RANDOM WALKS IN ONE DIMENSION

STEVEN P. LALLEY

1. THE GAMBLER'S RUIN PROBLEM

1.1. **Statement of the problem.** I have A dollars; my colleague Xinyi has B dollars. A cup of coffee at the Sacred Grounds in Swift Hall costs $A+B$ dollars, which neither Xinyi nor I can afford, so we decide to do the only fair and rational thing: we gamble for it. To make the game fair, we agree to each bet 1 dollar at a time, with the winner decided by a fair coin toss; if the coin comes up Heads, Xinyi wins my dollar, but if it comes up Tails then I win his dollar. We play this game repeatedly, until one or the other of us has won all $A+B$ dollars.

Problem: What is the probability that I win?

Here is a more formal way of stating the problem. Let $X_1, X_2, X_3, \dots$ be an infinite sequence of independent random variables, each with distribution

$$P\{X_i = +1\} = P\{X_i = -1\} = \frac{1}{2}, \tag{1}$$

and for each $n = 1, 2, 3, \dots$ set

$$S_n = \sum_{i=1}^{n} X_i. \tag{2}$$

The random variable S_n represents the *net change* in my fortune after n tosses of the coin. Since my fortune hasn't changed before we start playing, set $S_0 = 0$. Now define

$$T = T_{A,B} = \min\{n \geq 0 : S_n = -A \text{ or } S_n = +B\}. \tag{3}$$

Then the gambler's ruin problem can be re-formulated as follows:

Problem: Evaluate $P\{S_T = +B\}$.

1.2. **Almost sure termination.** How do we know the game actually ends? That is, how do we know that $P\{T < \infty\} = 1$? Here is an argument that proves this.

Suppose that in the first $A+B$ tosses we got nothing but Heads; then the game would be over after the Ath toss. If, on the other hand, the game has *not* ended after $A+B$ tosses, then my fortune will be something between 1 and $A+B-1$ dollars, and so if the *next* $A+B$ tosses resulted in nothing but Heads then the game would be over by the $2A+2B$th toss. If the game hasn't ended by the $2A+2B$th toss, then my fortune will again be something between 1 and $A+B-1$ dollars, and so if the *next* $A+B$ tosses resulted in nothing but Heads then the game would be over by the $3A+3B$th toss. And so on.

But if we toss a fair coin infinitely often, then in some block of $A+B$ consecutive tosses we will get nothing but Heads. Why? – Think of each block of $A+B$ tosses as a single success-failure experiment, with *success* being the event that we get nothing but Heads in that block of tosses. The probability of success is $p=(1/2)^{A+B}$, which, although possibly very small, is *positive*. Consequently, by the Law of Large Numbers, the limiting fraction of blocks that result in all Heads will (with probability one) be p, and so there must be at least one block where this occurs. Game over!

1.3. **The method of difference equations.** To solve the gambler's ruin problem, we will set up and solve a *difference equation*. As Xinyi and I play, my fortune will fluctuate, but up until the termination point T it will always be some integer x between 0 and $A+B$. Fix some such possible value x. If at some stage of the game my fortune is x, then my *conditional* probability of winning is the same as the *unconditional* probability of winning if I had started with x dollars and Xinyi with $A+B-x$, because the future coin tosses (after we have reached a stage where my fortune is x) are independent of the coin tosses that took us to the current stage of the game. With this in mind, define

$$u(x)=P(S_T=A+B\,|\,\text{my initial fortune is } x)=P(\text{I win}\,|\,\text{my initial fortune is } x).$$

Our original problem was to find $u(A)$, so if we can find a formula for every $u(x)$ we will have solved our original problem.

Our strategy (that is, our strategy for solving the problem, not my strategy for somehow cheating Xinyi out of his money – I still haven't figured out how to do this) will be to *find a relation among the different values* $u(x)$ and then to exploit this relation to figure out $u(x)$. The idea is this: if I start with x dollars and Xinyi with $A+B-x$, where x is some value between 1 and $A+B-1$, then at least one toss will be needed to settle the outcome. After this first toss I will have either $x-1$ or $x+1$ dollars. Each of these two possibilities occurs with probability 1/2; thus, by the law of total probability and the multiplication rule for conditional probability,

$$\begin{aligned} u(x) &= P(\text{I win}\,|\,\text{first toss H})P(\text{first toss H})+P(\text{I win}\,|\,\text{first toss T})P(\text{first toss T}) \\ &= \frac{1}{2}u(x-1)+\frac{1}{2}u(x+1). \end{aligned}$$

What happens at $x=0$ and $x=A+B$? – This is easy: if I start with $x=A+B$ dollars, then I have already won, with no coin tosses needed, so $u(A+B)=1$. On the other hand, if I start with $x=0$ then I have already lost, so $u(0)=0$. This shows that the function $u(x)$ satisfies the *difference equation*

$$u(x)=\frac{1}{2}u(x-1)+\frac{1}{2}u(x+1) \tag{4}$$

and the *boundary conditions*

$$u(0)=0 \quad \text{and} \quad u(A+B)=1. \tag{5}$$

1.4. **Solving the difference equation.** The boundary value problem we have just formulated is one of the simplest in the book, because it implies that there is a very simple relation between the *successive differences* $u(x+1)-u(x)$ and $u(x)-u(x-1)$. After multiplying both sides of (4) by 2, we obtain

$$\begin{aligned} 2u(x)=u(x-1)+u(x+1) \quad &\Longrightarrow \\ u(x)-u(x-1)=u(x+1)-u(x). \end{aligned}$$

This last equation implies that the successive line segments connecting points $(x, u(x))$ to $(x+1, u(x+1))$ all have the same slopes, and so all of the points $(x, u(x))$ lie on a straight line. Consequently,

$$u(x) = C + Dx$$

for some constants C, D. The boundary conditions (5) determine these unknown constants, because they imply that the points $(0,0)$ and $(A+B, 1)$ both lie on the line. Using this, we conclude that $C = 0$ and $D = 1/(A+B)$. Plugging in the value $x = A$ gives the solution to the original gambler's ruin problem:

$$P\{S_{T_{A,B}} = B\} = \frac{A}{A+B}. \tag{6}$$

2. Simple Random Walk

The sequence $0, S_1, S_2, \dots$ of successive partial sums of the random variables X_i with the coin-tossing ± 1 distribution (1) is known as *simple random walk* in one dimension. By design, the simple random walk evolves as follows: (i) start at $S_0 = 0$; (ii) at each step, if the current state is $S_n = x$, then choose one of the nearest neighbors $x-1$ or $x+1$ at random, and jump there. This recipe can be generalized to any lattice $\mathbb{Z}^d$ (the set of points in d dimensions with integer coordinates). In dimension $d = 2$, for instance, each lattice point (x, y) has 4 nearest neighbors

$$\begin{aligned}&(x+1, y),\\ &(x-1, y),\\ &(x, y+1),\\ &(x, y-1);\end{aligned}$$

the simple random walk in $\mathbb{Z}^2$ at each step chooses one of the 4 nearest neighbors of its current location at random, independently of all its previous moves, and then jumps there.

Problem: If the simple random walk on $\mathbb{Z}^d$ starts at the origin, what is the probability that it will eventually return to the origin?

If the probability is 1 (i.e., if the simple random walk is certain to return to its initial point) then the simple random walk is said to be *recurrent*; otherwise, it is said to be *transient*.

Problem: For which values of d is simple random walk recurrent?

This problem was solved by George Polya around 1920; his solution was published in a celebrated paper in the 1921 Math. Annalen. The answer, he found, is that recurrence holds only in dimensions $d = 1$ and $d = 2$; in all other dimensions, simple random walk is transient, that is, there is positive probability that it does not return to its initial point.

So who cares?

Proof of Polya's theorem in $d = 1$. I will begin by showing something slightly different, specifically, if a one-dimensional simple random walk is started at $S_0 = 0$ then the probability that it will eventually visit the state $+1$ is 1. To accomplish this, it suffices to show that for any value of $\varepsilon > 0$, no matter how small, the probability that simple random walk eventually visits the state $+1$ is at least $1-\varepsilon$. Fix a (large) integer $A \geq 1$. Our solution ot the gambler's ruin problem shows that the probability that the simple random walk will hit $+1$ before it hits $-A$ is $A/(A+1)$. If A

is sufficiently large then $A/(A+1) > 1-\varepsilon$, so it follows that the probability that simple random walk will eventually visit the state +1 is at least $1-\varepsilon$.

This proves that simple random walk must (with probability one) eventually visit the state +1. Virtually the same argument shows that it must also visit the state −1. The random walker cannot visit both +1 and −1 at the same time, so it must visit one before the other. But to get from one to the other, the random walker must pass through the state 0. Therefore, the probability of return is 1.

□

3. Gambler's Ruin: Expected Duration of the Game

Let's return to the original game ruin problem, where Xinyi and I bet 1 dollar at a time until one of us is broke. The number T of tosses needed until the game ends is obviously random. What is its expectation ET?

Once again our strategy will be to derive a difference equation relating the expected durations under different initial conditions x. If I started with x dollars and Xinyi with $A+B-x$, then the game would end after $T_{x,A+B-x}$ tosses, where

$$T_{x,A+B-x} = \min\{n \geq 0 : S_n = -x \text{ or } S_n = A+B-x\}$$

and S_n is the net change in my fortune after n tosses. Define

$$v(x) = E\,T_{x,A+B-x}.$$

Clearly, $v(0) = 0$ and $v(A+B) = 0$, because if I start with either 0 or $A+B$ then no tosses are needed — the game is already over. The values $v(0) = v(A+B) = 0$ are the *boundary values* for the problem. To obtain the *difference equation*, we reason as follows: if $0 < x < A+B$, at least one coin toss is needed. There are two possibilities for this toss: Heads, in which case I will have $x-1$ dollars, or Tails, in which case I will have $x+1$ dollars. Once this first toss is made, the game will start anew, but with my initial fortune now either $x+1$ or $x-1$, Therefore,

$$v(x) = 1 + \frac{1}{2}v(x-1) + \frac{1}{2}v(x+1) \quad \forall\, 1 \leq x \leq A-1. \tag{7}$$

The new feature is the additional term 1 on the right, which occurs because the first toss adds 1 to the duration of the game. This additional term makes the difference equation *inhomogeneous.*

To solve the boundary value problem, we set $d(x) = v(x) - v(x-1)$ and multiply each side of the equation the equation (7) by 2; then after rearrangement of terms the equation becomes

$$d(x) = d(x+1) + 2.$$

It's not hard to see what to do with this equation: for each x, add −2 to the current value to get the next one, so for each $k = 1, 2, \ldots, A+B$,

$$d(k) = d(0) - 2k. \tag{8}$$

The value of $d(0)$ is still unknown, but eventually we will use the two boundary conditions to pin it down. Next, given the formula (8) for the differences $d(k)$, we can work backwards to obtain a corresponding formula for the function $v(x)$:

$$v(k)-v(0)=\sum_{j=1}^{k} d(j)=kd(0)-2\sum_{j=1}^{k} j=kd(0)-k(k+1).$$

Since $v(0)=0$ (the first boundary condition), this leaves us with

$$v(k)=kd(0)-k(k+1).$$

Now the second boundary condition requires that $v(A+B)=0$, so

$$\begin{aligned}(A+B)d(0)&=(A+B)(A+B+1) \quad\Longrightarrow\\ d(0)&=(A+B+1),\end{aligned}$$

and so

$$v(k)=k(A+B+1)-k(k+1).$$

When $k=A$, as in the original game, the formula simplifies to

$$\boxed{ET=v(A)=AB.} \tag{9}$$

4. Biased Random Walk

4.1. **The model.** What if the coin used in the gambler's ruin game is not fair? How are the answers to the problems studied in sections 1,2,3 above affected?

To state the problem formally, let $X_1, X_2, \ldots$ be independent, identically distributed random variables with common distribution

$$\begin{aligned}P\{X_i=+1\}&=p,\\ P\{X_i=-1\}&=q:=1-p,\end{aligned} \tag{10}$$

and set $S_n=\sum_{i=1}^{n} X_i$, as in section 1. The sequence of random variables $0=S_0, S_1, S_2, \ldots$ is known as the p,q random walk, or more simply, *biased random walk*. As before, define

$$T=T_{A,B}=\min\{n\geq 0: S_n=-A \text{ or } S_n=+B\}$$

to be the first exit time from the interval $(-A,+B)$. The problems of interest are the same as for the case $p=1/2$:

(a) What is the probability that $S_T=B$?
(b) What is the expected exit time ET?

4.2. **Ruin probability.** First, consider Problem (a). As in the gambler's ruin problem with a fair coin, let $u(x)$ be the probability that I win if I start with x dollars and Xinyi starts with $A+B-x$. The boundary conditions are the same:

$$u(0)=0 \quad \text{and} \quad u(A+B)=1.$$

The *difference equation* is different, though, because the probability that the first toss brings my fortune to $x+1$ is now p, not $1/2$. Therefore, the appropriate difference equation is

$$u(x)=q\,u(x-1)+p\,u(x+1) \quad \text{for } 1\leq x\leq A+B-1, \tag{11}$$

by the same logic as we used to deduce equation (4) in the case of a fair coin.

To solve the difference equation (11), we will try (as we did to solve the boundary value problems in sections 1 and 3) to use (11) to derive a relation between successive differences

$d(x) = u(x) - u(x-1)$ and $d(x+1) = u(x+1) - u(x)$. Subtract $q\,u(x-1)$ from both sides of (11), then subtract $p\,u(x)$ from both sides; this gives

$$q(u(x) - u(x-1)) = p(u(x+1) - u(x)) \quad \Longrightarrow q\,d(x) = p\,d(x+1).$$

Thus, each $d(x+1)$ can be obtained by multiplying the previous difference $d(x)$ by q/p, and so

$$d(x) = d(1)\left(\frac{q}{p}\right)^{x-1}.$$

The value of $d(1) = u(1)$ is still unknown, but will eventually be determined by the second boundary condition $u(A+B) = 1$. We now know that the successive differences $d(x)$ form a geometric sequence with ratio q/p. Since the values $u(x)$ can be recovered by adding up the successive differences, we have (using again the boundary condition $u(0) = 0$)

$$u(x) = \sum_{k=1}^{x} d(k) = \sum_{k=1}^{x} d(1)\left(\frac{q}{p}\right)^{k-1}.$$

Summing the geometric series, using the assumption that $p \neq 1/2$ and the fact that $d(1) = u(1)$, yields

$$u(x) = u(1)\frac{1-(q/p)^x}{1-(q/p)}.$$

Finally, the boundary condition $u(A+B-1)$ implies that

$$1 = u(1)\frac{1-(q/p)^{A+B}}{1-(q/p)} \quad \Longrightarrow \quad u(1) = \frac{1-(q/p)}{1-(q/p)^{A+B}}.$$

Therefore,

$$\boxed{u(x) = \frac{1-(q/p)^x}{1-(q/p)^{A+B}}.} \tag{12}$$

4.3. **Wald's Identity.** To solve for the expected exit time ET we could try to write a difference equation and then solve it, as we have done for the other gamber's ruin problems. But there is another approach that relies more on "probabilistic" — as opposed to algebraic — reasoning, and this is the route we'll take here.

Recall that $S_1, S_2, \ldots$ is the sequence of partial sums of the increments X_i, where each random variable X_i has the distribution (10). The game ends at the random time $T = \min\{n : S_n = 0 \text{ or } S_n = A+B\}$. Clearly, for every $n = 1, 2, \ldots$,

$$ES_n = nEX_1 = n(p-q).$$

If we were to substitute the random time T for the nonrandom time n in this equation, the resulting formula would no longer make sense, because the quantity on the left side ES_T is a constant, but the quantity on the right side $T(p-q)$ would be a random variable. However, it turns out that if we then take expectations, we obtain an identity which is *true*.

Theorem 1. $ES_T = (p-q)ET$.

This is a special case of what is now known as *Wald's Identity*. Before looking at the proof of this identity, let's see how it can be used to determine the expected exit time ET when $p \neq 1/2$. For this, it suffices to determine the expectation ES_T. Now the random variable S_T takes only two possible values, 0 or $A+B$. We have already determined the probabilities of these two values, in equation (12). Using this, we have (assuming I start with $x = A$ dollars and Xinyi with B dollars)

$$ES_T = 0 \cdot P\{S_T = 0\} + (A+B)P\{S_T = A+B\} = (A+B)\frac{1-(q/p)^A}{1-(q/p)^{A+B}}.$$

Wald's identity now gives

$$ET = ES_T/(p-q) = \frac{A+B}{p-q}\frac{1-(q/p)^A}{1-(q/p)^{A+B}}.$$

Proof. We will use the following notation: for any event B, the random variable 1_B is the random variable that takes the value 1 on B (i.e., if the event B occurs) and takes the value 0 on the complement B^c. With this notation, we have

$$T = \sum_{m=1}^{\infty} 1_{\{T \geq m\}} \quad \text{and}$$

$$S_T = \sum_{m=1}^{\infty} X_m 1_{\{T \geq m\}}.$$

Taking expectations in both of these equations, we obtain

$$ET = \sum_{m=1}^{\infty} E(1_{\{T \geq m\}}) \quad \text{and}$$

$$ES_T = \sum_{m=1}^{\infty} E(X_m 1_{\{T \geq m\}}).$$

(Note: Since the sums on the right are both infinite, passing the expectation operator E under the summation sign requires some justification, but we will omit this.)

Now the key observation is this: for each $m = 1, 2, \dots$, the random variables X_m and $1_{\{T \geq m\}}$ are *independent*. Why? — The event $\{T \geq m\}$ can be rewritten as $\{T > m-1\}$, and whether or not this occurs is completely determined by the random variables $S_1, S_2, \dots, S_{m-1}$. These are functions of the first $m-1$ coin tosses, which by assumption are independent of the mth toss, which determines X_m; therefore, X_m is independent of $1_{\{T \geq m\}}$.

Once we know that X_m and $1_{\{T \geq m\}}$ are independent, we can use the product rule

$$E(X_m 1_{\{T \geq m\}}) = EX_m \cdot E1_{\{T \geq m\}} = (p-q)E1_{\{T \geq m\}}.$$

Consequently,

$$\begin{aligned} ES_T &= \sum_{m=1}^{\infty} E(X_m 1_{\{T \geq m\}}) \\ &= \sum_{m=1}^{\infty} (p-q)E1_{\{T \geq m\}}) \\ &= (p-q)\sum_{m=1}^{\infty} E1_{\{T \geq m\}}) \\ &= (p-q)ET. \end{aligned}$$

□

5. Simple Random Walk: First Passage Time Distribution

5.1. **First-Passage Times.** In this section we return to the *simple random walk*, that is, the sequence of sums $S_1, S_2, S_3, \dots$ of increments X_i with distribution $P\{X_i = 1\} = P\{X_i = -1\} = 1/2$. We have shown that simple random walk on the integers is recurrent, and in particular that if started in initial state $S_0 = 0$ it will, with probability 1, reach the level m, for any integer m, in finite (but random) time. Let $\tau(m)$ be the first passage time, that is,

$$\tau(m) := \min\{n \geq 0 : S_n = m\}, \tag{13}$$

and write $\tau = \tau(1)$. What can we say about the distribution of $\tau(m)$?

Consider first the *expectation* $E\tau(1)$. To reach the value $+1$, the simple random walk S_m must first exit the interval $(-A, 1)$; therefore, for any value A, no matter how large,

$$\tau(1) \geq T_{A,1}$$

where $T_{A,1}$ is the first time n such that $S_n = -A$ or $S_n = 1$. Now the expectation of $T_{A,1}$ is $E\,T_{A,1} = A$, by equation (9). The expectation of $\tau(1)$ must be at least this large. But since A is completely arbitrary, the expectation $E\tau(1)$ must be larger than any finite integer A, and so

$$E\tau(1) = +\infty. \tag{14}$$

For any value $m = 2, 3, \dots$, the number of tosses $\tau(m)$ needed until the first visit to the state m is larger than the number $\tau(1)$ needed to reach state 1, because the random walk must visit $+1$ before it can reach $+m$. Consequently, for every $m \geq 1$,

$$E\tau(m) = \infty.$$

5.2. **Probability Generating Function.** We will now look for the exact distribution of the random variable $\tau(1)$. To simplify the notation, we will write $\tau = \tau(1)$ throughout this derivation.

To get at the distribution of τ we'll look at its probability generating function

$$F(z) := Ez^\tau = \sum_{n=1}^{\infty} z^n P\{\tau = n\}. \tag{15}$$

This is defined for all real values of z less than or equal to 1 in absolute value. If we can obtain a formula for $F(z)$, then in principle we can recover all of the probabilities $P\{\tau = n\}$ by taking derivatives and setting $z = 1$.

The strategy for doing so is to condition on the first step of the random walk to obtain a functional equation for F. There are two possibilities for the first step: either $S_1 = +1$, in which case $\tau = 1$, or $S_1 = -1$. On the event that $S_1 = -1$, the random walk must first return to 0 before it can reach the level $+1$. But the amount of time τ' it takes to reach 0 starting from -1 has the same distribution as τ; and upon reaching 0, the amount of additional time τ'' to reach $+1$ again has the same distribution as τ, and is conditionally independent of the time taken to get from -1 to 0. Therefore,

$$F(z) = \frac{z}{2} + \frac{z}{2} E z^{\tau' + \tau''}, \tag{16}$$

where τ', τ'' are independent random variables each with the same distribution as τ. Now the multiplication rule for expectations implies that $Ez^{\tau'+\tau''} = Ez^{\tau'} Ez^{\tau''}$, and since each of τ' and τ'' has the same distribution as τ, it follows that

$$F(z) = (z + zF(z)^2)/2. \tag{17}$$

This is a quadratic equation in the unknown $F(z)$: the solution is $F(z) = (1 \pm \sqrt{1-z^2})/z$. But which is it: $\pm$? For this, observe that $F(z)$ must take values between 0 and 1 when $0 < z < 1$. It is a routine calculus exercise to show that only one of the two possibilities has this property, and so

$$\boxed{F(z) = \frac{1-\sqrt{1-z^2}}{z}} \tag{18}$$

Now we must somehow use this formula to recover the probabilities $P\{\tau = n\}$. This is done by taking successive derivatives, as equation (15) shows that these are (up to multiplication by $n!$) the probabilities $P\{\tau = n\}$. Alternatively, we can exploit the fact that Sir Isaac Newton did the calculation for us about 350 years ago. Newton discovered, in the process, that the binomial formula extends to fractional powers; in particular, for the square root function,

$$\sqrt{1-z^2} = \sum_{n=0}^{\infty} \binom{1/2}{n} (-z^2)^n, \tag{19}$$

where

$$\binom{1/2}{n} = \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(-\frac{2n-3}{2}\right)\Big/ n!.$$

Thus, after a small bit of unpleasant algebra, we obtain

$$P\{\tau = 2n-1\} = (-1)^{n-1}\binom{1/2}{n}. \tag{20}$$

Exercise 1. Verify that $P\{\tau = 2n-1\} = 2^{2n-1}(2n-1)^{-1}\binom{2n-1}{n}$. This implies that

$$P\{\tau = 2n-1\} = P\{S_{2n-1} = 1\}/(2n-1). \tag{21}$$

5.3. **Reflection Principle and First-Passage Distributions.** There is a completely different combinatorial approach to finding the distribution of the first-passage time $\tau(m)$. This uses the *Reflection Principle*, according to which a simple random walk path reflected in the line $y = m$ is still a simple random walk path. Here is a precise formulation: Let S_n be a simple random walk started at $S_0 = 0$, and let $\tau(m)$ be the first time that it reaches the state $m \ge 1$. Define a new path S_n^* by

$$\begin{aligned} S_n^* &= S_n && \text{if } n \le \tau(m); \\ S_n^* &= 2m - S_n && \text{if } n \ge \tau(m). \end{aligned} \tag{22}$$

See Figure 1 below for an example.

At first sight the Reflection Principle might look mysterious, but there is a way of viewing it so that it becomes totally transparent. Imagine that the successive steps $X_n = S_n - S_{n-1}$ of the random walk (the blue path in Figure 1) are determined by fair coin tosses, with Heads dictating a step of size $+1$ and Tails a step of size -1. Now imagine that a second random walker uses the same sequence of coin tosses to determine his/her steps, but after reaching the level $+m$ for the first time *reverses* the roles of Heads and Tails, that is, after time $\tau(m)$ the random walk makes steps of -1 on every Head and$+1$ on every Tail. The path $\{S_n^*\}_{n\ge 0}$ of this random walk is the reflected path (the red path in Figure 1). It should be relatively obvious that flipping the roles of

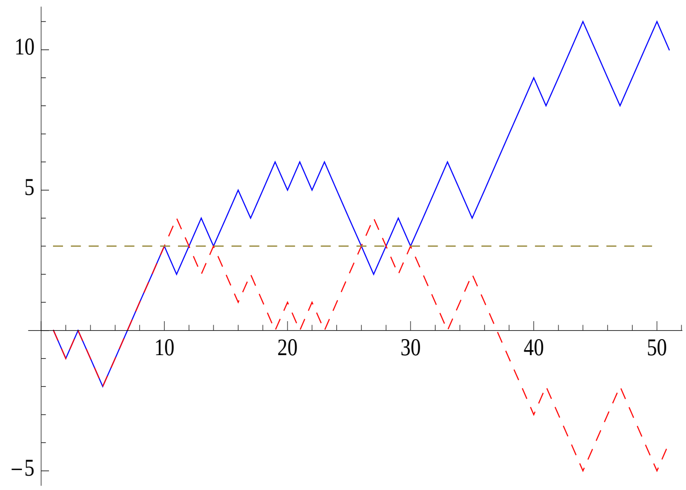

FIGURE 1. The Reflection Principle

Heads and Tails makes no difference when computing probabilities, so the new random path $\{S_n^*\}_{n\geq 0}$ is once again a simple random walk.[1] We record this fact as

Theorem 2. *(Reflection Principle) The sequence $\{S_n^*\}_{n\geq 0}$ is a simple random walk started at* 0.

Now consider the event $\tau(m) \leq n$. On this event, S_n and S_n^* are on opposite sides of m, unless they are both at m, and they correspond under reflection. Moreover, both processes are simple random walks, so for any $k \geq 0$,

$$P\{S_n^* = m+k\} = P\{S_n = m+k\}.$$

If $k \geq 0$, the event $S_n = m+k$ is impossible unless $\tau(m) \leq n$, so

$$\{S_n = m+k\} = \{S_n = m+k \text{ and } \tau(m) \leq n\}.$$

Hence,

$$\begin{aligned} P\{S_n = m+k\} &= P\{S_n = m+k \text{ and } \tau(m) \leq n\} \\ &= P\{S_n^* = m+k \text{ and } \tau(m) \leq n\} \\ &= P\{S_n = m-k \text{ and } \tau(m) \leq n\}. \end{aligned}$$

The event $\{\tau(m) \leq m\}$ is the disjoint union of the events $\{\tau(m) \leq m \text{ and } S_m = m+k\}$, where k ranges over the integers, and so

$$P\{\tau(m) \leq n\} = \sum_{k=-\infty}^{\infty} P\{S_n = m+k \text{ and } \tau(m) \leq n\} = P\{S_n = m\} + 2P\{S_n > m\}.$$

Exercise 2. Use this identity to derive the formula in exercise 1 for the density of $\tau(1)$. Derive a similar formula for $P\{\tau(m) = 2n-1\}$.

[1]Caution: It is not, however, *independent* of the original random walk. In fact the two random walks are highly correlated.