Chapter 12

Random Walks

12.1 Random Walks in Euclidean Space

In the last several chapters, we have studied sums of random variables with the goal being to describe the distribution and density functions of the sum. In this chapter, we shall look at sums of discrete random variables from a different perspective. We shall be concerned with properties which can be associated with the sequence of partial sums, such as the number of sign changes of this sequence, the number of terms in the sequence which equal 0, and the expected size of the maximum term in the sequence.

We begin with the following definition.

Definition 12.1 Let $\{X_k\}_{k=1}^\infty$ be a sequence of independent, identically distributed discrete random variables. For each positive integer n, we let S_n denote the sum $X_1 + X_2 + \cdots + X_n$. The sequence $\{S_n\}_{n=1}^\infty$ is called a *random walk*. If the common range of the X_k's is $\mathbf{R}^m$, then we say that $\{S_n\}$ is a random walk in $\mathbf{R}^m$. □

We view the sequence of X_k's as being the outcomes of independent experiments. Since the X_k's are independent, the probability of any particular (finite) sequence of outcomes can be obtained by multiplying the probabilities that each X_k takes on the specified value in the sequence. Of course, these individual probabilities are given by the common distribution of the X_k's. We will typically be interested in finding probabilities for events involving the related sequence of S_n's. Such events can be described in terms of the X_k's, so their probabilities can be calculated using the above idea.

There are several ways to visualize a random walk. One can imagine that a particle is placed at the origin in $\mathbf{R}^m$ at time $n = 0$. The sum S_n represents the position of the particle at the end of n seconds. Thus, in the time interval $[n-1, n]$, the particle moves (or jumps) from position S_{n-1} to S_n. The vector representing this motion is just $S_n - S_{n-1}$, which equals X_n. This means that in a random walk, the jumps are independent and identically distributed. If $m = 1$, for example, then one can imagine a particle on the real line that starts at the origin, and at the end of each second, jumps one unit to the right or the left, with probabilities given by the distribution of the X_k's. If $m = 2$, one can visualize the process as taking place in a city in which the streets form square city blocks. A person starts at one corner (i.e., at an intersection of two streets) and goes in one of the four possible directions according to the distribution of the X_k's. If $m = 3$, one might imagine being in a jungle gym, where one is free to move in any one of six directions (left, right, forward, backward, up, and down). Once again, the probabilities of these movements are given by the distribution of the X_k's.

Another model of a random walk (used mostly in the case where the range is $\mathbf{R}^1$) is a game, involving two people, which consists of a sequence of independent, identically distributed moves. The sum S_n represents the score of the first person, say, after n moves, with the assumption that the score of the second person is $-S_n$. For example, two people might be flipping coins, with a match or non-match representing $+1$ or -1, respectively, for the first player. Or, perhaps one coin is being flipped, with a head or tail representing $+1$ or -1, respectively, for the first player.

Random Walks on the Real Line

We shall first consider the simplest non-trivial case of a random walk in $\mathbf{R}^1$, namely the case where the common distribution function of the random variables X_n is given by

$$f_X(x) = \begin{cases} 1/2, & \text{if } x = \pm 1, \\ 0, & \text{otherwise.} \end{cases}$$

This situation corresponds to a fair coin being flipped, with S_n representing the number of heads minus the number of tails which occur in the first n flips. We note that in this situation, all paths of length n have the same probability, namely 2^{-n}.

It is sometimes instructive to represent a random walk as a polygonal line, or path, in the plane, where the horizontal axis represents time and the vertical axis represents the value of S_n. Given a sequence $\{S_n\}$ of partial sums, we first plot the points (n, S_n), and then for each $k < n$, we connect (k, S_k) and $(k+1, S_{k+1})$ with a straight line segment. The *length* of a path is just the difference in the time values of the beginning and ending points on the path. The reader is referred to Figure 12.1. This figure, and the process it illustrates, are identical with the example, given in Chapter 1, of two people playing heads or tails.

Returns and First Returns

We say that an *equalization* has occurred, or there is a *return to the origin* at time n, if $S_n = 0$. We note that this can only occur if n is an even integer. To calculate the probability of an equalization at time $2m$, we need only count the number of paths of length $2m$ which begin and end at the origin. The number of such paths is clearly

$$\binom{2m}{m} .$$

Since each path has probability 2^{-2m}, we have the following theorem.

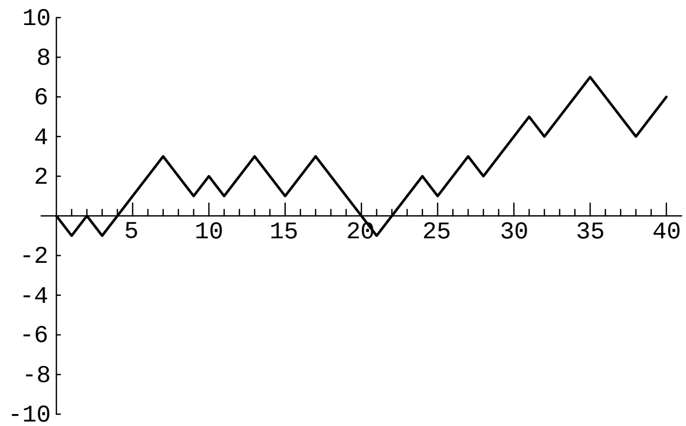

Figure 12.1: A random walk of length 40.

Theorem 12.1 The probability of a return to the origin at time $2m$ is given by

$$u_{2m} = \binom{2m}{m} 2^{-2m} .$$

The probability of a return to the origin at an odd time is 0. □

A random walk is said to have a *first return* to the origin at time $2m$ if $m > 0$, and $S_{2k} \neq 0$ for all $k < m$. In Figure 12.1, the first return occurs at time 2. We define f_{2m} to be the probability of this event. (We also define $f_0 = 0$.) One can think of the expression $f_{2m}2^{2m}$ as the number of paths of length $2m$ between the points $(0, 0)$ and $(2m, 0)$ that do not touch the horizontal axis except at the endpoints. Using this idea, it is easy to prove the following theorem.

Theorem 12.2 For $n \geq 1$, the probabilities $\{u_{2k}\}$ and $\{f_{2k}\}$ are related by the equation

$$u_{2n} = f_0 u_{2n} + f_2 u_{2n-2} + \cdots + f_{2n} u_0 .$$

Proof. There are $u_{2n}2^{2n}$ paths of length $2n$ which have endpoints $(0, 0)$ and $(2n, 0)$. The collection of such paths can be partitioned into n sets, depending upon the time of the first return to the origin. A path in this collection which has a first return to the origin at time $2k$ consists of an initial segment from $(0, 0)$ to $(2k, 0)$, in which no interior points are on the horizontal axis, and a terminal segment from $(2k, 0)$ to $(2n, 0)$, with no further restrictions on this segment. Thus, the number of paths in the collection which have a first return to the origin at time $2k$ is given by

$$f_{2k}2^{2k}u_{2n-2k}2^{2n-2k} = f_{2k}u_{2n-2k}2^{2n} .$$

If we sum over k, we obtain the equation

$$u_{2n}2^{2n} = f_0 u_{2n}2^{2n} + f_2 u_{2n-2}2^{2n} + \cdots + f_{2n} u_0 2^{2n} .$$

Dividing both sides of this equation by 2^{2n} completes the proof. □

The expression in the right-hand side of the above theorem should remind the reader of a sum that appeared in Definition 7.1 of the convolution of two distributions. The convolution of two sequences is defined in a similar manner. The above theorem says that the sequence $\{u_{2n}\}$ is the convolution of itself and the sequence $\{f_{2n}\}$. Thus, if we represent each of these sequences by an ordinary generating function, then we can use the above relationship to determine the value f_{2n}.

Theorem 12.3 For $m \geq 1$, the probability of a first return to the origin at time $2m$ is given by

$$f_{2m} = \frac{u_{2m}}{2m-1} = \frac{\binom{2m}{m}}{(2m-1)2^{2m}} .$$

Proof. We begin by defining the generating functions

$$U(x) = \sum_{m=0}^{\infty} u_{2m} x^m$$

and

$$F(x) = \sum_{m=0}^{\infty} f_{2m} x^m .$$

Theorem 12.2 says that

$$U(x) = 1 + U(x)F(x) . \qquad (12.1)$$

(The presence of the 1 on the right-hand side is due to the fact that u_0 is defined to be 1, but Theorem 12.2 only holds for $m \geq 1$.) We note that both generating functions certainly converge on the interval $(-1, 1)$, since all of the coefficients are at most 1 in absolute value. Thus, we can solve the above equation for $F(x)$, obtaining

$$F(x) = \frac{U(x) - 1}{U(x)} .$$

Now, if we can find a closed-form expression for the function $U(x)$, we will also have a closed-form expression for $F(x)$. From Theorem 12.1, we have

$$U(x) = \sum_{m=0}^{\infty} \binom{2m}{m} 2^{-2m} x^m .$$

In Wilf,[1] we find that

$$\frac{1}{\sqrt{1-4x}} = \sum_{m=0}^{\infty} \binom{2m}{m} x^m .$$

The reader is asked to prove this statement in Exercise 1. If we replace x by $x/4$ in the last equation, we see that

$$U(x) = \frac{1}{\sqrt{1-x}} .$$

[1] H. S. Wilf, *Generatingfunctionology,* (Boston: Academic Press, 1990), p. 50.

Therefore, we have

$$\begin{aligned} F(x) &= \frac{U(x)-1}{U(x)} \\ &= \frac{(1-x)^{-1/2}-1}{(1-x)^{-1/2}} \\ &= 1-(1-x)^{1/2} \ . \end{aligned}$$

Although it is possible to compute the value of f_{2m} using the Binomial Theorem, it is easier to note that $F'(x) = U(x)/2$, so that the coefficients f_{2m} can be found by integrating the series for $U(x)$. We obtain, for $m \geq 1$,

$$\begin{aligned} f_{2m} &= \frac{u_{2m-2}}{2m} \\ &= \frac{\binom{2m-2}{m-1}}{m2^{2m-1}} \\ &= \frac{\binom{2m}{m}}{(2m-1)2^{2m}} \\ &= \frac{u_{2m}}{2m-1} \ , \end{aligned}$$

since

$$\binom{2m-2}{m-1} = \frac{m}{2(2m-1)}\binom{2m}{m} \ .$$

This completes the proof of the theorem. □

Probability of Eventual Return

In the symmetric random walk process in $\mathbf{R}^m$, what is the probability that the particle eventually returns to the origin? We first examine this question in the case that $m = 1$, and then we consider the general case. The results in the next two examples are due to Pólya.[2]

Example 12.1 (Eventual Return in $\mathbf{R}^1$) One has to approach the idea of eventual return with some care, since the sample space seems to be the set of all walks of infinite length, and this set is non-denumerable. To avoid difficulties, we will define w_n to be the probability that a first return has occurred no later than time n. Thus, w_n concerns the sample space of all walks of length n, which is a finite set. In terms of the w_n's, it is reasonable to define the probability that the particle eventually returns to the origin to be

$$w_* = \lim_{n\to\infty} w_n \ .$$

This limit clearly exists and is at most one, since the sequence $\{w_n\}_{n=1}^{\infty}$ is an increasing sequence, and all of its terms are at most one.

[2] G. Pólya, "Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz," Math. Ann., vol. 84 (1921), pp. 149-160.

In terms of the f_n probabilities, we see that

$$w_{2n} = \sum_{i=1}^{n} f_{2i} \ .$$

Thus,

$$w_* = \sum_{i=1}^{\infty} f_{2i} \ .$$

In the proof of Theorem 12.3, the generating function

$$F(x) = \sum_{m=0}^{\infty} f_{2m} x^m$$

was introduced. There it was noted that this series converges for $x \in (-1, 1)$. In fact, it is possible to show that this series also converges for $x = \pm 1$ by using Exercise 4, together with the fact that

$$f_{2m} = \frac{u_{2m}}{2m-1} \ .$$

(This fact was proved in the proof of Theorem 12.3.) Since we also know that

$$F(x) = 1 - (1-x)^{1/2} \ ,$$

we see that

$$w_* = F(1) = 1 \ .$$

Thus, with probability one, the particle returns to the origin.

An alternative proof of the fact that $w_* = 1$ can be obtained by using the results in Exercise 2. □

Example 12.2 (Eventual Return in $\mathbf{R}^m$) We now turn our attention to the case that the random walk takes place in more than one dimension. We define $f_{2n}^{(m)}$ to be the probability that the first return to the origin in $\mathbf{R}^m$ occurs at time $2n$. The quantity $u_{2n}^{(m)}$ is defined in a similar manner. Thus, $f_{2n}^{(1)}$ and $u_{2n}^{(1)}$ equal f_{2n} and u_{2n}, which were defined earlier. If, in addition, we define $u_0^{(m)} = 1$ and $f_0^{(m)} = 0$, then one can mimic the proof of Theorem 12.2, and show that for all $m \geq 1$,

$$u_{2n}^{(m)} = f_0^{(m)} u_{2n}^{(m)} + f_2^{(m)} u_{2n-2}^{(m)} + \cdots + f_{2n}^{(m)} u_0^{(m)} \ . \qquad (12.2)$$

We continue to generalize previous work by defining

$$U^{(m)}(x) = \sum_{n=0}^{\infty} u_{2n}^{(m)} x^n$$

and

$$F^{(m)}(x) = \sum_{n=0}^{\infty} f_{2n}^{(m)} x^n \ .$$

Then, by using Equation 12.2, we see that

$$U^{(m)}(x) = 1 + U^{(m)}(x)F^{(m)}(x) \ ,$$

as before. These functions will always converge in the interval $(-1, 1)$, since all of their coefficients are at most one in magnitude. In fact, since

$$w_*^{(m)} = \sum_{n=0}^{\infty} f_{2n}^{(m)} \le 1$$

for all m, the series for $F^{(m)}(x)$ converges at $x = 1$ as well, and $F^{(m)}(x)$ is left-continuous at $x = 1$, i.e.,

$$\lim_{x \uparrow 1} F^{(m)}(x) = F^{(m)}(1) \ .$$

Thus, we have

$$w_*^{(m)} = \lim_{x \uparrow 1} F^{(m)}(x) = \lim_{x \uparrow 1} \frac{U^{(m)}(x) - 1}{U^{(m)}(x)} \ , \tag{12.3}$$

so to determine $w_*^{(m)}$, it suffices to determine

$$\lim_{x \uparrow 1} U^{(m)}(x) \ .$$

We let $u^{(m)}$ denote this limit.

We claim that

$$u^{(m)} = \sum_{n=0}^{\infty} u_{2n}^{(m)} \ .$$

(This claim is reasonable; it says that to find out what happens to the function $U^{(m)}(x)$ at $x = 1$, just let $x = 1$ in the power series for $U^{(m)}(x)$.) To prove the claim, we note that the coefficients $u_{2n}^{(m)}$ are non-negative, so $U^{(m)}(x)$ increases monotonically on the interval $[0, 1)$. Thus, for each K, we have

$$\sum_{n=0}^{K} u_{2n}^{(m)} \le \lim_{x \uparrow 1} U^{(m)}(x) = u^{(m)} \le \sum_{n=0}^{\infty} u_{2n}^{(m)} \ .$$

By letting $K \to \infty$, we see that

$$u^{(m)} = \sum_{2n}^{\infty} u_{2n}^{(m)} \ .$$

This establishes the claim.

From Equation 12.3, we see that if $u^{(m)} < \infty$, then the probability of an eventual return is

$$\frac{u^{(m)} - 1}{u^{(m)}} \ ,$$

while if $u^{(m)} = \infty$, then the probability of eventual return is 1.

To complete the example, we must estimate the sum

$$\sum_{n=0}^{\infty} u_{2n}^{(m)} \ .$$

In Exercise 12, the reader is asked to show that

$$u_{2n}^{(2)} = \frac{1}{4^{2n}} \binom{2n}{n}^2 .$$

Using Stirling's Formula, it is easy to show that (see Exercise 13)

$$\binom{2n}{n} \sim \frac{2^{2n}}{\sqrt{\pi n}} ,$$

so

$$u_{2n}^{(2)} \sim \frac{1}{\pi n} .$$

From this it follows easily that

$$\sum_{n=0}^{\infty} u_{2n}^{(2)}$$

diverges, so $w_*^{(2)} = 1$, i.e., in $\mathbf{R}^2$, the probability of an eventual return is 1.

When $m = 3$, Exercise 12 shows that

$$u_{2n}^{(3)} = \frac{1}{2^{2n}} \binom{2n}{n} \sum_{j,k} \left(\frac{1}{3^n} \frac{n!}{j!k!(n-j-k)!} \right)^2 .$$

Let M denote the largest value of

$$\frac{n!}{j!k!(n-j-k)!} ,$$

over all non-negative values of j and k with $j + k \le n$. It is easy, using Stirling's Formula, to show that

$$M \sim \frac{c}{n} ,$$

for some constant c. Thus, we have

$$u_{2n}^{(3)} \le \frac{1}{2^{2n}} \binom{2n}{n} \sum_{j,k} \left(\frac{M}{3^n} \frac{n!}{j!k!(n-j-k)!} \right) .$$

Using Exercise 14, one can show that the right-hand expression is at most

$$\frac{c'}{n^{3/2}} ,$$

where c' is a constant. Thus,

$$\sum_{n=0}^{\infty} u_{2n}^{(3)}$$

converges, so $w_*^{(3)}$ is strictly less than one. This means that in $\mathbf{R}^3$, the probability of an eventual return to the origin is strictly less than one (in fact, it is approximately .65).

One may summarize these results by stating that one should not get drunk in more than two dimensions. □

Expected Number of Equalizations

We now give another example of the use of generating functions to find a general formula for terms in a sequence, where the sequence is related by recursion relations to other sequences. Exercise 9 gives still another example.

Example 12.3 (Expected Number of Equalizations) In this example, we will derive a formula for the expected number of equalizations in a random walk of length $2m$. As in the proof of Theorem 12.3, the method has four main parts. First, a recursion is found which relates the mth term in the unknown sequence to earlier terms in the same sequence and to terms in other (known) sequences. An example of such a recursion is given in Theorem 12.2. Second, the recursion is used to derive a functional equation involving the generating functions of the unknown sequence and one or more known sequences. Equation 12.1 is an example of such a functional equation. Third, the functional equation is solved for the unknown generating function. Last, using a device such as the Binomial Theorem, integration, or differentiation, a formula for the mth coefficient of the unknown generating function is found.

We begin by defining g_{2m} to be the number of equalizations among all of the random walks of length $2m$. (For each random walk, we disregard the equalization at time 0.) We define $g_0 = 0$. Since the number of walks of length $2m$ equals 2^{2m}, the expected number of equalizations among all such random walks is $g_{2m}/2^{2m}$. Next, we define the generating function $G(x)$:

$$G(x) = \sum_{k=0}^{\infty} g_{2k} x^k \ .$$

Now we need to find a recursion which relates the sequence $\{g_{2k}\}$ to one or both of the known sequences $\{f_{2k}\}$ and $\{u_{2k}\}$. We consider m to be a fixed positive integer, and consider the set of all paths of length $2m$ as the disjoint union

$$E_2 \cup E_4 \cup \cdots \cup E_{2m} \cup H \ ,$$

where E_{2k} is the set of all paths of length $2m$ with first equalization at time $2k$, and H is the set of all paths of length $2m$ with no equalization. It is easy to show (see Exercise 3) that

$$|E_{2k}| = f_{2k} 2^{2m} \ .$$

We claim that the number of equalizations among all paths belonging to the set E_{2k} is equal to

$$|E_{2k}| + 2^{2k} f_{2k} g_{2m-2k} \ . \qquad (12.4)$$

Each path in E_{2k} has one equalization at time $2k$, so the total number of such equalizations is just $|E_{2k}|$. This is the first summand in expression Equation 12.4. There are $2^{2k} f_{2k}$ different initial segments of length $2k$ among the paths in E_{2k}. Each of these initial segments can be augmented to a path of length $2m$ in 2^{2m-2k} ways, by adjoining all possible paths of length $2m-2k$. The number of equalizations obtained by adjoining all of these paths to any one initial segment is g_{2m-2k}, by definition. This gives the second summand in Equation 12.4. Since k can range from 1 to m, we obtain the recursion

$$g_{2m} = \sum_{k=1}^{m} \Bigl(|E_{2k}| + 2^{2k} f_{2k} g_{2m-2k} \Bigr) \ . \tag{12.5}$$

The second summand in the typical term above should remind the reader of a convolution. In fact, if we multiply the generating function $G(x)$ by the generating function

$$F(4x) = \sum_{k=0}^{\infty} 2^{2k} f_{2k} x^k \ ,$$

the coefficient of x^m equals

$$\sum_{k=0}^{m} 2^{2k} f_{2k} g_{2m-2k} \ .$$

Thus, the product $G(x)F(4x)$ is part of the functional equation that we are seeking. The first summand in the typical term in Equation 12.5 gives rise to the sum

$$2^{2m} \sum_{k=1}^{m} f_{2k} \ .$$

From Exercise 2, we see that this sum is just $(1-u_{2m})2^{2m}$. Thus, we need to create a generating function whose mth coefficient is this term; this generating function is

$$\sum_{m=0}^{\infty} (1 - u_{2m}) 2^{2m} x^m \ ,$$

or

$$\sum_{m=0}^{\infty} 2^{2m} x^m + \sum_{m=0}^{\infty} u_{2m} x^m \ .$$

The first sum is just $(1-4x)^{-1}$, and the second sum is $U(4x)$. So, the functional equation which we have been seeking is

$$G(x) = F(4x)G(x) + \frac{1}{1-4x} - U(4x) \ .$$

If we solve this recursion for $G(x)$, and simplify, we obtain

$$G(x) = \frac{1}{(1-4x)^{3/2}} - \frac{1}{(1-4x)} \ . \tag{12.6}$$

We now need to find a formula for the coefficient of x^m. The first summand in Equation 12.6 is $(1/2)U'(4x)$, so the coefficient of x^m in this function is

$$u_{2m+2} 2^{2m+1} (m+1) \ .$$

The second summand in Equation 12.6 is the sum of a geometric series with common ratio $4x$, so the coefficient of x^m is 2^{2m}. Thus, we obtain

$$\begin{aligned} g_{2m} &= u_{2m+2}2^{2m+1}(m+1) - 2^{2m} \\ &= \frac{1}{2}\binom{2m+2}{m+1}(m+1) - 2^{2m}\ . \end{aligned}$$

We recall that the quotient $g_{2m}/2^{2m}$ is the expected number of equalizations among all paths of length $2m$. Using Exercise 4, it is easy to show that

$$\frac{g_{2m}}{2^{2m}} \sim \sqrt{\frac{2}{\pi}}\sqrt{2m}\ .$$

In particular, this means that the average number of equalizations among all paths of length $4m$ is not twice the average number of equalizations among all paths of length $2m$. In order for the average number of equalizations to double, one must quadruple the lengths of the random walks. □

It is interesting to note that if we define

$$M_n = \max_{0\le k\le n} S_k\ ,$$

then we have

$$E(M_n) \sim \sqrt{\frac{2}{\pi}}\sqrt{n}\ .$$

This means that the expected number of equalizations and the expected maximum value for random walks of length n are asymptotically equal as $n \to \infty$. (In fact, it can be shown that the two expected values differ by at most $1/2$ for all positive integers n. See Exercise 9.)

Exercises

1 Using the Binomial Theorem, show that

$$\frac{1}{\sqrt{1-4x}} = \sum_{m=0}^{\infty} \binom{2m}{m} x^m\ .$$

What is the interval of convergence of this power series?

2 (a) Show that for $m \ge 1$,

$$f_{2m} = u_{2m-2} - u_{2m}\ .$$

(b) Using part (a), find a closed-form expression for the sum

$$f_2 + f_4 + \cdots + f_{2m}\ .$$

(c) Using part (a), show that

$$\sum_{m=1}^{\infty} f_{2m} = 1\ .$$

(One can also obtain this statement from the fact that

$$F(x) = 1 - (1-x)^{1/2}\ .)$$

(d) Using Exercise 2, show that the probability of no equalization in the first $2m$ outcomes equals the probability of an equalization at time $2m$.

3 Using the notation of Example 12.3, show that

$$|E_{2k}| = f_{2k} 2^{2m} \ .$$

4 Using Stirling's Formula, show that

$$u_{2m} \sim \frac{1}{\sqrt{\pi m}} \ .$$

5 A *lead change* in a random walk occurs at time $2k$ if S_{2k-1} and S_{2k+1} are of opposite sign.

(a) Give a rigorous argument which proves that among all walks of length $2m$ that have an equalization at time $2k$, exactly half have a lead change at time $2k$.

(b) Deduce that the total number of lead changes among all walks of length $2m$ equals

$$\frac{1}{2}(g_{2m} - u_{2m}) \ .$$

(c) Find an asymptotic expression for the average number of lead changes in a random walk of length $2m$.

6 (a) Show that the probability that a random walk of length $2m$ has a last return to the origin at time $2k$, where $0 \le k \le m$, equals

$$\frac{\binom{2k}{k}\binom{2m-2k}{m-k}}{2^{2m}} = u_{2k} u_{2m-2k} \ .$$

(The case $k = 0$ consists of all paths that do not return to the origin at any positive time.) *Hint*: A path whose last return to the origin occurs at time $2k$ consists of two paths glued together, one path of which is of length $2k$ and which begins and ends at the origin, and the other path of which is of length $2m - 2k$ and which begins at the origin but never returns to the origin. Both types of paths can be counted using quantities which appear in this section.

(b) Using part (a), show that the probability that a walk of length $2m$ has no equalization in the last m outcomes is equal to $1/2$, regardless of the value of m. *Hint*: The answer to part a) is symmetric in k and $m - k$.

7 Show that the probability of no equalization in a walk of length $2m$ equals u_{2m}.

***8** Show that

$$P(S_1 \ge 0,\ S_2 \ge 0,\ \ldots,\ S_{2m} \ge 0) = u_{2m} \ .$$

Hint: First explain why

$$P(S_1 > 0,\ S_2 > 0,\ \ldots,\ S_{2m} > 0) \\ = \frac{1}{2} P(S_1 \neq 0,\ S_2 \neq 0,\ \ldots,\ S_{2m} \neq 0)\ .$$

Then use Exercise 7, together with the observation that if no equalization occurs in the first $2m$ outcomes, then the path goes through the point $(1,1)$ and remains on or above the horizontal line $x = 1$.

***9** In Feller,[3] one finds the following theorem: Let M_n be the random variable which gives the maximum value of S_k, for $1 \leq k \leq n$. Define

$$p_{n,r} = \binom{n}{\frac{n+r}{2}} 2^{-n}\ .$$

If $r \geq 0$, then

$$P(M_n = r) = \begin{cases} p_{n,r}, & \text{if } r \equiv n \pmod 2, \\ 1, & \text{if } p \geq q, \end{cases}$$

$$P(M_n = r) = \begin{cases} p_{n,r}\ , & \text{if } r \equiv n \pmod 2, \\ p_{n,r+1}\ , & \text{if } r \not\equiv n \pmod 2. \end{cases}$$

(a) Using this theorem, show that

$$E(M_{2m}) = \frac{1}{2^{2m}} \sum_{k=1}^{m} (4k-1) \binom{2m}{m+k}\ ,$$

and if $n = 2m + 1$, then

$$E(M_{2m+1}) = \frac{1}{2^{2m+1}} \sum_{k=0}^{m} (4k+1) \binom{2m+1}{m+k+1}\ .$$

(b) For $m \geq 1$, define

$$r_m = \sum_{k=1}^{m} k \binom{2m}{m+k}$$

and

$$s_m = \sum_{k=1}^{m} k \binom{2m+1}{m+k+1}\ .$$

By using the identity

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}\ ,$$

show that

$$s_m = 2r_m - \frac{1}{2}\left(2^{2m} - \binom{2m}{m}\right)$$

[3] W. Feller, *Introduction to Probability Theory and its Applications,* vol. I, 3rd ed. (New York: John Wiley & Sons, 1968).

and

$$r_m = 2s_{m-1} + \frac{1}{2}2^{2m-1} ,$$

if $m \geq 2$.

(c) Define the generating functions

$$R(x) = \sum_{k=1}^{\infty} r_k x^k$$

and

$$S(x) = \sum_{k=1}^{\infty} s_k x^k .$$

Show that

$$S(x) = 2R(x) - \frac{1}{2}\left(\frac{1}{1-4x}\right) + \frac{1}{2}\left(\sqrt{1-4x}\right)$$

and

$$R(x) = 2xS(x) + x\left(\frac{1}{1-4x}\right) .$$

(d) Show that

$$R(x) = \frac{x}{(1-4x)^{3/2}} ,$$

and

$$S(x) = \frac{1}{2}\left(\frac{1}{(1-4x)^{3/2}}\right) - \frac{1}{2}\left(\frac{1}{1-4x}\right) .$$

(e) Show that

$$r_m = m\binom{2m-1}{m-1} ,$$

and

$$s_m = \frac{1}{2}(m+1)\binom{2m+1}{m} - \frac{1}{2}(2^{2m}) .$$

(f) Show that

$$E(M_{2m}) = \frac{m}{2^{2m-1}}\binom{2m}{m} + \frac{1}{2^{2m+1}}\binom{2m}{m} - \frac{1}{2} ,$$

and

$$E(M_{2m+1}) = \frac{m+1}{2^{2m+1}}\binom{2m+2}{m+1} - \frac{1}{2} .$$

The reader should compare these formulas with the expression for $g_{2m}/2^{(2m)}$ in Example 12.3.

***10** (from K. Levasseur[4]) A parent and his child play the following game. A deck of $2n$ cards, n red and n black, is shuffled. The cards are turned up one at a time. Before each card is turned up, the parent and the child guess whether it will be red or black. Whoever makes more correct guesses wins the game. The child is assumed to guess each color with the same probability, so she will have a score of n, on average. The parent keeps track of how many cards of each color have already been turned up. If more black cards, say, than red cards remain in the deck, then the parent will guess black, while if an equal number of each color remain, then the parent guesses each color with probability 1/2. What is the expected number of correct guesses that will be made by the parent? *Hint*: Each of the $\binom{2n}{n}$ possible orderings of red and black cards corresponds to a random walk of length $2n$ that returns to the origin at time $2n$. Show that between each pair of successive equalizations, the parent will be right exactly once more than he will be wrong. Explain why this means that the average number of correct guesses by the parent is greater than n by exactly one-half the average number of equalizations. Now define the random variable X_i to be 1 if there is an equalization at time $2i$, and 0 otherwise. Then, among all relevant paths, we have

$$E(X_i) = P(X_i = 1) = \frac{\binom{2n-2i}{n-i}\binom{2i}{i}}{\binom{2n}{n}} \ .$$

Thus, the expected number of equalizations equals

$$E\biggl(\sum_{i=1}^n X_i\biggr) = \frac{1}{\binom{2n}{n}} \sum_{i=1}^n \binom{2n-2i}{n-i}\binom{2i}{i} \ .$$

One can now use generating functions to find the value of the sum.

It should be noted that in a game such as this, a more interesting question than the one asked above is what is the probability that the parent wins the game? For this game, this question was answered by D. Zagier.[5] He showed that the probability of winning is asymptotic (for large n) to the quantity

$$\frac{1}{2} + \frac{1}{2\sqrt{2}} \ .$$

***11** Prove that

$$u^{(2)}_{2n} = \frac{1}{4^{2n}} \sum_{k=0}^n \frac{(2n)!}{k!k!(n-k)!(n-k)!} \ ,$$

and

$$u^{(3)}_{2n} = \frac{1}{6^{2n}} \sum_{j,k} \frac{(2n)!}{j!j!k!k!(n-j-k)!(n-j-k)!} \ ,$$

[4] K. Levasseur, "How to Beat Your Kids at Their Own Game," *Mathematics Magazine* vol. 61, no. 5 (December, 1988), pp. 301-305.

[5] D. Zagier, "How Often Should You Beat Your Kids?" *Mathematics Magazine* vol. 63, no. 2 (April 1990), pp. 89-92.

where the last sum extends over all non-negative j and k with $j+k \le n$. Also show that this last expression may be rewritten as

$$\frac{1}{2^{2n}}\binom{2n}{n}\sum_{j,k}\left(\frac{1}{3^n}\frac{n!}{j!k!(n-j-k)!}\right)^2 .$$

***12** Prove that if $n \ge 0$, then

$$\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n} .$$

Hint: Write the sum as

$$\sum_{k=0}^{n}\binom{n}{k}\binom{n}{n-k}$$

and explain why this is a coefficient in the product

$$(1+x)^n(1+x)^n .$$

Use this, together with Exercise 11, to show that

$$u_{2n}^{(2)} = \frac{1}{4^{2n}}\binom{2n}{n}\sum_{k=0}^{n}\binom{n}{k}^2 = \frac{1}{4^{2n}}\binom{2n}{n}^2 .$$

***13** Using Stirling's Formula, prove that

$$\binom{2n}{n} \sim \frac{2^{2n}}{\sqrt{\pi n}} .$$

***14** Prove that

$$\sum_{j,k}\left(\frac{1}{3^n}\frac{n!}{j!k!(n-j-k)!}\right) = 1 ,$$

where the sum extends over all non-negative j and k such that $j+k \le n$. *Hint*: Count how many ways one can place n labelled balls in 3 labelled urns.

***15** Using the result proved for the random walk in $\mathbf{R}^3$ in Example 12.2, explain why the probability of an eventual return in $\mathbf{R}^n$ is strictly less than one, for all $n \ge 3$. *Hint*: Consider a random walk in $\mathbf{R}^n$ and disregard all but the first three coordinates of the particle's position.

12.2 Gambler's Ruin

In the last section, the simplest kind of symmetric random walk in $\mathbf{R}^1$ was studied. In this section, we remove the assumption that the random walk is symmetric. Instead, we assume that p and q are non-negative real numbers with $p+q=1$, and that the common distribution function of the jumps of the random walk is

$$f_X(x) = \begin{cases} p, & \text{if } x = 1, \\ q, & \text{if } x = -1. \end{cases}$$

One can imagine the random walk as representing a sequence of tosses of a weighted coin, with a head appearing with probability p and a tail appearing with probability q. An alternative formulation of this situation is that of a gambler playing a sequence of games against an adversary (sometimes thought of as another person, sometimes called "the house") where, in each game, the gambler has probability p of winning.

The Gambler's Ruin Problem

The above formulation of this type of random walk leads to a problem known as the Gambler's Ruin problem. This problem was introduced in Exercise 23, but we will give the description of the problem again. A gambler starts with a "stake" of size s. She plays until her capital reaches the value M or the value 0. In the language of Markov chains, these two values correspond to absorbing states. We are interested in studying the probability of occurrence of each of these two outcomes.

One can also assume that the gambler is playing against an "infinitely rich" adversary. In this case, we would say that there is only one absorbing state, namely when the gambler's stake is 0. Under this assumption, one can ask for the probability that the gambler is eventually ruined.

We begin by defining q_k to be the probability that the gambler's stake reaches 0, i.e., she is ruined, before it reaches M, given that the initial stake is k. We note that $q_0 = 1$ and $q_M = 0$. The fundamental relationship among the q_k's is the following:

$$q_k = pq_{k+1} + qq_{k-1} \ ,$$

where $1 \le k \le M-1$. This holds because if her stake equals k, and she plays one game, then her stake becomes $k+1$ with probability p and $k-1$ with probability q. In the first case, the probability of eventual ruin is q_{k+1} and in the second case, it is q_{k-1}. We note that since $p + q = 1$, we can write the above equation as

$$p(q_{k+1} - q_k) = q(q_k - q_{k-1}) \ ,$$

or

$$q_{k+1} - q_k = \frac{q}{p}(q_k - q_{k-1}) \ .$$

From this equation, it is easy to see that

$$q_{k+1} - q_k = \left(\frac{q}{p}\right)^k (q_1 - q_0) \ . \qquad (12.7)$$

We now use telescoping sums to obtain an equation in which the only unknown is q_1:

$$\begin{aligned} -1 &= q_M - q_0 \\ &= \sum_{k=0}^{M-1} (q_{k+1} - q_k) \ , \end{aligned}$$

so

$$\begin{aligned} -1 &= \sum_{k=0}^{M-1} \left(\frac{q}{p}\right)^k (q_1 - q_0) \\ &= (q_1 - q_0) \sum_{k=0}^{M-1} \left(\frac{q}{p}\right)^k . \end{aligned}$$

If $p \neq q$, then the above expression equals

$$(q_1 - q_0)\frac{(q/p)^M - 1}{(q/p) - 1} ,$$

while if $p = q = 1/2$, then we obtain the equation

$$-1 = (q_1 - q_0)M .$$

For the moment we shall assume that $p \neq q$. Then we have

$$q_1 - q_0 = -\frac{(q/p) - 1}{(q/p)^M - 1} .$$

Now, for any z with $1 \leq z \leq M$, we have

$$\begin{aligned} q_z - q_0 &= \sum_{k=0}^{z-1} (q_{k+1} - q_k) \\ &= (q_1 - q_0) \sum_{k=0}^{z-1} \left(\frac{q}{p}\right)^k \\ &= -(q_1 - q_0)\frac{(q/p)^z - 1}{(q/p) - 1} \\ &= -\frac{(q/p)^z - 1}{(q/p)^M - 1} . \end{aligned}$$

Therefore,

$$\begin{aligned} q_z &= 1 - \frac{(q/p)^z - 1}{(q/p)^M - 1} \\ &= \frac{(q/p)^M - (q/p)^z}{(q/p)^M - 1} . \end{aligned}$$

Finally, if $p = q = 1/2$, it is easy to show that (see Exercise 10)

$$q_z = \frac{M - z}{M} .$$

We note that both of these formulas hold if $z = 0$.

We define, for $0 \leq z \leq M$, the quantity p_z to be the probability that the gambler's stake reaches M without ever having reached 0. Since the game might continue indefinitely, it is not obvious that $p_z + q_z = 1$ for all z. However, one can use the same method as above to show that if $p \neq q$, then

$$q_z = \frac{(q/p)^z - 1}{(q/p)^M - 1} \ ,$$

and if $p = q = 1/2$, then

$$q_z = \frac{z}{M} \ .$$

Thus, for all z, it is the case that $p_z + q_z = 1$, so the game ends with probability 1.

Infinitely Rich Adversaries

We now turn to the problem of finding the probability of eventual ruin if the gambler is playing against an infinitely rich adversary. This probability can be obtained by letting M go to ∞ in the expression for q_z calculated above. If $q < p$, then the expression approaches $(q/p)^z$, and if $q > p$, the expression approaches 1. In the case $p = q = 1/2$, we recall that $q_z = 1 - z/M$. Thus, if $M \to \infty$, we see that the probability of eventual ruin tends to 1.

Historical Remarks

In 1711, De Moivre, in his book *De Mesura Sortis*, gave an ingenious derivation of the probability of ruin. The following description of his argument is taken from David.[6] The notation used is as follows: We imagine that there are two players, A and B, and the probabilities that they win a game are p and q, respectively. The players start with a and b counters, respectively.

> Imagine that each player starts with his counters before him in a pile, and that nominal values are assigned to the counters in the following manner. A's bottom counter is given the nominal value q/p; the next is given the nominal value $(q/p)^2$, and so on until his top counter which has the nominal value $(q/p)^a$. B's top counter is valued $(q/p)^{a+1}$, and so on downwards until his bottom counter which is valued $(q/p)^{a+b}$. After each game the loser's top counter is transferred to the top of the winner's pile, and it is always the top counter which is staked for the next game. Then *in terms of the nominal values* B's stake is always q/p times A's, so that at every game each player's nominal expectation is nil. This remains true throughout the play; therefore A's chance of winning all B's counters, multiplied by his nominal gain if he does so, must equal B's chance multiplied by B's nominal gain. Thus,

$$P_a\left(\left(\frac{q}{p}\right)^{a+1} + \cdots + \left(\frac{q}{p}\right)^{a+b}\right) = P_b\left(\left(\frac{q}{p}\right) + \cdots + \left(\frac{q}{p}\right)^{a}\right) \ . \qquad (12.8)$$

[6] F. N. David, *Games, Gods and Gambling* (London: Griffin, 1962).

Using this equation, together with the fact that

$$P_a + P_b = 1 \ ,$$

it can easily be shown that

$$P_a = \frac{(q/p)^a - 1}{(q/p)^{a+b} - 1} \ ,$$

if $p \neq q$, and

$$P_a = \frac{a}{a+b} \ ,$$

if $p = q = 1/2$.

In terms of modern probability theory, de Moivre is changing the values of the counters to make an unfair game into a fair game, which is called a martingale. With the new values, the expected fortune of player A (that is, the sum of the nominal values of his counters) after each play equals his fortune before the play (and similarly for player B). (For a simpler martingale argument, see Exercise 9.) De Moivre then uses the fact that when the game ends, it is still fair, thus Equation 12.8 must be true. This fact requires proof, and is one of the central theorems in the area of martingale theory.

Exercises

1 In the gambler's ruin problem, assume that the gambler initial stake is 1 dollar, and assume that her probability of success on any one game is p. Let T be the number of games until 0 is reached (the gambler is ruined). Show that the generating function for T is

$$h(z) = \frac{1 - \sqrt{1 - 4pqz^2}}{2pz} \ ,$$

and that

$$h(1) = \begin{cases} q/p, & \text{if } q \leq p, \\ 1, & \text{if } q \geq p, \end{cases}$$

and

$$h'(1) = \begin{cases} 1/(q-p), & \text{if } q > p, \\ \infty, & \text{if } q = p. \end{cases}$$

Interpret your results in terms of the time T to reach 0. (See also Example 10.7.)

2 Show that the Taylor series expansion for $\sqrt{1-x}$ is

$$\sqrt{1-x} = \sum_{n=0}^{\infty} \binom{1/2}{n} x^n \ ,$$

where the binomial coefficient $\binom{1/2}{n}$ is

$$\binom{1/2}{n} = \frac{(1/2)(1/2-1)\cdots(1/2-n+1)}{n!} \ .$$

Using this and the result of Exercise 1, show that the probability that the gambler is ruined on the nth step is

$$p_T(n) = \begin{cases} \frac{(-1)^{k-1}}{2p}\binom{1/2}{k}(4pq)^k, & \text{if } n = 2k-1, \\ 0, & \text{if } n = 2k. \end{cases}$$

3 For the gambler's ruin problem, assume that the gambler starts with k dollars. Let T_k be the time to reach 0 for the first time.

(a) Show that the generating function $h_k(t)$ for T_k is the kth power of the generating function for the time T to ruin starting at 1. *Hint*: Let $T_k = U_1 + U_2 + \cdots + U_k$, where U_j is the time for the walk starting at j to reach $j-1$ for the first time.

(b) Find $h_k(1)$ and $h'_k(1)$ and interpret your results.

4 (The next three problems come from Feller.[7]) As in the text, assume that M is a fixed positive integer.

(a) Show that if a gambler starts with an stake of 0 (and is allowed to have a negative amount of money), then the probability that her stake reaches the value of M before it returns to 0 equals $p(1-q_1)$.

(b) Show that if the gambler starts with a stake of M then the probability that her stake reaches 0 before it returns to M equals qq_{M-1}.

5 Suppose that a gambler starts with a stake of 0 dollars.

(a) Show that the probability that her stake never reaches M before returning to 0 equals $1 - p(1-q_1)$.

(b) Show that the probability that her stake reaches the value M exactly k times before returning to 0 equals $p(1-q_1)(1-qq_{M-1})^{k-1}(qq_{M-1})$. *Hint*: Use Exercise 4.

6 In the text, it was shown that if $q < p$, there is a positive probability that a gambler, starting with a stake of 0 dollars, will never return to the origin. Thus, we will now assume that $q \geq p$. Using Exercise 5, show that if a gambler starts with a stake of 0 dollars, then the expected number of times her stake equals M before returning to 0 equals $(p/q)^M$, if $q > p$ and 1, if $q = p$. (We quote from Feller: "The truly amazing implications of this result appear best in the language of fair games. A perfect coin is tossed until the first equalization of the accumulated numbers of heads and tails. The gambler receives one penny for every time that the accumulated number of heads exceeds the accumulated number of tails by m. *The 'fair entrance fee' equals 1 independent of m.*")

[7] W. Feller, op. cit., pg. 367.

7 In the game in Exercise 6, let $p = q = 1/2$ and $M = 10$. What is the probability that the gambler's stake equals M at least 20 times before it returns to 0?

8 Write a computer program which simulates the game in Exercise 6 for the case $p = q = 1/2$, and $M = 10$.

9 In de Moivre's description of the game, we can modify the definition of player A's fortune in such a way that the game is still a martingale (and the calculations are simpler). We do this by assigning nominal values to the counters in the same way as de Moivre, but each player's current fortune is defined to be just the value of the counter which is being wagered on the next game. So, if player A has a counters, then his current fortune is $(q/p)^a$ (we stipulate this to be true even if $a = 0$). Show that under this definition, player A's expected fortune after one play equals his fortune before the play, if $p \neq q$. Then, as de Moivre does, write an equation which expresses the fact that player A's expected final fortune equals his initial fortune. Use this equation to find the probability of ruin of player A.

10 Assume in the gambler's ruin problem that $p = q = 1/2$.

(a) Using Equation 12.7, together with the facts that $q_0 = 1$ and $q_M = 0$, show that for $0 \leq z \leq M$,

$$q_z = \frac{M - z}{M} \ .$$

(b) In Equation 12.8, let $p \to 1/2$ (and since $q = 1 - p$, $q \to 1/2$ as well). Show that in the limit,

$$q_z = \frac{M - z}{M} \ .$$

Hint: Replace q by $1 - p$, and use L'Hopital's rule.

11 In American casinos, the roulette wheels have the integers between 1 and 36, together with 0 and 00. Half of the non-zero numbers are red, the other half are black, and 0 and 00 are green. A common bet in this game is to bet a dollar on red. If a red number comes up, the bettor gets her dollar back, and also gets another dollar. If a black or green number comes up, she loses her dollar.

(a) Suppose that someone starts with 40 dollars, and continues to bet on red until either her fortune reaches 50 or 0. Find the probability that her fortune reaches 50 dollars.

(b) How much money would she have to start with, in order for her to have a 95% chance of winning 10 dollars before going broke?

(c) A casino owner was once heard to remark that "If we took 0 and 00 off of the roulette wheel, we would still make lots of money, because people would continue to come in and play until they lost all of their money." Do you think that such a casino would stay in business?

12.3 Arc Sine Laws

In Exercise 12.1.6, the distribution of the time of the last equalization in the symmetric random walk was determined. If we let $\alpha_{2k,2m}$ denote the probability that a random walk of length $2m$ has its last equalization at time $2k$, then we have

$$\alpha_{2k,2m} = u_{2k}u_{2m-2k}\ .$$

We shall now show how one can approximate the distribution of the α's with a simple function. We recall that

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}}\ .$$

Therefore, as both k and m go to ∞, we have

$$\alpha_{2k,2m} \sim \frac{1}{\pi\sqrt{k(m-k)}}\ .$$

This last expression can be written as

$$\frac{1}{\pi m\sqrt{(k/m)(1-k/m)}}\ .$$

Thus, if we define

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}\ ,$$

for $0 < x < 1$, then we have

$$\alpha_{2k,2m} \approx \frac{1}{m} f\left(\frac{k}{m}\right)\ .$$

The reason for the $\approx$ sign is that we no longer require that k get large. This means that we can replace the discrete $\alpha_{2k,2m}$ distribution by the continuous density $f(x)$ on the interval $[0,1]$ and obtain a good approximation. In particular, if x is a fixed real number between 0 and 1, then we have

$$\sum_{k<xm} \alpha_{2k,2m} \approx \int_0^x f(t)\,dt\ .$$

It turns out that $f(x)$ has a nice antiderivative, so we can write

$$\sum_{k<xm} \alpha_{2k,2m} \approx \frac{2}{\pi} \arcsin\sqrt{x}\ .$$

One can see from the graph of this last function that it has a minimum at $x = 1/2$ and is symmetric about that point. As noted in the exercise, this implies that half of the walks of length $2m$ have no equalizations after time m, a fact which probably would not be guessed.

It turns out that the arc sine density comes up in the answers to many other questions concerning random walks on the line. Recall that in Section 12.1, a random walk could be viewed as a polygonal line connecting $(0,0)$ with (m, S_m). Under this interpretation, we define $b_{2k,2m}$ to be the probability that a random walk of length $2m$ has exactly $2k$ of its $2m$ polygonal line segments above the t-axis.

The probability $b_{2k,2m}$ is frequently interpreted in terms of a two-player game. (The reader will recall the game Heads or Tails, in Example 1.4.) Player A is said to be in the lead at time n if the random walk is above the t-axis at that time, or if the random walk is on the t-axis at time n but above the t-axis at time $n-1$. (At time 0, neither player is in the lead.) One can ask what is the most probable number of times that player A is in the lead, in a game of length $2m$. Most people will say that the answer to this question is m. However, the following theorem says that m is the least likely number of times that player A is in the lead, and the most likely number of times in the lead is 0 or $2m$.

Theorem 12.4 If Peter and Paul play a game of Heads or Tails of length $2m$, the probability that Peter will be in the lead exactly $2k$ times is equal to

$$\alpha_{2k,2m} \ .$$

Proof. To prove the theorem, we need to show that

$$b_{2k,2m} = \alpha_{2k,2m} \ . \tag{12.9}$$

Exercise 12.1.7 shows that $b_{2m,2m} = u_{2m}$ and $b_{0,2m} = u_{2m}$, so we only need to prove that Equation 12.9 holds for $1 \le k \le m-1$. We can obtain a recursion involving the b's and the f's (defined in Section 12.1) by counting the number of paths of length $2m$ that have exactly $2k$ of their segments above the t-axis, where $1 \le k \le m-1$. To count this collection of paths, we assume that the first return occurs at time $2j$, where $1 \le j \le m-1$. There are two cases to consider. Either during the first $2j$ outcomes the path is above the t-axis or below the t-axis. In the first case, it must be true that the path has exactly $(2k-2j)$ line segments above the t-axis, between $t = 2j$ and $t = 2m$. In the second case, it must be true that the path has exactly $2k$ line segments above the t-axis, between $t = 2j$ and $t = 2m$.

We now count the number of paths of the various types described above. The number of paths of length $2j$ all of whose line segments lie above the t-axis and which return to the origin for the first time at time $2j$ equals $(1/2)2^{2j}f_{2j}$. This also equals the number of paths of length $2j$ all of whose line segments lie below the t-axis and which return to the origin for the first time at time $2j$. The number of paths of length $(2m-2j)$ which have exactly $(2k-2j)$ line segments above the t-axis is $b_{2k-2j,2m-2j}$. Finally, the number of paths of length $(2m-2j)$ which have exactly $2k$ line segments above the t-axis is $b_{2k,2m-2j}$. Therefore, we have

$$b_{2k,2m} = \frac{1}{2}\sum_{j=1}^{k} f_{2j} b_{2k-2j,2m-2j} + \frac{1}{2}\sum_{j=1}^{m-k} f_{2j} b_{2k,2m-2j} \ .$$

We now assume that Equation 12.9 is true for $m < n$. Then we have

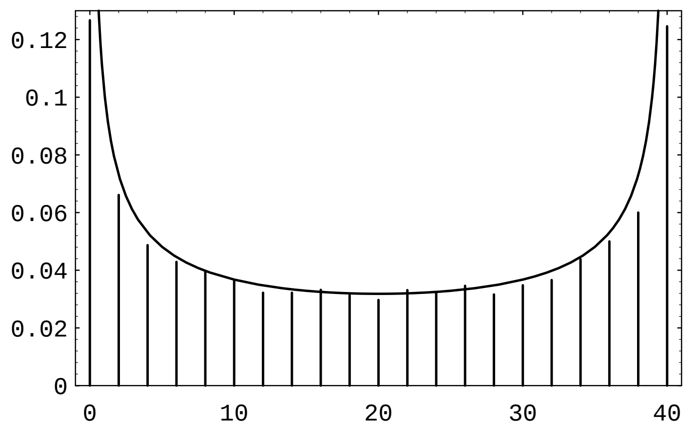

Figure 12.2: Times in the lead.

$$\begin{aligned}
b_{2k,2n} &= \frac{1}{2}\sum_{j=1}^{k} f_{2j}\alpha_{2k-2j,2m-2j} + \frac{1}{2}\sum_{j=1}^{m-k} f_{2j}\alpha_{2k,2m-2j} \\
&= \frac{1}{2}\sum_{j=1}^{k} f_{2j}u_{2k-2j}u_{2m-2k} + \frac{1}{2}\sum_{j=1}^{m-k} f_{2j}u_{2k}u_{2m-2j-2k} \\
&= \frac{1}{2}u_{2m-2k}\sum_{j=1}^{k} f_{2j}u_{2k-2j} + \frac{1}{2}u_{2k}\sum_{j=1}^{m-k} f_{2j}u_{2m-2j-2k} \\
&= \frac{1}{2}u_{2m-2k}u_{2k} + \frac{1}{2}u_{2k}u_{2m-2k}\ ,
\end{aligned}$$

where the last equality follows from Theorem 12.2. Thus, we have

$$b_{2k,2n} = \alpha_{2k,2n}\ ,$$

which completes the proof. □

We illustrate the above theorem by simulating 10,000 games of Heads or Tails, with each game consisting of 40 tosses. The distribution of the number of times that Peter is in the lead is given in Figure 12.2, together with the arc sine density.

We end this section by stating two other results in which the arc sine density appears. Proofs of these results may be found in Feller.[8]

Theorem 12.5 Let J be the random variable which, for a given random walk of length $2m$, gives the smallest subscript j such that $S_j = S_{2m}$. (Such a subscript j must be even, by parity considerations.) Let $\gamma_{2k,2m}$ be the probability that $J = 2k$. Then we have

$$\gamma_{2k,2m} = \alpha_{2k,2m}\ .$$

□

[8]W. Feller, op. cit., pp. 93–94.

The next theorem says that the arc sine density is applicable to a wide range of situations. A continuous distribution function $F(x)$ is said to be *symmetric* if $F(x) = 1 - F(-x)$. (If X is a continuous random variable with a symmetric distribution function, then for any real x, we have $P(X \le x) = P(X \ge -x)$.) We imagine that we have a random walk of length n in which each summand has the distribution $F(x)$, where F is continuous and symmetric. The subscript of the *first maximum* of such a walk is the unique subscript k such that

$$S_k > S_0,\ \ldots,\ S_k > S_{k-1},\ S_k \ge S_{k+1},\ \ldots,\ S_k \ge S_n\ .$$

We define the random variable K_n to be the subscript of the first maximum. We can now state the following theorem concerning the random variable K_n.

Theorem 12.6 Let F be a symmetric continuous distribution function, and let α be a fixed real number strictly between 0 and 1. Then as $n \to \infty$, we have

$$P(K_n < n\alpha) \to \frac{2}{\pi} \arcsin \sqrt{\alpha}\ .$$

□

A version of this theorem that holds for a symmetric random walk can also be found in Feller.

Exercises

1 For a random walk of length $2m$, define ϵ_k to equal 1 if $S_k > 0$, or if $S_{k-1} = 1$ and $S_k = 0$. Define ϵ_k to equal -1 in all other cases. Thus, ϵ_k gives the side of the t-axis that the random walk is on during the time interval $[k-1, k]$. A "law of large numbers" for the sequence $\{\epsilon_k\}$ would say that for any $\delta > 0$, we would have

$$P\left(-\delta < \frac{\epsilon_1 + \epsilon_2 + \cdots + \epsilon_n}{n} < \delta\right) \to 1$$

as $n \to \infty$. Even though the ϵ's are not independent, the above assertion certainly appears reasonable. Using Theorem 12.4, show that if $-1 \le x \le 1$, then

$$\lim_{n\to\infty} P\left(\frac{\epsilon_1 + \epsilon_2 + \cdots + \epsilon_n}{n} < x\right) = \frac{2}{\pi} \arcsin \sqrt{\frac{1+x}{2}}\ .$$

2 Given a random walk W of length m, with summands

$$\{X_1, X_2, \ldots, X_m\}\ ,$$

define the *reversed* random walk to be the walk W^* with summands

$$\{X_m, X_{m-1}, \ldots, X_1\}\ .$$

(a) Show that the kth partial sum S^*_k satisfies the equation

$$S^*_k = S_m - S_{n-k}\ ,$$

where S_k is the kth partial sum for the random walk W.

(b) Explain the geometric relationship between the graphs of a random walk and its reversal. (It is not in general true that one graph is obtained from the other by reflecting in a vertical line.)

(c) Use parts (a) and (b) to prove Theorem 12.5.