Massachusetts Institute of Technology | Lecture 26
6.042J/18.062J: Mathematics for Computer Science | 9 Dec 97
Professor Tom Leighton

Lecture Notes

The topic of this lecture is random walks. The abstract idea of a random walk is that there is some value that randomly goes up, goes down, or stays the same over a sequence of time steps. Many natural phenomena are nicely modeled by random walks.

However, for some reason, random walks are traditionally discussed in the context of some social vice. For example, the value is often regarded as the position of a drunkard who randomly staggers left, staggers right, or just wobbles in place during each time step.

In this lecture, we will discuss random walks in the context of gambling. In this case, the value is a gambler's cash-on-hand. This value goes up or down depending on whether he wins or loses each bet.

This lecture makes extensive use of linear recurrences, which are discussed in Lecture 11.

1 The Gambler's Ruin

A gambler starts with n dollars. He makes a sequence of \$1 bets. In each bet, he wins with probability p and loses with probability $1 - p$. The gambler plays until he is either bankrupt or is m dollars ahead. What is the probability that the gambler makes a profit?

A few naming conventions are used throughout the lecture. The amount of money that the gambler has at any time is called his *capital*. Let n be the gambler's starting capital, let m be his desired profit, and let $T = n + m$ be his total capital when he achieves his desired profit. If the gambler's capital reaches T dollars, then he is called a *winner*. If the gambler's capital reaches zero dollars, then we say that he *goes broke*.

For example, suppose that the gambler is playing roulette, and always bets red or black. In this game, the probability of winning is $p = \frac{18}{38} \approx 0.47$, since a roulette wheel has 18 black numbers, 18 red numbers, and 2 uncolored numbers. Suppose that the gambler's starting capital is $n = 500$ dollars and he wants to make $m = 100$ dollars. That is, he plays until he goes broke or has $T = n + m = \$600$. What is the probability that he is a winner?

Since the game is almost fair, one might expect that the gambler's chances are at least better than even. In fact, we will see that his odds are less than 1 in 37,000!

1.1 Gambler's Ruin as a Random Walk

In a *random walk* there is some value that randomly goes up, goes down, or stays the same during each time step. If the value is as likely to go up as to go down, then the random walk is called *unbiased*. If one direction or the other is more likely, then the random walk is said to be *biased*.

We can model the Gambler's Ruin problem as a random walk. In this case, the value is the gambler's capital. During each time step, the value goes up by 1 with probability p (gambler wins a bet) and goes down by 1 with probability $1-p$ (gambler loses a bet). In this case, the value never stays the same during a time step. This situation is shown graphically in Figure 1.

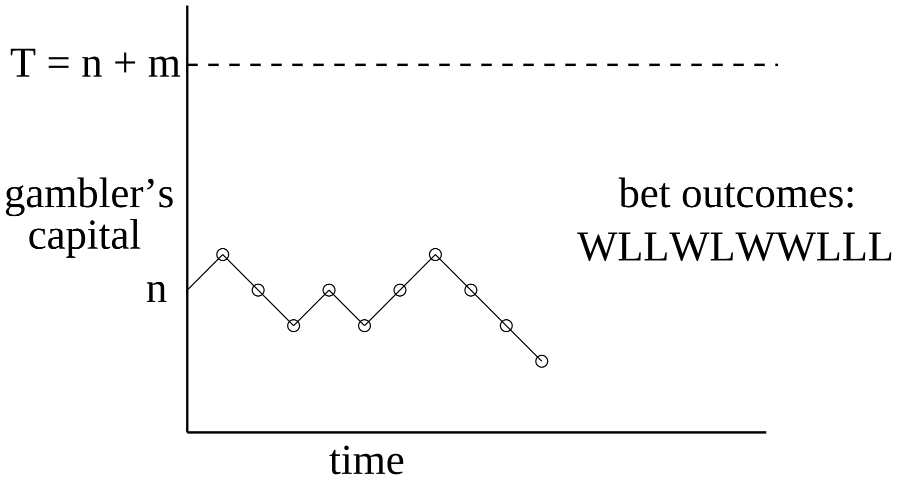

Figure 1: *This is a graph of the gambler's capital versus time for one possible sequence of bet outcomes. At each time step, the graph goes up with probability p and down with probability* $1-p$. *The gambler continues betting until the graph reaches either 0 or* $T = n + m$.

If the gambler plays a fair game then $p = 1 - p = \frac{1}{2}$, and so the corresponding random walk is *unbiased*. Otherwise, the gambler is either more likely to win or more likely to lose each bet, and so the corresponding random walk is *biased*.

A random walk is a *martingale* if each value change is independent of all past changes. For example, the Gambler's Ruin random walk is a martingale because the outcome of the current bet is independent of the outcomes of all preceding bets. On the other hand, if the casino cheats against customers who win too often, then the problem is no longer described by a martingale; the outcome of the current bet would depend on whether it was preceded by a long sequence of wins that irked the management.

In the Gambler's Ruin problem, the random walk has boundaries at 0 and $T = n + m$. If the random walk ever reaches either of these boundary values, then it terminates. The problem is to determine the probability that the walk terminates at boundary T, in which case the gambler is a winner.

2 Solving the Gambler's Ruin Problem

The sample space for the Gambler's Ruin problem consists of all strings of W's and L's, such as $WLLWLWWLLL\ldots$. Each string indicates a possible sequence of won and lost bets. As in earlier problems where the sample space consisted of infinite strings, we can not assign a probability to each sample point, but can still compute event probabilities.

A direct approach to the Gambler's Ruin problem is to tally up all strings of W's and L's that make the gambler a winner. For example, the gambler is a winner for all strings starting with m consecutive W's. This approach works, but we will instead solve the problem using probability laws and recurrence techniques developed over the last few weeks.

2.1 A Recurrence for the Probability of Winning

The following claim gives a recurrence for the probability that the gambler is a winner if he starts with n dollars.

Claim 2.1 If the gambler starts with n dollars and wins each bet with probability p, then the probability that he reaches $T = n + m$ dollars before going broke is P_n, which is defined by the following recurrence.

$$P_n = \begin{cases} 0 & \text{if } n = 0 \\ 1 & \text{if } n = T \\ pP_{n+1} + (1-p)P_{n-1} & \text{if } 0 < n < T \end{cases}$$

If this claim is correct, then we can solve the recurrence to find the probability that the gambler is a winner.

The claim *looks* correct. If the gambler starts with 0 dollars, then he is already broke, so $P_0 = 0$. Similarly, if he starts with T dollars, then he is immediately a winner, so $P_T = 1$. Otherwise, the gambler starts with n dollars, where $0 < n < T$. Consider the outcome of his first bet. The gambler wins the first bet with probability p. In this case, he is left with $n+1$ dollars and becomes a winner with probability P_{n+1}. On the other hand, he loses the first bet with probability $1-p$. Now he is left with $n-1$ dollars and becomes a winner with probability P_{n-1}. Overall, he is a winner with probability $pP_{n+1} + (1-p)P_{n-1}$ as claimed.

This informal argument is probably more convincing than the actual proof, which is rather intricate. In particular, the proof uses an identity which can be confusing when many events and random variables are floating around. From the definition of conditional probability, we have $\Pr(A \cap B) = \Pr(B)\Pr(A \mid B)$ where A and B are events. Conditioning on another event C gives the identity:

$$\Pr(A \cap B \mid C) = \Pr(B \mid C)\Pr(A \mid B \cap C)$$

With this tool in hand, we are ready to prove Claim 2.1.

Proof. If $n = 0$, then the gambler is immediately bankrupt, so $P_0 = 0$. If $n = T$, then the gambler is immediately a winner, so $P_T = 1$. In the remaining case, we have $0 < n < T$. Define the following events and random variable:

- Let E be the event that the gambler wins the first bet.
- Let W be the event that the gambler is a winner.

- Let the random variable D_i be the gambler's capital after i bets.

We start with the observation that P_n is defined as the probability that the gambler is a winner, given that his initial capital is n dollars, and reason as follows.

$$\begin{aligned}
P_n &= \Pr(W \mid D_0 = n) \\
&= \Pr(W \cap E \mid D_0 = n) + \Pr(W \cap \bar{E} \mid D_0 = n) \\
&= \Pr(E \mid D_0 = n)\Pr(W \mid E \cap D_0 = n) + \Pr(\bar{E} \mid D_0 = n)\Pr(W \mid \bar{E} \cap D_0 = n) \\
&= p\Pr(W \mid E \cap D_0 = n) + (1-p)\Pr(W \mid \bar{E} \cap D_0 = n) \\
&= p\Pr(W \mid D_1 = n+1) + (1-p)\Pr(W \mid D_1 = n-1)
\end{aligned}$$

The second equality uses the theorem of total probability. In the third step, we apply the identity described above to both terms. In the fourth step, we use the fact that the gambler wins the first bet with probability p, regardless of his initial capital D_0. In the last step, we observe that the event $E \cap D_0 = n$ is the same as the event $D_1 = n+1$; that is, if the gambler wins his first bet and started with n dollars, then he has $n+1$ dollars after the first bet. We apply a similar observation in the second term.

By assumption, the outcome of a bet is independent of the outcomes of all previous bets. Therefore, a gambler with $n+1$ dollars after his first best is in the same position as a gambler just starting with $n+1$ dollars. As a consequence, the events $W \mid D_1 = n+1$ and $W \mid D_0 = n+1$ have equal probability. (These are, however, distinct events.) We can use this fact to continue simplifying the above expression for P_n as follows.

$$\begin{aligned}
&= p\Pr(W \mid D_0 = n+1) + (1-p)\Pr(W \mid D_0 = n-1) \\
&= pP_{n+1} + (1-p)P_{n-1}
\end{aligned}$$

The second equation uses the definition of P_n. This establishes the recurrence. ■

2.2 Solving the Recurrence

We now have a recurrence for the probability that a gambler starting with n dollars is a winner. Solving the Gambler's Ruin problem just requires solving this recurrence, which is restated below.

$$P_n = pP_{n+1} + (1-p)P_{n-1} \qquad \text{where } P_0 = 0 \text{ and } P_T = 1$$

We have seen many recurrences where the n-th term depends on preceding terms. However, here P_n depends on the term before *and* the term after! This is really only a superficial problem; looked at another way, the recurrence gives P_{n+1} as a function of the two preceding terms. Moving everything to one side gives:

$$pP_{n+1} - P_n + (1-p)P_{n-1} = 0 \qquad \text{where } P_0 = 0 \text{ and } P_T = 1$$

This is now a regular linear homogeneous recurrence. The only snag is that the highest index is $n+1$ instead of the usual n. We could correct this problem by substituting $k = n+1$. However, this turns out to be unnecessary. Recall that the first step in solving a linear recurrence is to find the characteristic equation. And regardless of the indexing scheme, the characteristic equation of this recurrence is:

$$pr^2 - r + (1-p) = 0$$

One surprising fact about the Gambler's Ruin problem is the enormous difference between an almost-fair game and a perfectly fair game. Looking at the roots of the characteristic equation makes this discontinuity apparent. We can find the roots as shown below.

$$\begin{aligned} r &= \frac{1 \pm \sqrt{1-4p(1-p)}}{2p} \\ &= \frac{1 \pm (1-2p)}{2p} \\ &= \frac{1-p}{p} \text{ or } 1 \end{aligned}$$

If $p \neq \frac{1}{2}$, then the characteristic equation has two distinct roots. This corresponds to an unfair gambling game, modeled by a biased random walk. This case is considered in Section 2.3. However, if $p = \frac{1}{2}$, then the characteristic equation has a double root at $r = 1$. This means that the recurrence has a completely different form in the case of a fair game! We address this case in Section 2.4.

2.3 An Unfair Game

We first consider the case of an unfair game. In this case, the characteristic equation has two distinct roots, $\frac{1-p}{p}$ and 1. Therefore, the solution to the recurrence has the following general form:

$$P_n = A\left(\frac{1-p}{p}\right)^n + B(1)^n$$

We can find constants A and B from the boundary conditions, $P_0 = 0$ and $P_T = 1$. Substituting these two conditions into the general form gives a system of linear equations.

$$\begin{aligned} 0 &= A + B \\ 1 &= A\left(\frac{1-p}{p}\right)^T + B \end{aligned}$$

The constants A and B are found by solving the linear system.

$$\begin{aligned} A &= \frac{1}{\left(\frac{1-p}{p}\right)^T - 1} \\ B &= -\frac{1}{\left(\frac{1-p}{p}\right)^T - 1} \end{aligned}$$

Substituting these values for A and B into the general form gives the solution to the recurrence.

$$P_n = \frac{1}{\left(\frac{1-p}{p}\right)^T - 1} \cdot \left(\frac{1-p}{p}\right)^n - \frac{1}{\left(\frac{1-p}{p}\right)^T - 1} \cdot (1)^n = \frac{\left(\frac{1-p}{p}\right)^n - 1}{\left(\frac{1-p}{p}\right)^T - 1}$$

We now have a solution to the Gambler's Ruin problem. If the gambler starts with n dollars, then he goes home a winner with probability P_n, which is given by the expression above.

A Simpler Upper Bound

The exact expression for P_n is hard to interpret. There is a simpler upper bound which is nearly tight when the gambler's starting capital is large. Suppose that $p < \frac{1}{2}$; that is, the game is biased *against* the gambler. Then both the numerator and denominator in the expression for P_n are positive and the ratio is less than one. Under these conditions, adding 1 to both the numerator and denominator gives an upper bound as follows.

$$\begin{aligned} P_n &< \frac{\left(\frac{1-p}{p}\right)^n}{\left(\frac{1-p}{p}\right)^T} \\ &= \left(\frac{p}{1-p}\right)^{T-n} \\ &= \left(\frac{p}{1-p}\right)^m \end{aligned}$$

This shows that the gambler goes up m dollars before going broke with probability at most $\left(\frac{p}{1-p}\right)^m$.

Examples

Suppose that the gambler starts with \$500 and plays red-black in roulette until he is up \$100 or goes broke. The probability that he is a winner is at most:

$$P_{500} < \left(\frac{\frac{18}{38}}{1-\frac{18}{38}}\right)^{100} = \left(\frac{9}{10}\right)^{100} < \frac{1}{37,648}$$

The amazing thing is that the upper bound on P_n does not depend on n, the gambler's starting capital. That is, if the gambler starts with a trillion dollars, then his chance of going up a mere \$100 before losing the whole trillion is still less than 1 in 37,648!

Furthermore, the bound is exponential in m. Regardless of his starting capital, the probability that the gambler is ever up $m = 1000$ dollars before he goes broke is at most $\left(\frac{9}{10}\right)^{1000} < 10^{-45}$.

The odds of winning a little money are not so bad. Applying the exact formula for P_n, we can find the probability of winning \$10 before losing \$10.

$$P_{10} = \frac{\left(\frac{1-\frac{18}{38}}{\frac{18}{38}}\right)^{10} - 1}{\left(\frac{1-\frac{18}{38}}{\frac{18}{38}}\right)^{20} - 1} = 0.2585\ldots$$

We will give some intuition for why the gambler generally has such rotten odds in Section 2.5.

2.4 A Fair Game

Now suppose that the gambler plays a fair game. In this case, the characteristic equation has a double root at 1. This means that the solution to the recurrence has the following general form:

$$\begin{aligned} P_n &= An(1)^n + B(1)^n \\ &= An + B \end{aligned}$$

To find the constants A and B, we substitute the boundary conditions, $P_0 = 0$ and $P_T = 1$, into the general form. This gives a system of two linear equations in two unknowns.

$$\begin{aligned} 0 &= B \\ 1 &= AT + B \end{aligned}$$

Solving this system of equations, we conclude that $A = \frac{1}{T}$ and $B = 0$. The solution to the recurrence is obtained by substituting these values back into the general form.

$$\begin{aligned} P_n &= \frac{n}{T} \\ &= \frac{n}{n+m} \end{aligned}$$

This means that if the gambler starts with n dollars and bets until he wins m dollars or goes broke, then he is a winner with probability $\frac{n}{n+m}$.

Examples

Suppose that the gambler again starts with \$500 and plays a fair game until he is up \$100 or goes broke. Now he is a winner with probability $\frac{500}{500+100} = \frac{5}{6}$. This is much better than the 1 in 37,648 probability that he faced in a slightly unfair game!

Suppose that the gambler starts with a trillion dollars and tries to make \$1000. Earlier we saw that if the game is slightly unfair, then the probability that he is a winner is less than 10^{-45}. However, by the above formula, if the game is fair, then the probability that he is *not* at winner is now less than 10^{-9}!

Similarly, the probability of winning \$10 before losing \$10 is now $\frac{10}{10+10} = \frac{1}{2}$. This is somewhat better than the 1 in 4 chance in the unfair game, but not greatly so.

2.5 Intuition

Why is the gambler so unlikely to make money when the game is slightly biased against him? Intuitively, there are two forces at work. First, the gambler's capital has random upward and downward *swings* due to runs of good and bad luck. Second, the gambler's capital will have a steady, downward *drift*, because he has a small, negative expected return on every bet. The situation is shown in Figure 2.

For example, in roulette the gambler wins a dollar with probability p and loses a dollar with probability $(1-p)$. Therefore, his expected return on each bet is $1 \cdot p + (-1) \cdot (1-p) = 2p - 1 = -\frac{1}{19} \approx -0.053$. That is, on each bet his capital is expect to drift downward by a little over 5 cents.

Our intuition is that if the gambler starts with a trillion dollars, then he will play for a very long time, so at some point there should be a lucky, upward swing that puts him \$100 ahead. The problem is that his capital is steadily drifting downward. If the gambler does not have a lucky, upward swing early on, then he is doomed. After his capital drifts downward a few hundred dollars, he needs a huge upward swing to save himself. And such a huge swing is extremely improbable. As a rule of thumb, *drift* dominates *swings* in the long term.

In a fair game, there is no drift; swings are the only effect. In the absence of downward drift, our earlier intuition is correct. If the gambler starts with a trillion dollars then almost certainly there will eventually be a lucky swing that puts him \$100 ahead.

If we start with \$10 and play to win only \$10 more, then the difference between the fair and unfair games is relatively small. (The probability of winning is $\frac{1}{2}$ versus about $\frac{1}{4}$.) Since swings of \$10 are relatively common, the game usually ends before the gambler's capital can drift very far. That is, the game does not last long enough for drift to dominate the swings.

3 Duration of the Gambling Game

We now know the probability that the gambler is a winner in both fair and unfair games. In this section, we address a related question; namely, how many bets does the gambler need on average to either win or go broke? A preliminary question is, what is the probability that the gambler *never* wins or goes broke?

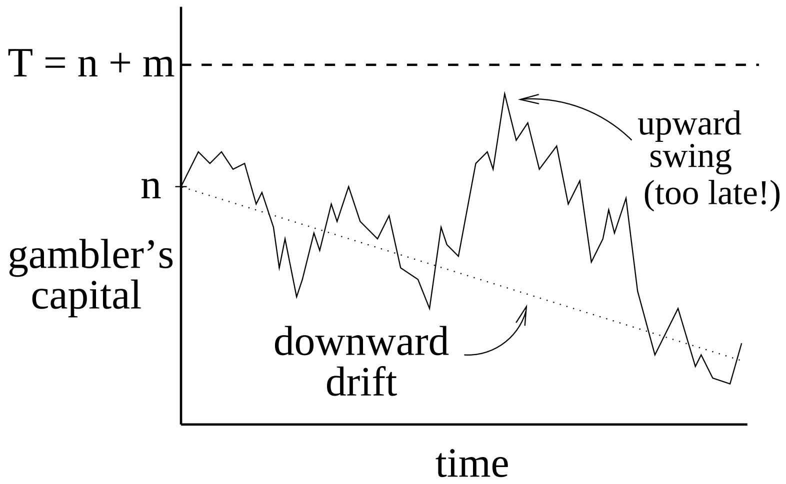

Figure 2: *In an unfair game, the gambler's capital swings randomly up and down, but steadily drifts downward. If the gambler does not have a winning swing early on, then his capital drifts downward, and later upward swings are insufficient to make him a winner.*

3.1 Probabilty that the Game Never Terminates

In an unfair game, we know that the gambler will almost certainly not be a winner. This seems to imply that he will almost surely go broke. However, there is a third possibility— he may go on gambling forever! For example, in the outcome $WLWLWLWLWL\ldots$, the gambler's capital fluctuates between n and $n+1$ forever. In light of this example, the following theorem is surprising.

Theorem 3.1 The event that the gambling game lasts forever has probability 0 for all $0 \leq p \leq 1$.

Even though the game can go on forever, the probability of this event is zero! Strange phenomena like this crop up when the sample space is uncountable. These matters are sorted out in an area of mathematics called measure theory. There is a little intuition here, however. In some sense, every individual sample point (such as $WLWLWLWLWL\ldots$) has probability zero. Not surprisingly, an event also has probability zero, provided it consists of "not too many" sample points. For example, an event has zero probability if it consists of a finite number or even a countable infinity of sample points. Now in this case, the gambling game last forever for an uncountably infinite number of sample points. However, this still turns out to be "not too many", and so the game lasts forever with probability zero.

The proof of the theorem requires no unfamiliar math.

Proof. If $p = 0$, then the gambler loses every bet, and the gambling game terminates when he is bankrupt after n bets. If $p = 1$, then the gambler wins every bet, and the game ends after he makes m dollars. Therefore, we can assume hereafter that p is strictly between 0 and 1.

Let Q_n be the probability that the game lasts forever given that the gambler starts with n dollars. We can express Q_n recursively as follows:

$$Q_n = \begin{cases} 0 & \text{if } n = 0 \\ 0 & \text{if } n = T \\ pQ_{n+1} + (1-p)Q_{n-1} & \text{if } 0 < n < T \end{cases}$$

(The proof for this recurrence is the same as in Claim 2.1. The only difference is that $Q_T = 0$, whereas $P_T = 1$. As a result, the characteristic equation is the same, and we just have to plug in the new boundary condition.)

First, we consider an unfair game. If $p \neq \frac{1}{2}$, then the solution to the recurrence has the following general form:

$$Q_n = A\left(\frac{1-p}{p}\right)^n + B$$

We find constants the A and B by substituting the boundary conditions, $Q_0 = 0$ and $Q_T = 0$, into this general form.

$$\begin{aligned} 0 &= A + B \\ 0 &= A\left(\frac{1-p}{p}\right)^T + B \end{aligned}$$

Solving the linear system, we find that $A = B = 0$. This implies that $Q_n = 0$ for all n. Therefore, the probability that an unfair gambling game lasts forever is 0.

(This also shows that a small change in a boundary condition can radically alter the solution to a linear recurrence. Recall that Q_n and P_n are defined by the same recurrence equation. They differ only in a boundary condition; we have $Q_T = 0$, but $P_T = 1$. This difference in the boundary condition greatly altered the solution to the recurrence. While $Q_n = 0$, we have $P_n = \frac{((1-p)/p)^n - 1}{((1-p)/p)^T - 1}$.)

Now we consider a fair game. If $p = \frac{1}{2}$, then the solution to the recurrence has the following general form.

$$Q_n = An + B$$

Again, we find A and B by substituting in the boundary conditions.

$$\begin{aligned} 0 &= B \\ 0 &= AT + B \end{aligned}$$

Once more, solving the linear system shows that $A = B = 0$. This implies that $Q_n = 0$ for all n, and so the probability that a fair gambling game lasts forever is also 0. ■

3.2 Expected Duration of a Game

We have shown that there is zero probability that the gambling game lasts forever, but what is the expected duration? We can answer this question by setting up and solving a recurrence as in the last section. The key difference is that the new recurrence involves expectations instead of probabilities.

Claim 3.2 If the gambler starts with n dollars and wins each bet with probability p, then the expected number of bets before he reaches either 0 dollars or $T \geq n$ dollars is E_n, which is defined by the recurrence below.

$$E_n = \begin{cases} 0 & \text{if } n = 0 \\ 0 & \text{if } n = T \\ 1 + pE_{n+1} + (1-p)E_{n-1} & \text{if } 0 < n < T \end{cases}$$

As with Claim 2.1, the intuition here is simple and the proof is a little hairy. If the gambler starts with 0 or T dollars, then the game ends immediately, since he is either broke or a winner already. Otherwise, the gambler makes his first bet. If he wins (probability p), then he has $n+1$ dollars, and the game is expected to continue for another E_{n+1} bets. If he loses (probability $1-p$), then he has $n-1$ dollars, and the game is expected to continue for another E_{n-1} bets. Overall, the expected number of bets is $1 + pE_{n+1} + (1-p)E_{n-1}$.

Proof. If the gambler starts with 0 or T dollars, then the game ends immediately. Therefore, $E_0 = E_T = 0$ as claimed. In the remaining case, we have $0 < n < T$. Define the following event and random variables:

- Let E be the event that the gambler wins the first bet.
- Let the random variable S be number of bets in the game.
- Let the random variable D_i be the gambler's capital after i bets.

Since E_n is defined as the expected number of bets, given that the gambler starts with n dollars, we have:

$$\begin{aligned} E_n &= \text{Ex}(S \mid D_0 = n) \\ &= \text{Ex}(S \mid D_0 = n \cap E)\Pr(E) + \text{Ex}(S \mid D_0 = n \cap \bar{E})\Pr(\bar{E}) \\ &= \text{Ex}(S \mid D_0 = n \cap E)p + \text{Ex}(S \mid D_0 = n \cap \bar{E})(1-p) \\ &= \text{Ex}(S \mid D_1 = n+1)p + \text{Ex}(S \mid D_1 = n-1)(1-p) \end{aligned}$$

The second step uses the analogue to the Theorem of Total Probability that was introduced in Tutorial 13. The third step uses the fact that the gambler wins the first bet with probability p.

The last step uses the observation that if the gambler starts with n dollars and wins his first bet, then he has $n+1$ dollars after the first bet. A similar observation is used for the second term.

By assumption, the outcome of a bet is independent of all previous outcomes. Therefore, a gambler with $n+1$ dollars after the first bet is in the same position as a gambler who started with $n+1$ dollars, except that he has already made one bet. With this in mind, we can continue rewriting E_n as follows.

$$\begin{aligned}
&= (1 + \mathrm{Ex}(S \mid D_0 = n+1))p + (1 + \mathrm{Ex}(S \mid D_0 = n-1))(1-p) \\
&= (1 + E_{n+1})p + (1 + E_{n-1})(1-p) \\
&= 1 + pE_{n+1} + (1-p)E_{n-1}
\end{aligned}$$

The second step uses the definition of E_n, and the last step is simplification. ■

3.3 Duration of an Unfair Game

We can find the expected number of bets in the gambling game by solving the recurrence in Claim 3.2. Rearranging terms, the recurrence is:

$$pE_{n+1} - E_n + (1-p)E_{n-1} = -1 \qquad \text{where } E_0 = E_T = 0$$

This is the same recurrence as for Q_n, except that there is a -1 on the right side. This makes the recurrence inhomogeneous. The solution to an inhomogeneous linear recurrence is the sum of the homogeneous solution and a particular inhomogeneous solution. We now know that the homogeneous solution for this recurrence is radically different for fair and unfair games. In this section, we analyze the duration of an unfair game ($p \neq \frac{1}{2}$). The next section considers the duration of a fair game ($p = \frac{1}{2}$).

By solving the characteristic equation, we found that for $p \neq \frac{1}{2}$ the homogeneous solution has the following form.

$$E_n = A\left(\frac{1-p}{p}\right)^n + B$$

Since the inhomogeneous term is a constant, we guess a particular solution that is also a constant, say a. Substituting $E_n = a$ into the recurrence gives:

$$\begin{aligned}
pa - a + (1-p)a &= -1 \\
a(p - 1 + 1 - p) &= -1 \\
0 &= -1
\end{aligned}$$

This does not look entirely promising. Let's try a particular solution of higher degree, $an + b$.

$$\begin{aligned}
p(a(n+1)+b) - (an+b) + (1-p)(a(n-1)+b) &= -1 \\
pan + pa + pb - an - b + an - a + b - pan + pa - pb &= -1 \\
2pa - a &= -1 \\
a &= \frac{1}{1-2p}
\end{aligned}$$

The recurrence has a particular solution of the form $an + b$ when $a = \frac{1}{1-2p}$ and b is anything. Setting $b = 0$ gives the particular solution $E_n = \frac{n}{1-2p}$. Adding the homogeneous solution and the particular solution gives a complete solution of the form:

$$E_n = A\left(\frac{1-p}{p}\right)^n + B + \frac{n}{1-2p}$$

Substituting in the boundary conditions, $E_0 = E_T = 0$, gives a system of linear equations.

$$\begin{aligned}
0 &= A + B \\
0 &= A\left(\frac{1-p}{p}\right)^T + B + \frac{T}{1-2p}
\end{aligned}$$

Now we find constants A and B by solving the linear system.

$$A = -\frac{\frac{T}{1-2p}}{\left(\frac{1-p}{p}\right)^T - 1}$$

$$B = \frac{\frac{T}{1-2p}}{\left(\frac{1-p}{p}\right)^T - 1}$$

Substituting the values of constants A and B into the general form gives the solution to the recurrence.

$$\begin{aligned}
E_n &= -\frac{\frac{T}{1-2p}}{\left(\frac{1-p}{p}\right)^T - 1}\left(\frac{1-p}{p}\right)^n + \frac{\frac{T}{1-2p}}{\left(\frac{1-p}{p}\right)^T - 1} + \frac{n}{1-2p} \\
&= \frac{n}{1-2p} - \frac{T}{1-2p} \cdot \frac{\left(\frac{1-p}{p}\right)^n - 1}{\left(\frac{1-p}{p}\right)^T - 1}
\end{aligned}$$

Sure, this isn't something to trot out on a first date, but this ugly formula does lead to some nice conclusions.

If the game is biased against the gambler ($p < \frac{1}{2}$), then the second term is always positive. Therefore, $E_n \leq \frac{n}{1-2p}$; that is, if the gambler starts with n dollars, then the expected number of bets is at most $\frac{n}{1-2p}$. This makes some intuitive sense. On every bet, the gambler's expected loss is $1-2p$. Therefore, the gambler should need about $\frac{n}{1-2p}$ bets to lose his initial capital of n dollars.

The actual formula for E_n is complicated by the fact that the gambling game also ends if the gambler wins by reaching T dollars. We can remove this complication by making T very large; in fact, we can take the limit as T goes to infinity. This corresponds the situation where the gambler never quits while he is ahead, but keeps playing until he is broke. In the limit, the second term of E_n goes to zero because T appears as an exponent in the denominator. Therefore, the expected number of bets approaches exactly $\frac{n}{1-2p}$, just as one would think.

3.4 Duration of a Fair Game

Now we will find the expected duration of a fair gambling game ($p = \frac{1}{2}$). In this case, the characteristic equation has a double root at 1. As we saw earlier, the homogeneous solution has the general form shown below.

$$E_n = An + B$$

For the particular solution, we guess $E_n = an + b$ as before. Substituting this guess into the recurrence gives:

$$\begin{aligned}
\frac{1}{2}(a(n+1)+b) - (an+b) + (1-\frac{1}{2})(a(n-1)+b) &= -1 \\
0 &= -1
\end{aligned}$$

Since this flopped, we try adding a higher order order term. Substituting $E_n = an^2 + bn + c$ gives a particular solution with $a = -1$ and $b = c = 0$. Therefore, the complete solution to the recurrence has the following general form.

$$E_n = An + B - n^2$$

As usual, we plug in the boundary conditions, $E_0 = E_T = 0$, to get a system of linear equations.

$$\begin{aligned}
0 &= B \\
0 &= AT + B - T^2
\end{aligned}$$

Solving the system, we find that $A = T$ and $B = 0$. Substituting these values into the general form gives the complete solution to the recurrence.

$$E_n = Tn - n^2$$

We can rewrite this solution using the definition $T = n + m$.

$$\begin{aligned} E_n &= (n+m)n - n^2 \\ &= nm \end{aligned}$$

This is a very simple answer. If the gambler starts with n dollars and plays until he is broke or ahead m dollars, then he can expect to make nm bets. For example, if he starts with \$10 dollars and plays until he is broke or ahead \$10, then 100 bets are required on average.

3.5 Quit While You Are Ahead

Suppose that the gambler never quits while he is ahead. That is, he starts with n dollars, plays a fair game, and only quits when he is flat broke. Now what is the expected duration of the game?

Claim 3.3 If the gambler starts with n dollars and plays a fair game until he is broke, then he can expect to play forever.

Proof. We defined E_n to be the expected number of bets until the gambler is ahead m dollars or goes broke, given that he starts with n dollars. Let E_n^* be the expected number of bets until the gambler goes broke, given that he starts with n dollars. (Whether the gambler is ever ahead m dollars is irrelevant in the definition of E_n^*.) Clearly, $E_n^* > E_n = nm$ for all m. Since we can make m arbitrarily large, this implies that E_n^* is infinity for all $n > 0$. Therefore, the gambler can expect to play forever. ∎

Even if the gambler starts with just one dollar, he can expect to play forever! Of course, this does not mean that he will always play forever. For example, he might lose the very first bet and be done right away. More remarkably, we have the following result.

Claim 3.4 If the gambler starts with n dollars and plays a fair game until he is broke, then he will go broke with probability 1.

Even if the gambler starts with a million dollars and plays a perfectly fair game, he will eventually lose it all with probability 1.

The last two claims are paradoxical; the first says that the gambler can expect to play forever and the second says that with probability 1 he will go broke. These statements may sound contradictory, but actually they are not.

Proof. Let T be any number of dollars greater than n. We can partition the sample space into three disjoint events:

1. The gambler holds T dollars at least once.
2. The gambler never holds T dollars and goes broke.
3. The gambler never holds T dollars and never goes broke.

We can lower bound the probability that the gambler goes broke as follows.

$$\begin{aligned}
\Pr(\text{gambler goes broke}) &\geq \Pr(\text{event 2}) \\
&= 1 - \Pr(\text{event 1}) \\
&= 1 - \frac{n}{T}
\end{aligned}$$

The first inequality holds because the second event is a subset of the event that the gambler goes broke. In step two, we use the fact that the probabilities of the first two events must sum to 1, since we proved that the third event has probability 0 in Theorem 3.1.

Since T can be any number greater than n, we can lower bound the probability that the gambler goes broke by values arbitrarily close to 1. Therefore, the gambler must go broke with probability 1.

■

Interestingly, if the gambler plays a game that is biased slightly in his favor, then there is actually a non-zero probability that he plays forever and never goes broke.