In general, equation (18.2) is more useful than the defining equation (18.1) 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 (18.1) sometimes yields easier proofs about general properties of expectation.

Medians

The mean of a random variable is not the same as the *median*. The median is the *midpoint* of a distribution.

Definition 18.4.4. The *median* of a random variable R is the value $x \in \text{range}(R)$ such that

$$\Pr[R \leq x] \leq \frac{1}{2} \quad \text{and}$$
$$\Pr[R > x] < \frac{1}{2}.$$

We won't devote much attention to the median. The expected value is more useful and has much more interesting properties.

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

$$\text{Ex}[R \mid A] ::= \sum_{r \in \text{range}(R)} r \cdot \Pr\big[R = r \mid A\big]. \tag{18.3}$$

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 (18.3),

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

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 49.8% 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

$$\begin{aligned}
\text{Ex}[H] &= \text{Ex}[H \mid M]\,\text{Pr}[M] + \text{Ex}[H \mid F]\,\text{Pr}[F] \\
&= (5 + 11/12) \cdot 0.498 + (5 + 5/12) \cdot 0.502 \\
&= 5.665
\end{aligned}$$

which is a little less than 5' 8."

This method is justified by:

Theorem 18.4.6 (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*

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

Proof.

$$\begin{aligned}
\text{Ex}[R] &= \sum_{r \in \text{range}(R)} r \cdot \text{Pr}[R = r] && \text{(by 18.2)} \\
&= \sum_r r \cdot \sum_i \text{Pr}\left[R = r \mid A_i\right] \text{Pr}[A_i] && \text{(Law of Total Probability)} \\
&= \sum_r \sum_i r \cdot \text{Pr}\left[R = r \mid A_i\right] \text{Pr}[A_i] && \text{(distribute constant } r\text{)} \\
&= \sum_i \sum_r r \cdot \text{Pr}\left[R = r \mid A_i\right] \text{Pr}[A_i] && \text{(exchange order of summation)} \\
&= \sum_i \text{Pr}[A_i] \sum_r r \cdot \text{Pr}\left[R = r \mid A_i\right] && \text{(factor constant Pr}[A_i]\text{)} \\
&= \sum_i \text{Pr}[A_i]\,\text{Ex}[R \mid A_i]. && \text{(Def 18.4.5 of cond. expectation)}
\end{aligned}$$

■

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 18.4.6. 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 let 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

$$\text{Ex}[C] = \text{Ex}[C \mid A]\Pr[A] + \text{Ex}[C \mid \overline{A}]\Pr[\overline{A}]. \tag{18.4}$$

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

$$\text{Ex}[C \mid A] = 1. \tag{18.5}$$

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,

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

Plugging (18.5) and (18.6) into (18.4):

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

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 wants to have a baby girl. For simplicity assume there is a 50% chance that each child they have is a girl, and the genders of their children are mutually independent. If the couple insists on having children until they get a girl, then how many baby boys should they expect first?

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 girl?" In this case, a crash corresponds to having a girl, so we should set $p = 1/2$. By the preceding analysis, the couple should expect a baby girl after having $1/p = 2$ children. Since the last of these will be the girl, they should expect just one boy.

Something to think about: If every couple follows the strategy of having children until they get a girl, what will eventually happen to the fraction of girls born in this world?

Using the Law of Total Expectation to find expectations is a worthwhile approach to keep in mind, but it's good review to derive the same formula directly from the definition of expectation. Namely, the probability that the first crash occurs in the ith hour for some $i > 0$ is the probability, $(1-p)^{i-1}$, that it does not crash in each of the first $i - 1$ hours, times the probability, p, that it does crash in the ith hour. So

$$\begin{aligned} \text{Ex}[C] &= \sum_{i \in \mathbb{N}} i \cdot \Pr[C = i] && \text{(by (18.2))} \\ &= \sum_{i \in \mathbb{N}} i(1-p)^{i-1} p \\ &= \frac{p}{1-p} \cdot \sum_{i \in \mathbb{N}} i(1-p)^i. && (18.7) \end{aligned}$$

But we've already seen a sum like this last one (you did remember this, right?), namely, equation (14.13):

$$\sum_{i \in \mathbb{N}} i x^i = \frac{x}{(1-x)^2}.$$

Combining (14.13) with (18.7) gives

$$\text{Ex}[C] = \frac{p}{1-p} \cdot \frac{1-p}{(1-(1-p))^2} = \frac{1}{p}$$

as expected.

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 paramter p.

Definition 18.4.7. A random variable, C, has a *geometric distribution* with paramter p iff $\text{codomain}(C) = \mathbb{Z}^+$ and

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

Lemma 18.4.8. *If a random variable C had a geometric distribution with paramter 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.

Some gambling games are more complicated and thus more interesting. The following game where the winners split a pot is representative of many poker games, betting pools, and lotteries.

Splitting the Pot

After your last encounter with biker dude, one thing led to another and you have dropped out of school and become a Hell's Angel. It's late on a Friday night and, feeling nostalgic for the old days, you drop by your old hangout, where you encounter two of your former TAs, Eric and Nick. Eric and Nick propose that you join them in a simple wager. Each player will put $2 on the bar and secretly write "heads" or "tails" on their napkin. Then one player will flip a fair coin. The $6 on the bar will then be divided equally among the players who correctly predicted the outcome of the coin toss.

After your life-altering encounter with strange dice, you are more than a little skeptical. So Eric and Nick agree to let you be the one to flip the coin. This certainly seems fair. How can you lose?