds like a fair game, but after your regrettable encounter with strange dice (Section 16.3), you are definitely skeptical about gambling with bikers. So before agreeing to play, you go through the four-step method and write out the

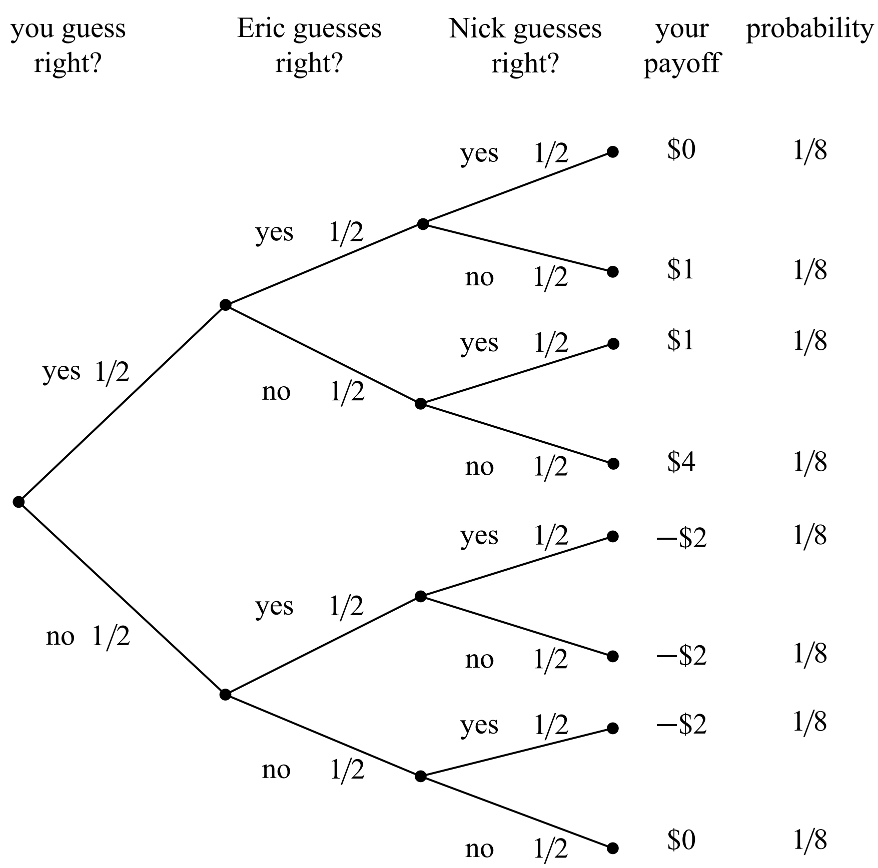


**Figure 18.6** The tree diagram for the game where three players each wager $2 and then guess the outcome of a fair coin toss. The winners split the pot.

tree diagram to compute your expected return. The tree diagram is shown in Figure 18.6.

The “payoff” values in Figure 18.6 are computed by dividing the \$6 pot[1] among those players who guessed correctly and then subtracting the \$2 that you put into the pot at the beginning. For example, if all three players guessed correctly, then your payoff is \$0, since you just get back your \$2 wager. If you and Nick guess correctly and Eric guessed wrong, then your payoff is

$$\frac{6}{2} - 2 = 1.$$

In the case that everyone is wrong, you all agree to split the pot and so, again, your payoff is zero.

To compute your expected return, you use equation (18.3):

$$\begin{aligned}\text{Ex}[\text{payoff}] &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} \\ &\quad + (-2) \cdot \frac{1}{8} + (-2) \cdot \frac{1}{8} + (-2) \cdot \frac{1}{8} + 0 \cdot \frac{1}{8} \\ &= 0.\end{aligned}$$

This confirms that the game is fair. So, for old time’s sake, you break your solemn vow to never ever engage in strange gambling games.

### The Impact of Collusion

Needless to say, things are not turning out well for you. The more times you play the game, the more money you seem to be losing. After 1000 wagers, you have lost over \$500. As Nick and Eric are consoling you on your “bad luck,” you do a back-of-the-envelope calculation and decide that the probability of losing \$500 in 1000 fair \$2 wagers is very, very small.

Now it is possible of course that you are very, very unlucky. But it is more likely that something fishy is going on. Somehow the tree diagram in Figure 18.6 is not a good model of the game.

The “something” that’s fishy is the opportunity that Nick and Eric have to collude against you. The fact that the coin flip is fair certainly means that each of Nick and Eric can only guess the outcome of the coin toss with probability $1/2$. But when you look back at the previous 1000 bets, you notice that Eric and Nick never made the same guess. In other words, Nick always guessed “tails” when Eric guessed “heads,” and vice-versa. Modelling this fact now results in a slightly different tree diagram, as shown in Figure 18.7.

[1] The money invested in a wager is commonly referred to as the *pot*.

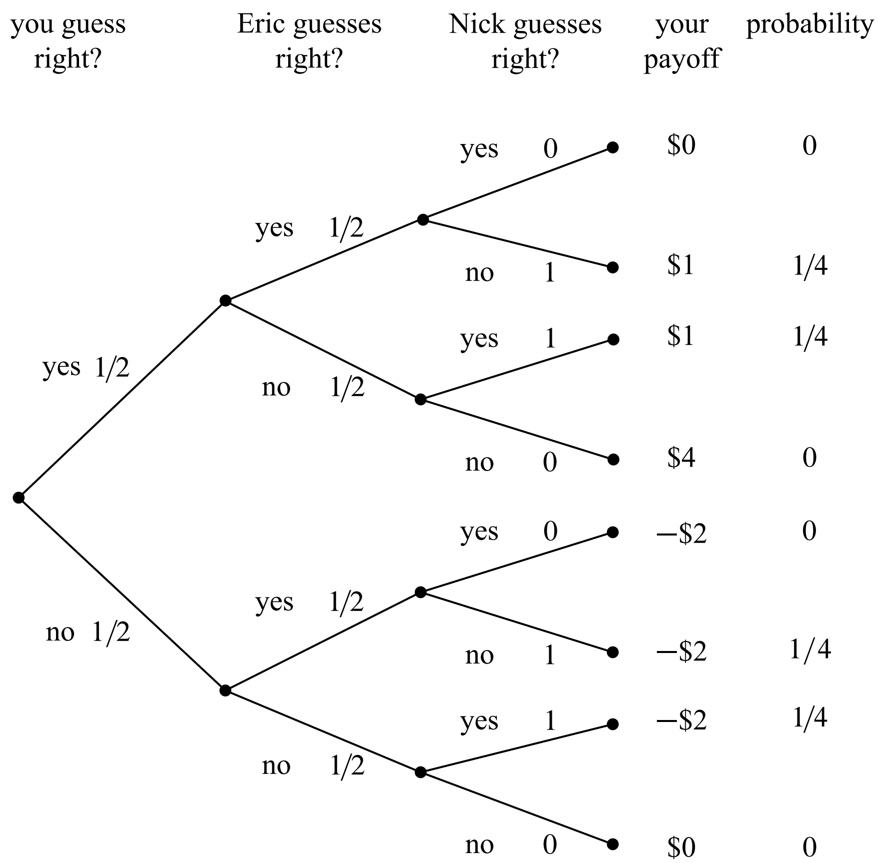


**Figure 18.7** The revised tree diagram reflecting the scenario where Nick always guesses the opposite of Eric.

The payoffs for each outcome are the same in Figures 18.6 and 18.7, but the probabilities of the outcomes are different. For example, it is no longer possible for all three players to guess correctly, since Nick and Eric are always guessing differently. More importantly, the outcome where your payoff is \$4 is also no longer possible. Since Nick and Eric are always guessing differently, one of them will always get a share of the pot. As you might imagine, this is not good for you!

When we use equation (18.3) to compute your expected return in the collusion scenario, we find that

$$\begin{aligned}\text{Ex}[\text{payoff}] &= 0 \cdot 0 + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 4 \cdot 0 \\ &\quad + (-2) \cdot 0 + (-2) \cdot \frac{1}{4} + (-2) \cdot \frac{1}{4} + 0 \cdot 0 \\ &= -\frac{1}{2}.\end{aligned}$$

So watch out for these biker dudes! By colluding, Nick and Eric have made it so that you expect to lose \$.50 every time you play. No wonder you lost \$500 over the course of 1000 wagers.

### How to Win the Lottery

Similar opportunities to collude arise in many betting games. For example, consider the typical weekly football betting pool, where each participant wagers \$10 and the participants that pick the most games correctly split a large pot. The pool seems fair if you think of it as in Figure 18.6. But, in fact, if two or more players collude by guessing differently, they can get an "unfair" advantage at your expense!

In some cases, the collusion is inadvertent and you can profit from it. For example, many years ago, a former MIT Professor of Mathematics named Herman Chernoff figured out a way to make money by playing the state lottery. This was surprising since the state usually takes a large share of the wagers before paying the winners, and so the expected return from a lottery ticket is typically pretty poor. So how did Chernoff find a way to make money? It turned out to be easy!

In a typical state lottery,

- all players pay \$1 to play and select 4 numbers from 1 to 36,
- the state draws 4 numbers from 1 to 36 uniformly at random,
- the states divides 1/2 of the money collected among the people who guessed correctly and spends the other half redecorating the governor's residence.

This is a lot like the game you played with Nick and Eric, except that there are more players and more choices. Chernoff discovered that a small set of numbers was selected by a large fraction of the population. Apparently many people think the same way; they pick the same numbers not on purpose as in the previous game with Nick and Eric, but based on the Red Sox winning average or today's date. The result is as though the players were intentionally colluding to lose. If any one of them guessed correctly, then they'd have to split the pot with many other players. By selecting numbers uniformly at random, Chernoff was unlikely to get one of these favored sequences. So if he won, he'd likely get the whole pot! By analyzing actual state lottery data, he determined that he could win an average of 7 cents on the dollar. In other words, his expected return was not −$.50 as you might think, but +$.07.[2] Inadvertent collusion often arises in betting pools and is a phenomenon that you can take advantage of.

## 18.5 Linearity of Expectation

Expected values obey a simple, very helpful rule called *Linearity of Expectation*. Its simplest form says that the expected value of a sum of random variables is the sum of the expected values of the variables.

**Theorem 18.5.1.** *For any random variables $R_1$ and $R_2$,*

$$\text{Ex}[R_1 + R_2] = \text{Ex}[R_1] + \text{Ex}[R_2].$$

*Proof.* Let $T ::= R_1 + R_2$. The proof follows straightforwardly by rearranging terms in equation (18.2) in the definition of expectation:

$$\begin{aligned}
\text{Ex}[T] &::= \sum_{\omega \in \mathcal{S}} T(\omega) \cdot \Pr[\omega] \\
&= \sum_{\omega \in \mathcal{S}} (R_1(\omega) + R_2(\omega)) \cdot \Pr[\omega] && \text{(def of } T\text{)} \\
&= \sum_{\omega \in \mathcal{S}} R_1(\omega) \Pr[\omega] + \sum_{\omega \in \mathcal{S}} R_2(\omega) \Pr[\omega] && \text{(rearranging terms)} \\
&= \text{Ex}[R_1] + \text{Ex}[R_2]. && \text{(by (18.2))}
\end{aligned}$$

■

A small extension of this proof, which we leave to the reader, implies

[2] Most lotteries now offer randomized tickets to help smooth out the distribution of selected sequences.

**Theorem 18.5.2.** *For random variables $R_1$, $R_2$ and constants $a_1, a_2 \in \mathbb{R}$,*

$$\text{Ex}[a_1 R_1 + a_2 R_2] = a_1 \text{Ex}[R_1] + a_2 \text{Ex}[R_2].$$

In other words, expectation is a linear function. A routine induction extends the result to more than two variables:

**Corollary 18.5.3** (Linearity of Expectation)**.** *For any random variables $R_1, \dots, R_k$ and constants $a_1, \dots, a_k \in \mathbb{R}$,*

$$\text{Ex}\left[\sum_{i=1}^{k} a_i R_i\right] = \sum_{i=1}^{k} a_i \text{Ex}[R_i].$$

The great thing about linearity of expectation is that *no independence is required*. This is really useful, because dealing with independence is a pain, and we often need to work with random variables that are not known to be independent.

As an example, let's compute the expected value of the sum of two fair dice.

### 18.5.1 Expected Value of Two Dice

What is the expected value of the sum of two fair dice?

Let the random variable $R_1$ be the number on the first die, and let $R_2$ be the number on the second die. We observed earlier that the expected value of one die is 3.5. We can find the expected value of the sum using linearity of expectation:

$$\text{Ex}[R_1 + R_2] = \text{Ex}[R_1] + \text{Ex}[R_2] = 3.5 + 3.5 = 7.$$

Assuming that the dice were independent, we could use a tree diagram to prove that this expected sum is 7, but this would be a bother since there are 36 cases. And without assuming independence, it's not apparent how to apply the tree diagram approach at all. But notice that we did *not* have to assume that the two dice were independent. The expected sum of two dice is 7—even if they are controlled to act together in some way—as long as each individual controlled die remains fair.

### 18.5.2 Sums of Indicator Random Variables

Linearity of expectation is especially useful when you have a sum of indicator random variables. As an example, suppose there is a dinner party where $n$ men check their hats. The hats are mixed up during dinner, so that afterward each man receives a random hat. In particular, each man gets his own hat with probability $1/n$. What is the expected number of men who get their own hat?

Letting $G$ be the number of men that get their own hat, we want to find the expectation of $G$. But all we know about $G$ is that the probability that a man gets his own hat back is $1/n$. There are many different probability distributions of hat permutations with this property, so we don't know enough about the distribution of $G$ to calculate its expectation directly using equation (18.2) or (18.3). But linearity of expectation lets us sidestep this issue.

We'll use a standard, useful trick to apply linearity, namely, we'll express $G$ as a sum of indicator variables. In particular, let $G_i$ be an indicator for the event that the $i$th man gets his own hat. That is, $G_i = 1$ if the $i$th man gets his own hat, and $G_i = 0$ otherwise. The number of men that get their own hat is then the sum of these indicator random variables:

$$G = G_1 + G_2 + \cdots + G_n. \tag{18.9}$$

These indicator variables are *not* mutually independent. For example, if $n - 1$ men all get their own hats, then the last man is certain to receive his own hat. But again, we don't need to worry about this dependence, since linearity holds regardless.

Since $G_i$ is an indicator random variable, we know from Lemma 18.4.2 that

$$\text{Ex}[G_i] = \Pr[G_i = 1] = 1/n. \tag{18.10}$$

By Linearity of Expectation and equation (18.9), this means that

$$\begin{aligned}
\text{Ex}[G] &= \text{Ex}[G_1 + G_2 + \cdots + G_n] \\
&= \text{Ex}[G_1] + \text{Ex}[G_2] + \cdots + \text{Ex}[G_n] \\
&= \overbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}^{n} \\
&= 1.
\end{aligned}$$

So even though we don't know much about how hats are scrambled, we've figured out that on average, just one man gets his own hat back, regardless of the number of men with hats!

More generally, Linearity of Expectation provides a very good method for computing the expected number of events that will happen.

**Theorem 18.5.4.** *Given any collection of events $A_1, A_2, \ldots, A_n$, the expected number of events that will occur is*

$$\sum_{i=1}^{n} \Pr[A_i].$$

For example, $A_i$ could be the event that the $i$th man gets the right hat back. But in general, it could be any subset of the sample space, and we are asking for the expected number of events that will contain a random sample point.

*Proof.* Define $R_i$ to be the indicator random variable for $A_i$, where $R_i(\omega) = 1$ if $w \in A_i$ and $R_i(\omega) = 0$ if $w \notin A_i$. Let $R = R_1 + R_2 + \cdots + R_n$. Then

$$\begin{aligned}
\text{Ex}[R] &= \sum_{i=1}^{n} \text{Ex}[R_i] && \text{(by Linearity of Expectation)} \\
&= \sum_{i=1}^{n} \Pr[R_i = 1] && \text{(by Lemma 18.4.2)} \\
&= \sum_{i=1}^{n} \Pr[A_i]. && \text{(def of indicator variable)}
\end{aligned}$$

So whenever you are asked for the expected number of events that occur, all you have to do is sum the probabilities that each event occurs. Independence is not needed.

### 18.5.3 Expectation of a Binomial Distribution

Suppose that we independently flip $n$ biased coins, each with probability $p$ of coming up heads. What is the expected number of heads?

Let $J$ be the random variable denoting the number of heads. Then $J$ has a binomial distribution with parameters $n$, $p$, and

$$\Pr[J = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Applying equation (18.3), this means that

$$\text{Ex}[J] = \sum_{k=0}^{n} k \Pr[J = k] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}. \tag{18.11}$$

This sum looks a tad nasty, but linearity of expectation leads to an easy derivation of a simple closed form. We just express $J$ as a sum of indicator random variables, which is easy. Namely, let $J_i$ be the indicator random variable for the $i$th coin coming up heads, that is,

$$J_i ::= \begin{cases} 1 & \text{if the } i\text{th coin is heads} \\ 0 & \text{if the } i\text{th coin is tails.} \end{cases}$$

Then the number of heads is simply

$$J = J_1 + J_2 + \cdots + J_n.$$

By Theorem 18.5.4,

$$\text{Ex}[J] = \sum_{i=1}^{n} \Pr[J_i] = pn. \tag{18.12}$$

That really was easy. If we flip $n$ mutually independent coins, we expect to get $pn$ heads. Hence the expected value of a binomial distribution with parameters $n$ and $p$ is simply $pn$.

But what if the coins are not mutually independent? It doesn't matter—the answer is still $pn$ because Linearity of Expectation and Theorem 18.5.4 do not assume any independence.

If you are not yet convinced that Linearity of Expectation and Theorem 18.5.4 are powerful tools, consider this: without even trying, we have used them to prove a complicated looking identity, namely,

$$\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = pn, \tag{18.13}$$

which follows by combining equations (18.11) and (18.12) (see also Exercise 18.26).

The next section has an even more convincing illustration of the power of linearity to solve a challenging problem.

### 18.5.4 The Coupon Collector Problem

Every time we purchase a kid's meal at Taco Bell, we are graciously presented with a miniature "Racin' Rocket" car together with a launching device which enables us to project our new vehicle across any tabletop or smooth floor at high velocity. Truly, our delight knows no bounds.

There are different colored Racin' Rocket cars. The color of car awarded to us by the kind server at the Taco Bell register appears to be selected uniformly and independently at random. What is the expected number of kid's meals that we must purchase in order to acquire at least one of each color of Racin' Rocket car?

The same mathematical question shows up in many guises: for example, what is the expected number of people you must poll in order to find at least one person with each possible birthday? The general question is commonly called the *coupon collector problem* after yet another interpretation.

A clever application of linearity of expectation leads to a simple solution to the coupon collector problem. Suppose there are five different colors of Racin' Rocket cars, and we receive this sequence:

blue green green red blue orange blue orange gray.

Let's partition the sequence into 5 segments:

$$\underbrace{\text{blue}}_{X_0} \quad \underbrace{\text{green}}_{X_1} \quad \underbrace{\text{green} \quad \text{red}}_{X_2} \quad \underbrace{\text{blue} \quad \text{orange}}_{X_3} \quad \underbrace{\text{blue} \quad \text{orange} \quad \text{gray}}_{X_4}.$$

The rule is that a segment ends whenever we get a new kind of car. For example, the middle segment ends when we get a red car for the first time. In this way, we can break the problem of collecting every type of car into stages. Then we can analyze each stage individually and assemble the results using linearity of expectation.

In the general case there are $n$ colors of Racin' Rockets that we're collecting. Let $X_k$ be the length of the $k$th segment. The total number of kid's meals we must purchase to get all $n$ Racin' Rockets is the sum of the lengths of all these segments:

$$T = X_0 + X_1 + \cdots + X_{n-1}.$$

Now let's focus our attention on $X_k$, the length of the $k$th segment. At the beginning of segment $k$, we have $k$ different types of car, and the segment ends when we acquire a new type. When we own $k$ types, each kid's meal contains a type that we already have with probability $k/n$. Therefore, each meal contains a new type of car with probability $1 - k/n = (n-k)/n$. Thus, the expected number of meals until we get a new kind of car is $n/(n-k)$ by the Mean Time to Failure rule. This means that

$$\text{Ex}[X_k] = \frac{n}{n-k}.$$

Linearity of expectation, together with this observation, solves the coupon collector problem:

$$\begin{aligned}
\text{Ex}[T] &= \text{Ex}[X_0 + X_1 + \cdots + X_{n-1}] \\
&= \text{Ex}[X_0] + \text{Ex}[X_1] + \cdots + \text{Ex}[X_{n-1}] \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \cdots + \frac{n}{3} + \frac{n}{2} + \frac{n}{1} \\
&= n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) \\
&= n\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1} + \frac{1}{n}\right) \\
&= nH_n \qquad (18.14) \\
&\sim n \ln n.
\end{aligned}$$

Cool! It's those Harmonic Numbers again.

We can use equation (18.14) to answer some concrete questions. For example, the expected number of die rolls required to see every number from 1 to 6 is:

$$6H_6 = 14.7\ldots.$$

And the expected number of people you must poll to find at least one person with each possible birthday is:

$$365H_{365} = 2364.6\ldots.$$

### 18.5.5 Infinite Sums

Linearity of expectation also works for an infinite number of random variables provided that the variables satisfy an absolute convergence criterion.

**Theorem 18.5.5** (Linearity of Expectation)**.** *Let $R_0$, $R_1$, ..., be random variables such that*

$$\sum_{i=0}^{\infty} \text{Ex}[\,|R_i|\,]$$

*converges. Then*

$$\text{Ex}\left[\sum_{i=0}^{\infty} R_i\right] = \sum_{i=0}^{\infty} \text{Ex}[R_i].$$

*Proof.* Let $T ::= \sum_{i=0}^{\infty} R_i$.

We leave it to the reader to verify that, under the given convergence hypothesis, all the sums in the following derivation are absolutely convergent, which justifies rearranging them as follows:

$$\begin{aligned}
\sum_{i=0}^{\infty} \text{Ex}[R_i] &= \sum_{i=0}^{\infty} \sum_{s \in \mathcal{S}} R_i(s) \cdot \Pr[s] && \text{(Def. 18.4.1)} \\
&= \sum_{s \in \mathcal{S}} \sum_{i=0}^{\infty} R_i(s) \cdot \Pr[s] && \text{(exchanging order of summation)} \\
&= \sum_{s \in \mathcal{S}} \left[\sum_{i=0}^{\infty} R_i(s)\right] \cdot \Pr[s] && \text{(factoring out } \Pr[s]) \\
&= \sum_{s \in \mathcal{S}} T(s) \cdot \Pr[s] && \text{(Def. of } T) \\
&= \text{Ex}[T] && \text{(Def. 18.4.1)} \\
&= \text{Ex}\left[\sum_{i=0}^{\infty} R_i\right]. && \text{(Def. of } T).
\end{aligned}$$

■

### 18.5.6 A Gambling Paradox

One of the simplest casino bets is on "red" or "black" at the roulette table. In each play at roulette, a small ball is set spinning around a roulette wheel until it lands in a red, black, or green colored slot. The payoff for a bet on red or black matches the bet; for example, if you bet \$10 on red and the ball lands in a red slot, you get back your original \$10 bet plus another matching \$10.

The casino gets its advantage from the green slots, which make the probability of both red and black each less than 1/2. In the US, a roulette wheel has 2 green slots among 18 black and 18 red slots, so the probability of red is $18/38 \approx 0.473$. In Europe, where roulette wheels have only 1 green slot, the odds for red are a little better—that is, $18/37 \approx 0.486$—but still less than even.

Of course you can't expect to win playing roulette, even if you had the good fortune to gamble against a *fair* roulette wheel. To prove this, note that with a fair wheel, you are equally likely win or lose each bet, so your expected win on any spin is zero. Therefore if you keep betting, your expected win is the sum of your expected wins on each bet: still zero.

Even so, gamblers regularly try to develop betting strategies to win at roulette despite the bad odds. A well known strategy of this kind is *bet doubling*, where you bet, say, \$10 on red and keep doubling the bet until a red comes up. This means you stop playing if red comes up on the first spin, and you leave the casino with a \$10 profit. If red does not come up, you bet \$20 on the second spin. Now if the second spin comes up red, you get your \$20 bet plus \$20 back and again walk away with a net profit of $\$20 - 10 = \$10$. If red does not come up on the second spin, you next bet \$40 and walk away with a net win of $\$40 - 20 - 10 = \$10$ if red comes up on on the third spin, and so on.

Since we've reasoned that you can't even win against a fair wheel, this strategy against an unfair wheel shouldn't work. But wait a minute! There is a 0.486 probability of red appearing on each spin of the wheel, so the mean time until a red occurs is less than three. What's more, red will come up *eventually* with probability one, and as soon as it does, you leave the casino \$10 ahead. In other words, by bet doubling you are *certain* to win \$10, and so your expectation is \$10, not zero!

Something's wrong here.

### 18.5.7 Solution to the Paradox

The argument claiming the expectation is zero against a fair wheel is flawed by an implicit, invalid use of linearity of expectation for an infinite sum.

To explain this carefully, let $B_n$ be the number of dollars you win on your $n$th bet, where $B_n$ is defined to be zero if red comes up before the $n$th spin of the wheel.

Now the dollar amount you win in any gambling session is

$$\sum_{n=1}^{\infty} B_n,$$

and your expected win is

$$\operatorname{Ex}\left[\sum_{n=1}^{\infty} B_n\right]. \tag{18.15}$$

Moreover, since we're assuming the wheel is fair, it's true that $\operatorname{Ex}[B_n] = 0$, so

$$\sum_{n=1}^{\infty} \operatorname{Ex}[B_n] = \sum_{n=1}^{\infty} 0 = 0. \tag{18.16}$$

The flaw in the argument that you can't win is the implicit appeal to linearity of expectation to conclude that the expectation (18.15) equals the sum of expectations in (18.16). This is a case where linearity of expectation fails to hold—even though the expectation (18.15) is 10 and the sum (18.16) of expectations converges. The problem is that the expectation of the sum of the absolute values of the bets diverges, so the condition required for infinite linearity fails. In particular, under bet doubling your $n$th bet is $10 \cdot 2^{n-1}$ dollars while the probability that you will make an $n$th bet is $2^{-n}$. So

$$\operatorname{Ex}[|B_n|] = 10 \cdot 2^{n-1} 2^{-n} = 20.$$

Therefore the sum

$$\sum_{n=1}^{\infty} \operatorname{Ex}[|B_n|] = 20 + 20 + 20 + \cdots$$

diverges rapidly.

So the presumption that you can't beat a fair game, and the argument we offered to support this presumption, are mistaken: by bet doubling, you can be sure to walk away a winner. Probability theory has led to an apparently absurd conclusion.

But probability theory shouldn't be rejected because it leads to this absurd conclusion. If you only had a finite amount of money to bet with—say enough money to make $k$ bets before going bankrupt—then it would be correct to calculate your expection by summing $B_1 + B_2 + \cdots + B_k$, and your expectation would be zero for the fair wheel and negative against an unfair wheel. In other words, in order to follow the bet doubling strategy, you need to have an infinite bankroll. So it's absurd to assume you could actually follow a bet doubling strategy, and it's entirely reasonable that an absurd assumption leads to an absurd conclusion.

### 18.5.8 Expectations of Products

While the expectation of a sum is the sum of the expectations, the same is usually not true for products. For example, suppose that we roll a fair 6-sided die and denote the outcome with the random variable $R$. Does $\text{Ex}[R \cdot R] = \text{Ex}[R] \cdot \text{Ex}[R]$?

We know that $\text{Ex}[R] = 3\frac{1}{2}$ and thus $\text{Ex}[R]^2 = 12\frac{1}{4}$. Let's compute $\text{Ex}[R^2]$ to see if we get the same result.

$$\begin{aligned}\text{Ex}\left[R^2\right] &= \sum_{\omega \in \mathcal{S}} R^2(\omega) \Pr[w] = \sum_{i=1}^{6} i^2 \cdot \Pr[R_i = i] \\ &= \frac{1^2}{6} + \frac{2^2}{6} + \frac{3^2}{6} + \frac{4^2}{6} + \frac{5^2}{6} + \frac{6^2}{6} = 15\ 1/6 \neq 12\ 1/4.\end{aligned}$$

That is,

$$\text{Ex}[R \cdot R] \neq \text{Ex}[R] \cdot \text{Ex}[R].$$

So the expectation of a product is not always equal to the product of the expectations.

There is a special case when such a relationship *does* hold however; namely, when the random variables in the product are *independent*.

**Theorem 18.5.6.** *For any two* independent *random variables $R_1$, $R_2$,*

$$\text{Ex}[R_1 \cdot R_2] = \text{Ex}[R_1] \cdot \text{Ex}[R_2].$$

The proof follows by rearrangement of terms in the sum that defines $\text{Ex}[R_1 \cdot R_2]$. Details appear in Problem 18.25.

Theorem 18.5.6 extends routinely to a collection of mutually independent variables.

**Corollary 18.5.7.** *[Expectation of Independent Product]*

*If random variables $R_1, R_2, \ldots, R_k$ are mutually independent, then*

$$\text{Ex}\left[\prod_{i=1}^{k} R_i\right] = \prod_{i=1}^{k} \text{Ex}[R_i].$$

6.042J / 18.062J Mathematics for Computer Science
Spring 2015