ONE-DIMENSIONAL RANDOM WALKS

1. Simple Random Walk

Definition 1. A *random walk* on the integers $\mathbb{Z}$ with step distribution F and initial state $x \in \mathbb{Z}$ is a sequence S_n of random variables whose increments are independent, identically distributed random variables ξ_i with common distribution F, that is,

$$S_n = x + \sum_{i=1}^{n} \xi_i. \tag{1}$$

The definition extends in an obvious way to random walks on the $d-$dimensional integer lattice $\mathbb{Z}^d$: the increments are then random $d-$vectors. *Simple random walk* on $\mathbb{Z}^d$ is the particular case where the step distribution is the uniform distribution on the $2d$ nearest neighbors of the origin; in one dimension, this is the *Rademacher*-$\frac{1}{2}$ distribution, the distribution that puts mass $1/2$ at each of the two values ± 1. The moves of a simple random walk in 1D are determined by independent fair coin tosses: For each Head, jump one to the right; for each Tail, jump one to the left.

1.1. Gambler's Ruin.

Simple random walk describes (among other things) the fluctuations in a speculator's wealth when he/she is fully invested in a risky asset whose value jumps by either ± 1 in each time period. Although this seems far too simple a model to be of any practical value, when the unit of time is small (e.g., seconds) it isn't so bad, at least over periods on the order of days or weeks, and in fact it is commonly used as the basis of the so-called *tree models* for valuing options.

Gambler's Ruin Problem: Suppose I start with x dollars. What is the probability that my fortune will grow to A dollars before I go broke? More precisely, if

$$T = T_{[0,A]} := \min\{n : S_n = 0 \text{ or } A\} \tag{2}$$

then what is $P^x\{S_T = A\}$?[1] Before we try to answer this, we need to verify that $T < \infty$ with probability 1. To see that this is so, observe that if at any time during the course of the game there is a run of A consecutive Heads, then the game must end, because my fortune will have increased by at least A dollars. But if I toss a fair coin forever, a run of A consecutive Heads will certainly occur. (Why?)

Difference Equations: To solve the gambler's ruin problem, we'll set up and solve a *difference equation* for the quantity of interest

$$u(x) := P^x\{S_T = A\}. \tag{3}$$

First, if I start with A dollars then I have already reached my goal, so $u(A) = 1$; similarly, $u(0) = 0$. Now consider what happens on the very first play, if $0 < x < A$: either I toss a Head, in which case

[1] Here and throughout the course, the superscript x denotes the initial state of the process S_n. When there is no superscript, the initial state is $x = 0$. Thus, $P = P^0$.

my fortune increases by 1, or I toss a tail, in which case it decreases by 1. At this point, it is like starting the game from scratch, but with initial fortune either $x+1$ or $x-1$. Hence, u satisfies the difference equation

$$u(x) = \frac{1}{2}u(x+1) + \frac{1}{2}u(x-1) \quad \forall\, 1 \le x \le A-1 \tag{4}$$

and the boundary conditions

$$\begin{aligned} u(A) &= 1; \\ u(0) &= 0. \end{aligned} \tag{5}$$

How do we solve this? The most direct approach is to translate the difference equation into a relation between the successive differences $u(x+1) - u(x)$ and $u(x) - u(x-1)$:

$$\frac{1}{2}(u(x+1) - u(x)) = \frac{1}{2}(u(x) - u(x-1)). \tag{6}$$

This equation says that the successive differences in the function u are all the same, and it is easy to see (exercise!) that the only functions with this property are *linear* functions $u(x) = Bx + C$. Conversely, any linear function solves (4). To determine the coefficients B, C, use the boundary conditions: these imply $C = 0$ and $B = 1/A$. This proves

Proposition 1. $P^x\{S_T = A\} = x/A$.

Remark 1. We will see later in the course that first-passage problems for Markov chains and continuous-time Markov processes are, in much the same way, related to boundary value problems for other difference and differential operators. This is the basis for what has become known as *probabilistic potential theory*. The connection is also of practical importance, because it leads to the possibility of *simulating* the solutions to boundary value problems by running random walks and Markov chains on computers.

Remark 2. In solving the difference equation (4) , we used it to obtain a relation (6) between successive differences of the unknown function u. This doesn't always work. However, in general, if a difference equation is of order m, then it relates $u(x)$ to the last m values $u(x-1), \dots, u(x-m)$. Thus, it relates the *vector*

$$\begin{aligned} U(x) &:= (u(x), u(x-1), \dots, u(x-m+1)) \quad \text{to the vector} \\ U(x-1) &:= (u(x-1), u(x-2), \dots, u(x-m)). \end{aligned}$$

If the difference equation is *linear*, as is usually the case in Markov chain problems, then this relation can be formulated as a matrix equation $MU(x-1) = U(x)$. This can then be solved by matrix multiplication. Following is a simple example where this point of view is helpful.

Expected Duration of the Game: Now that we know the probabilities of winning and losing, it would be nice to know how long the game will take. This isn't a well-posed problem, because the duration T of the game is random, but we can at least calculate $E^x T$. Once again, we will use difference equations: Set

$$v(x) := E^x T; \tag{7}$$

then $v(0) = v(A) = 0$ and, by reasoning similar to that used above,

$$v(x) = 1 + \frac{1}{2}v(x-1) + \frac{1}{2}v(x+1) \quad \forall\, 1 \le x \le A-1. \tag{8}$$

The new feature is the additional term 1 on the right — this makes the equation *inhomogeneous*. To solve this, we'll convert the equation to a matrix equation. Set $d(x) = v(x) - v(x-1)$; then after multiplication by 2 the equation (8) becomes

$$\begin{pmatrix} d(x+1) \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d(x) \\ -2 \end{pmatrix},$$

and so

$$\begin{pmatrix} d(m) \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{m-1} \begin{pmatrix} d(1) \\ -2 \end{pmatrix}$$

Exercise 1. Check that

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{m} = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$$

Given Exercise 1, we can now conclude that $d(m) = d(1) - 2(m-1)$. Since by definition $d(m) = v(m) - v(m-1)$ and $v(0) = 0$, it follows that $d(1) = v(1)$ and

$$v(m) = \sum_{k=1}^{m} d_k = m v(1) - 2 \sum_{j=1}^{m-1} j = m v(1) - m(m-1).$$

The value $v(1) = A - 1$ is now forced by the second boundary condition $v(A) = 0$. This proves

Proposition 2. $E^x T = m(A - m)$.

Exercise 2. Consider the $p-q$ random walk on the integers, that is, the random walk whose step distribution is $P\{\xi_1 = +1\} = p$ and $P\{\xi_1 = -1\} = q$ where $p + q = 1$. Solve the gambler's ruin problem for $p-q$ random walk by setting up and solving a difference equation. (Reformulate the difference equation as a matrix equation, and use this to represent the solution as a matrix multiplication. To get a simple formula for the matrix product, *diagonalize* the matrix.)

1.2. **Recurrence of Simple Random Walk.** The formula for the ruin probability (Proposition 1) has an interesting qualititative consequence. Suppose we start a simple random walk at some integer x. By Proposition 1, the probability that we reach 0 before hitting A is $1 - x/A$, and so the probability that we will eventually reach state 0 is at least $1 - x/A$. But this is true for every value of $A > x$; sending $A \to \infty$ shows that

$$P^x\{\text{reach 0 eventually}\} = 1. \tag{9}$$

Clearly, if S_n is a random walk that starts at x, then for any integer y the process $S_n + y$ is a random walk that starts at $x + y$; hence, hitting probabilities are invariant under translations. Similarly, they are invariant under reflection of the integer lattice through 0 (that is, changing x to $-x$), because reversing the roles of Heads and Tails doesn't change the probability of any event. Therefore, (9) implies that for any two integers x, y,

$$P^x\{\text{reach y eventually}\} = 1. \tag{10}$$

Define

$$\nu_y = \nu(y) = \min\{n : S_n = y\}$$

to be the first passage time to state y. We have just shown that, regardless of the initial point, $\nu_y < \infty$ with probability one. Now of course ν_y is random, but since the coin tosses after time ν_y are unaffected by the course of the random walk up to time ν_y, it seems clear, intuitively, that the random walk "restarts" at its first visit to y. The next definition abstracts the essential property of the random time ν_y that justifies this.

Definition 2. A *stopping time* for the random walk S_n is a nonnegative integer-valued random variable τ such that for every integer $n \geq 0$ the indicator function of the event $\{\tau = n\}$ is a (measurable)[2] function of $S_1, S_2, \dots, S_n$.

Proposition 3. *(Strong Markov Property) If τ is a stopping time for a random walk $\{S_n\}_{n\geq 0}$, then the post-τ sequence $\{S_{\tau+j}\}_{j\geq 0}$ is also a random walk, with the same step distribution, started at S_τ, and is independent of the random path $\{S_j\}_{j\leq\tau}$.*

Proof. Exercise. Hint: What you must show is that for any two sequences $\{\omega_j\}$ and $\{\omega_j^*\}$ of ± 1, and for all positive integers k, m,

$$P^x(\{\xi_j = \omega_j \ \forall j \leq k\} \cap \{\tau = k\} \cap \{\xi_{k+j} = \omega_j^* \ \forall j \leq m\})$$
$$= P^x(\{\xi_j = \omega_j \ \forall j \leq k;\} \cap \{\nu(y) = k\}) P^y\{\xi_j = \omega_j^* \ \forall j \leq m\}.$$

□

The first-passsage times ν_y are clearly stopping times. Consequently, by Proposition 3, the post-ν_y process is just an independent simple random walk started at y. But (10) (with the roles of x, y reversed) implies that this random walk must eventually visit x. When this happens, the random walk restarts again, so it will go back to y, and so on. Thus, by an easy induction argument (see Corollary 14 below):

Theorem 4. *With probability one, simple random walk visits every state y infinitely often.*

1.3. **First-Passage Time Distribution.** We now know that simple random walk on the integers is recurrent, and in particular that if started in initial state $S_0 = 0$ will reach the level m, for any integer m, in finite (but random) time. Let $\tau(m)$ be the first passage time, that is,

$$\tau(m) := \min\{n \geq 0 : S_n = m\}, \tag{11}$$

and write $\tau = \tau(1)$. What can we say about the distribution of $\tau(m)$? Suppose $m \geq 1$; then to reach m, the random walk must first reach $+1$, so $\tau(m) \geq \tau$. At this time, the random walk restarts (Proposition 3). The *additional* time needed to reach m has the same distribution as $\tau(m-1)$, and is independent of τ. Consequently, $\tau(m)$ is the sum of m independent copies of τ.

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

$$F(z) := Ez^\tau = \sum_{n=1}^{\infty} z^n P\{\tau = n\}. \tag{12}$$

This is defined for all real values of z less than 1 in absolute value. By elementary rules governing independence and generating functions, the probability generating function of $\tau(m)$ is $F(z)^m$, so if we can find $F(z)$ then we'll have a handle on the distributions of all the first passage times.

The strategy 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 it takes to reach 0 starting from -1 has the same distribution as τ; and upon reaching 0, the amount of additional time to reach $+1$ again has the same distribution

[2] Any reasonable function is measurable. Nonmeasurable functions exist only if you believe in the Axiom of Choice.

as τ, and is conditionally indepedent of the time taken to get from -1 to 0 (by Proposition 3). Therefore,

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

where τ', τ'' are independent random variables each with the same distribution as τ. Because the probability generating function of a sum of independent random variables is the product of their p.g.f.s, it follows that

(14) $$F(z) = (z + zF(z)^2)/2.$$

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

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

Consequences: First, F is continuous at $z = 1$, but not differentiable at $z = 1$; therefore, $E\tau = \infty$. (If a nonnegative random variable has finite expectation, then its probability generating function is differentiable at $z = 1$, and the derivative is the expectation.) Second, the explicit formula (15) allows us to write an explicit expression for the discrete density of τ. According to Newton's binomial formula,

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

and so, after a small bit of unpleasant algebra, we obtain

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

Exercise 3. Verify that $P\{\tau = 2n-1\} = 2^{2n-1}(2n-1)^{-1}\binom{2n-1}{n}$. This implies that

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

Exercise 4. Show that $P\{\tau = 2n-1\} \sim C/n^{3/2}$ for some constant C, and identify C.

Remark 3. Exercise 4 asserts that the density of τ obeys a *power law* with exponent 3/2.

Exercise 5. (a) Show that the generating function $F(z)$ given by equation (15) satisfies the relation

(19) $$1 - F(z) \sim \sqrt{2}\sqrt{1-z} \qquad \text{as } z \to 1-.$$

(b) The random variable $\tau(m) = \min\{n : S_n = m\}$ is the sum of m independent copies of $\tau = \tau(1)$, and so its probability generating function is the nth power of $F(z)$. Use this fact and the result of part (a) to show that for every real number $\lambda > 0$,

(20) $$\lim_{m \to \infty} E \exp\{-\lambda \tau(m)/m^2\} = e^{-\sqrt{2\lambda}}$$

Remark 4. The function $\varphi(\lambda) = \exp\{-\sqrt{2\lambda}\}$ is the Laplace transform of a probability density called the *one-sided stable law of exponent* 1/2. You will hear more about this density in connection with Brownian motion later in the course. The result of exercise 2b, together with the *continuity theorem* for Laplace transforms, implies that the rescaled random variables $\tau(m)/m^2$ converge in distribution to the one-sided stable law of exponent 1/2.

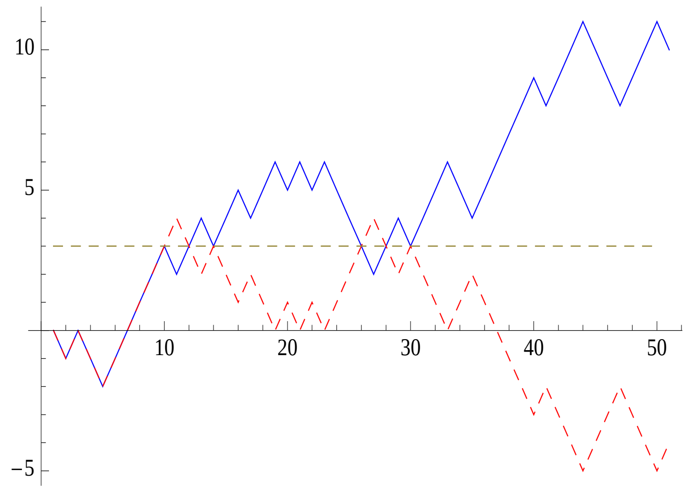

FIGURE 1. The Reflection Principle

1.4. **Reflection Principle and First-Passage Distributions.** There is another approach to finding the distribution of the first passage time $\tau(m)$ that does not use generating functions. This is based on 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 $m \geq 1$. Define a new path S_n^* by

$$
\begin{aligned}
S_n^* &= S_n && \text{if } n \leq \tau(m); \\
S_n^* &= 2m - S_n && \text{if } n \geq \tau(m).
\end{aligned}
\tag{21}
$$

See Figure 1.4 for an example.

Proposition 5. *(Reflection Principle) The sequence $\{S_n^*\}_{n\geq 0}$ is a simple random walk started at* 0.

Proof. Exercise. HINT: The path S_n^* is what you get if you reverse the roles of Heads and Tails after reaching m. □

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,

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

and so

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

Exercise 6. Use this identity to derive the formula in exercise 3 for the density of $\tau(1)$. Derive a similar formula for $P\{\tau(m) = 2n-1\}$.

1.5. **Skip-Free Random Walk and Lagrange Inversion.** There is a *third* approach (and also a fourth — see section 2.4 below) to determining the distribution of the first-passage time $\tau(1)$. Having already seen two derivations of the basic formula (18) you may already be inclined to believe that it is true, in which case you should feel free to skip this section. However, the approach developed here has the advantage that it works for a much larger class of random walks, called *skip-free*, or sometimes *right-continuous* random walks. A *skip-free* random walk is one whose step distribution puts no mass on integers ≥ 2. Equivalently,

$$\xi_n = 1 - Y_n$$

where $Y_1, Y_2, \dots$ are independent, identically distributed with common distribution

$$q_k := P\{Y_n = k\} \quad \text{for } k = 0, 1, 2, \dots$$

Let $Q(w) = \sum_{k \ge 0} q_k w^k$ be the generating function of Y_n, and let τ be the first passage time to the level 1 by the random walk $S_n = \sum_{j=1}^n \xi_j$.

Exercise 7. Show that the probability generating function $F(z) := Ez^\tau$ satisfies the functional equation

$$F(z) = zQ(F(z)). \tag{22}$$

NOTE: The random variable τ need not be finite with probability one. If this is the case, then interpret Ez^τ to mean $Ez^\tau \mathbf{1}\{\tau < \infty\}$, equivalently,

$$Ez^\tau := \sum_{n=1}^{\infty} z^n P\{\tau = n\}.$$

Exercise 8. Let $x_1, x_2, \dots, x_n$ be a sequence of integers ≤ 1 with sum 1. Show there is a unique cyclic permutation π of the integers $1, 2, \dots, n$ such that

$$\sum_{j=1}^{k} x_{\pi(j)} \le 0 \quad \forall\, k = 1, 2, \dots, n-1. \tag{23}$$

HINT: The trick is to guess where the cycle should begin. Try drawing a picture.

Exercise 9. Use the result of Exercise 8 to prove that

$$P\{\tau = n\} = n^{-1} P\{S_n = 1\}. \tag{24}$$

Exercise 10. The *Lagrange Inversion Formula* states that if $F(z) = \sum_{n=1}^{\infty} a_n z^n$ is a power series with no constant term that satisfies the functional equation (22) then

$$n a_n = (n-1)\text{th coefficient of the power series } Q(w)^n$$

Show that when $Q(w)$ is a probability generating function, this is equivalent to the result of Exercise 9.

2. The Wald Identities

2.1. **Stopping times.** Recall (Definition 2) that a stopping time for a random walk S_n is a nonnegative integer-valued random variable such that for every $n = 0, 1, 2, \dots$ the event that $\tau = n$ depends only on the values $S_1, S_2, \dots, S_n$, or, equivalently, on the values $\xi_1, \xi_2, \dots, \xi_n$. In general, first-passage times, or first times that some event of interest occurs, are stopping times. A nonrandom time n is trivially a stopping time. On the other hand, the *last* time that (say) a random walk visits the state 0 is not a stopping time.

Lemma 6. *If τ and ν are stopping times for a random walk S_n then so are $\tau \wedge \nu$ and $\tau + \nu$.*

Lemma 7. *If τ is a stopping time for a random walk S_n then for each nonnegative integer n the event $\{\tau \geq n\}$ depends only on $\xi_1, \xi_2, \dots, \xi_{n-1}$.*

Exercise 11. Supply the proofs.

Consequently, if τ is a stopping time, then for every nonnegative integer n the random variable $\tau \wedge n$ is also a stopping time. Hence, every stopping time is the increasing limit of a sequence of *bounded* stopping times.

2.2. **Wald Identities: Statements.** In the following statements, assume that S_n is a one-dimensional random walk with initial value $S_0 = 0$.

First Wald Identity . *Assume that the random variables ξ_j have finite first moment, and let $\mu = E\xi_1$. Then for any stopping time τ with finite expectation,*

$$ES_\tau = \mu E\tau. \tag{25}$$

Second Wald Identity . *Assume that the random variables ξ_j have finite second moment, and let $\mu = E\xi_1$ and $\sigma^2 = E(\xi_1 - \mu)^2$. Then for any stopping time τ with finite expectation,*

$$E(S_\tau - m\tau)^2 = \sigma^2 E\tau. \tag{26}$$

Third Wald Identity . *Assume that the moment generating function $\varphi(\theta) = Ee^{\theta\xi_1}$ of the random variables ξ_j is finite at the argument θ. Then for any bounded stopping time τ,*

$$E\left(\frac{\exp\{\theta S_\tau\}}{\varphi(\theta)^\tau}\right) = 1. \tag{27}$$

The hypothesis on the stopping time τ is stronger in the Third Wald Identity than in the first two. Later we will see an example where equation (27) fails even though $E\tau < \infty$. When $E\tau = \infty$, the Wald identities can fail in a big way:

Example 1. Let S_n be simple random walk on $\mathbb{Z}$ and let τ be the first time that the random walk visits the state 1. Then

$$1 = ES_\tau \neq \mu E\tau = 0 \times \infty.$$

2.3. **Proofs of Wald identities 1 and 3.** When you study martingales later you will learn that all three Wald identities are special cases of a general theorem about martingales, Doob's Optional Sampling Formula. But it's instructive to see direct proofs. Everyone should know:

Lemma 8. *For any nonnegative integer-valued random variable Y,*

$$EY = \sum_{n=1}^{\infty} P\{Y \geq n\}$$

Proof of the First Wald Identity. The essential idea is clearest in the special case where τ is bounded, say $\tau \leq M$ for some integer M. In this case, S_τ can be decomposed as a finite sum

$$S_\tau = \sum_{n=0}^{M} S_n \mathbf{1}\{\tau = n\} = \sum_{n=1}^{M} \xi_n \mathbf{1}\{\tau \geq n\}.$$

Since the sum is finite, there is no obstacle to moving the expectation operator E inside the sum, and so

$$ES_\tau = \sum_{n=1} E\xi_n \mathbf{1}\{\tau \geq n\}$$

But the event $\{\tau \geq n\}$ depends only on the first $n-1$ increments (Lemma 7), so it is independent of ξ_n. Consequently,

$$E\xi_n \mathbf{1}\{\tau \geq n\} = \mu P\{\tau \geq n\},$$

and so

$$ES_\tau = \mu \sum_{n=1}^{M} P\{\tau \geq n\} = \mu E\tau.$$

When τ is not bounded, the analogous decomposition of S_τ leaves us with an infinite sum, and passing expectations through infinite sums must be done with some care. Here it is possible to use either the DCT (dominated convergence theorem) or the Fubini-Tonelli theorem to justify the interchange. Let's try DCT: Since $|\xi_n|$ and $\mathbf{1}\{\tau \geq n\}$ are independent,

$$\sum_{n=1}^{\infty} E|\xi_n| \mathbf{1}\{\tau \geq n\} = \sum_{n=1}^{\infty} E|\xi_1| P\{\tau \geq n\} = E|\xi_1| E\tau < \infty.$$

Hence, by DCT,

$$\begin{aligned} ES_\tau &= E\sum_{n=1}^{\infty} \xi_n \mathbf{1}\{\tau \geq n\} \\ &= \sum_{n=1}^{\infty} E\xi_n \mathbf{1}\{\tau \geq n\} \\ &= \sum_{n=1}^{\infty} \mu P\{\tau \geq n\} \\ &= \mu E\tau. \end{aligned}$$

□

Proof of the Third Wald Identity. The key to this is that the expectation of a product is the product of the expectations, *provided* that the factors in the product are independent. Fix indices $0 \leq k < m$. The event $\{\tau = k\}$ depends only on the random variables $\xi_1, \xi_2, \ldots, \xi_k$, and so is independent of the random variable ξ_m. Similarly, the product $e^{\theta S_k} \mathbf{1}\{\tau = k\}$ is independent of $\sum_{m=k+1}^{n} \xi_m$. Consequently, by the product rule, for any $n \geq k$,

$$\begin{aligned} E\exp\{\theta S_n\}\mathbf{1}\{\tau = k\} &= E\exp\{\theta S_k\}\exp\{\theta(S_n - S_k)\}\mathbf{1}\{\tau = k\} \\ &= E\exp\{\theta(S_n - S_k)\}E\exp\{\theta S_k\}\mathbf{1}\{\tau = k\} \\ &= \varphi(\theta)^{n-k} E e^{\theta S_k} \mathbf{1}\{\tau = k\}. \end{aligned} \tag{28}$$

Here $\mathbf{1}_F$ denotes the *indicator random variable* for the event F, that is, the random variable that takes the value 1 on F and 0 on F^c.

Suppose now that τ is a *bounded* stopping time, that is, that there is a nonrandom integer $n<\infty$ such that $\tau \leq n$. Then by equation (28),

$$\begin{aligned}
E\left(\frac{\exp\{\theta S_\tau\}}{\varphi(\theta)^\tau}\right) &= \sum_{k=0}^{n} E\left(\frac{\exp\{\theta S_\tau\}}{\varphi(\theta)^\tau}\right)\mathbf{1}\{\tau=k\} \\
&= \sum_{k=0}^{n} E\left(\frac{\exp\{\theta S_k\}}{\varphi(\theta)^k}\right)\mathbf{1}\{\tau=k\} \\
&= \sum_{k=0}^{n} E\left(\frac{\exp\{\theta S_k\}}{\varphi(\theta)^k}\right)\left(\frac{\exp\{\theta S_n - S_k\}}{\varphi(\theta)^{n-k}}\right)\mathbf{1}\{\tau=k\} \\
&= \sum_{k=0}^{n} E\left(\frac{\exp\{\theta S_n\}}{\varphi(\theta)^n}\right)\mathbf{1}\{\tau=k\} \\
&= E\left(\frac{\exp\{\theta S_n\}}{\varphi(\theta)^n}\right) \\
&= 1.
\end{aligned}$$

2.4. **Gambler's Ruin, Revisited.** Consider once again the simple random walk on $\mathbb{Z}$ with initial point $S_0=x$, and let $T=T_{[0,A]}$ be the first exit time from the interval $[1,A-1]$. To use the Wald identities, we must subtract x. We also need to know *a priori* that $ET<\infty$, but this follows by essentially the same argument that we used earlier to show that $T<\infty$. (Exercise: Fill in the gap.) The first Wald identity implies that

$$E^x(S_T-x)=\mu ET=0.$$

Now the random variable S_T takes only two values, 0 and A, with probabilities $u(x)$ and $1-u(x)$ respectively. Hence,

$$\begin{aligned}
(A-x)u(x)+(-x)(1-u(x))=0 \quad &\Longrightarrow \\
u(x)=x/A.
\end{aligned}$$

Next, apply the *second* Wald identity, using $\sigma^2=E\xi_1^2=1$:

$$E(S_T-x)^2=\sigma^2 ET=ET.$$

Since we know the distribution of S_T, by the first Wald identity, we can use it to compute the left side. The result:

$$(A-x)^2\frac{x}{A}+x^2\frac{A-x}{A}=x(A-x)=ET.$$

2.5. **First-Passage Time Distribution.** Let S_n be simple random walk with initial state $S_0=0$, and let $\tau=\tau(1)$ be the first passage time to the level 1. Earlier we derived explicit formulas for the distribution and probability generating function of τ using the Reflection Principle and algebra. Here we'll see that the probability generating function can also be obtained by using the third Wald identity. For this, we need the moment generating function of ξ_1:

$$\varphi(\theta)=Ee^{\theta\xi_1}=\cosh\theta.$$

Set $s = 1/\varphi(\theta)$; then by solving a quadratic equation (exercise) you find that for $\theta > 0$,

$$e^{-\theta} = \frac{1-\sqrt{1-4s^2}}{2s}.$$

Now let's use the third Wald identity. Since this only applies directly to *bounded* stopping times, we'll use it on $\tau \wedge n$ and then hope for the best in letting $n \to \infty$. The identity gives

$$E\left(\frac{\exp\{\theta S_{\tau\wedge n}\}}{\varphi(\theta)^{\tau\wedge n}}\right) = 1.$$

We will argue below that if $\theta > 0$ then it is permissible to take $n \to \infty$ in this identity. Suppose for the moment that it is; then since $S_\tau \equiv 1$, the limiting form of the identity will read, after the substitution $s = 1/\varphi(\theta)$,

$$e^\theta E s^\tau = 1.$$

Using the formula for $e^{-\theta}$ obtained above, we conclude that

$$\boxed{E s^\tau = \frac{1-\sqrt{1-4s^2}}{2s}} \tag{29}$$

To justify letting $n \to \infty$ above, we use the dominated convergence theorem. First, since $\tau < \infty$ (at least with probability one),

$$\lim_{n\to\infty} \frac{\exp\{\theta S_{\tau\wedge n}\}}{\varphi(\theta)^{\tau\wedge n}} = \frac{\exp\{\theta S_\tau\}}{\varphi(\theta)^\tau}.$$

Hence, by the DCT, it will follow that interchange of limit and expectation is allowable provided the integrands are dominated by an integrable random variable. For this, examine the numerator and the denominator separately. Since $\theta > 0$, the random variable $e^{\theta S_{\tau\wedge n}}$ cannot be larger than e^θ, because on the one hand, $S_\tau = 1$, and on the other, if $\tau > n$ then $S_n \le 0$ and so $e^{S_{\tau\wedge n}} \le 1$. The denominator is even easier: since $\varphi(\theta) = \cosh\theta \ge 1$, it is always the case that $\varphi(\theta)^{\tau\wedge n} \ge 1$. Thus,

$$\frac{\exp\{\theta S_{\tau\wedge n}\}}{\varphi(\theta)^{\tau\wedge n}} \le e^\theta,$$

and so the integrands are uniformly bounded.

Exercise 12. A probability distribution $F = \{p_x\}_{x\in\mathbb{Z}}$ on the integers is said to have a *geometric right tail* if for some values of $\alpha > 0$ and $0 < \varrho < 1$,

$$p_x = \alpha\varrho^x \quad \text{for all } x \ge 1. \tag{30}$$

Let $S_n = \sum_{j=1}^n \xi_j$ be a random walk whose step distribution F has a geometric right tail (30). For each $x \ge 0$, define

$$\begin{aligned}\tau_x = \tau(x) &= \min\{n : S_n > x\}\\ &= \infty \quad \text{if } S_n \le x \ \forall\, n.\end{aligned}$$

(A) Show that the conditional distribution of $S_{\tau(x)} - x$, given that $\tau(x) < \infty$, is the geometric distribution with parameter ϱ.

(B) Suppose that $E\xi_j = \mu > 0$. Calculate $E\tau(x)$.

Exercise 13. Let $\{S_n\}_{n\geq 0}$ be simple random walk started at $S_0 = 0$. Fix $-A < 0 < B$ and let $T = T_{[-A,B]}$ be the first time that the random walk visits either $-A$ or $+B$. Use the third Wald identity to evaluate the generating functions

$$\psi_+(s) := Es^T \mathbf{1}\{S_T = +B\} \quad \text{and}$$
$$\psi_-(s) := Es^T \mathbf{1}\{S_T = -A\}.$$

Use your formulas to deduce as much as you can about the distribution of T. HINT: For each $0 < s < 1$ there are *two* solutions $\theta \in \mathbb{R}$ of the equation $\cosh\theta = 1/s$. Use the third Wald identity for each of these: this gives two equations in two unknowns.

3. The Strong Law Of Large Numbers and Random Walk

3.1. **The SLLN and the Ergodic Theorem.** Three of the most fundamental theorems concerning one-dimensional random walks — the Strong Law of Large Numbers, the Recurrence Theorem, and the Renewal Theorem — are all "first-moment" theorems, that is, they require only that the step distribution have finite first moment. The most basic of these theorems is the Strong Law of Large Numbers; we will see, later, that the others are consequences of the Strong Law. We will also see that the SLLN is useful in other ways, in particular for doing certain calculations (see Exercise 15 below for an example). Here is a precise statement:

Theorem 9. *(SLLN) Let $\xi_1, \xi_2, \dots$ be independent, identically distributed random variables with finite first moment $E|\xi_1| < \infty$ and mean $\mu := E\xi_1$, and let $S_n = \sum_{k=1}^n \xi_k$. Then with probability one,*

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu. \tag{31}$$

We'll take this as known, even though we haven't proved it. Here is an equivalent way to state it: Fix $\varepsilon > 0$ small, and let $L_\pm$ be the lines through the origin of slopes $\mu \pm \varepsilon$, respectively. Then with probability one, the points (n, S_n) on the graph of the random walk eventually all fall between the lines L_+ and L_-. See the figure below for a simulation of 2500 steps of the $p-q$ random walk with $p = .6$.

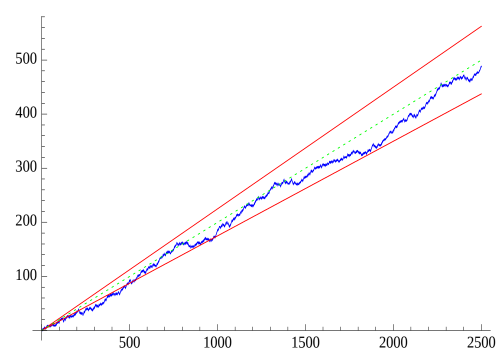

FIGURE 2. The Strong Law of Large Numbers

Corollary 10. *If the step distribution of the random walk S_n has finite,* nonzero *mean μ, then with probability one $S_n \to \infty$ if $\mu > 0$, and with probability one $S_n \to -\infty$ if $\mu < 0$. Therefore, random walk with nonzero mean is* transient*: it makes at most finitely many visits to any state $x \in \mathbb{Z}$.*

Without the hypothesis of finite first moment, the SLLN may fail to hold. An instructive example is provided by the *Cauchy distribution*: If the random variables ξ_i are i.i.d. with the standard Cauchy density

$$p(x) = \frac{1}{\pi(1+x^2)}$$

then with probability one the sequence S_n/n not only fails to converge, but has the entire real line $\mathbb{R}$ as its set of accumulation points.

Exercise 14. Prove this. HINT: First, show that for every $n \geq 1$ the sample average S_n/n has density $p(x)$. This is most easily done by using characteristic functions (Fourier transforms).

Exercise 15. Deduce the first Wald identity from the SLLN. HINT: String together infinitely many independent copies of the random sequence

$$X_1, X_2, \dots, X_T.$$

There is an important and useful generalization of the Strong Law of Large Numbers, called the *Ergodic Theorem*, due to Birkhoff. Following is a special case tailored to applications in random walk theory and the study of Markov chains. Let $g : \mathbb{R}^\infty \to \mathbb{R}$ be a bounded (measurable) function mapping infinite sequences to real numbers, and set

$$Y_n = g(X_n, X_{n+1}, X_{n+2}, \dots). \tag{32}$$

(For example, Y_n might be the indicator of the event that the random walk $\{S_{m+n} - S_n\}_{m\geq 1}$ ever visits the state 0. This is the particular case that will come into play in section 3.2 below.) The random variables $Y_1, Y_2, \dots,$ although not independent, are identically distributed; in fact, the sequence Y_n is *stationary*, that is, for every $m \geq 1$,

$$(Y_1, Y_2, \dots) \stackrel{\mathscr{D}}{=} (Y_{m+1}, Y_{m+2}, \dots). \tag{33}$$

(The notation $\stackrel{\mathscr{D}}{=}$ means that the two sides have the same joint distribution.)

Theorem 11. *(Ergodic Theorem) Let Y_n be defined by* (32). *If $E|Y_1| < \infty$ then with probability one*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = EY_1. \tag{34}$$

Remark 5. Any reasonable function g — in particular, any function that is a limit of functions depending on only finitely many coordinates — is measurable. What we'll need to know about measurable functions is this: For any $\varepsilon > 0$ there exists a bounded function h that depends on only finitely many coordinates such that

$$E|g(X_1, X_2, \dots) - h(X_1, X_2, \dots)| < \varepsilon \tag{35}$$

The proof of Theorem 11 relies on ideas and techniques that won't be needed elsewhere in the course, so it is relegated to Appendix 5 below. However, the *weak* form of Theorem 11, which states that the convergence (34) takes place in probability, can be deduced easily from the Weak Law of Large Numbers and the Chebyshev-Markov inequality, as follows.

Proof of the Weak Ergodic Theorem. This will be accomplished in two steps: First, we'll show that the theorem is true for functions g that depend on only finitely many coordinates. This, it turns out, is easy, given the SLLN. Then we'll use an approximation argument to show that it holds in general.

Step 1: Suppose that g depends on only the first m coordinates, that is,

$$g(x_1, x_2, \dots) = g(x, x_2, \dots, x_m).$$

If we break the sequence $\xi_1, \xi_2, \dots$ into blocks of m, and then apply g to each block, the resulting random variables are independent. Hence, each of the m sequences

$$\begin{aligned} &Y_1, Y_{m+1}, Y_{2m+1}, \dots \\ &Y_2, Y_{m+2}, Y_{2m+2}, \dots \\ &\dots \\ &Y_m, Y_{m+m}, Y_{2m+m}, \dots \end{aligned} \tag{36}$$

consists of independent, identically distributed random variables. Consequently, the SLLN applies to each row separately: for each $k = 1, 2, \dots, m$, with probability one,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} Y_{k+jm} = EY_1. \tag{37}$$

If n is a multiple of m, say $n = mn'$, then the sample average on the left side of (34) is just the average of sample averages of the rows (36), and so (37) implies that the convergence (34) holds when the limit is taken through the subsequence $n = mn'$ of multiples of m. It then follows routinely that the whole sequence converges. (Exercise: Fill in the details. You will need to know that $n^{-1}Y_n \to 0$ almost surely. This can be proved with the help of the Borel-Cantelli lemma, using the hypothesis that $E|Y_1| < \infty$.) □

Step 2: Now let g be an arbitrary bounded measurable function of the sequence $x_1, x_2, \dots$. By Remark 5 above, for each choice of $0 < \varepsilon = \delta^2 < 1$, there is a function h depending only on finitely many coordinates such that inequality (35) holds. Set

$$U_n = h(X_n, X_{n+1}, \dots).$$

The Chebyshev-Markov inequality inequality and inequality (35) imply that for each $n = 1, 2, \dots$,

$$P\left\{\left|n^{-1}\sum_{k=1}^{n} Y_k - n^{-1}\sum_{k=1}^{n} U_k\right| > \delta\right\} < \varepsilon/\delta = \delta.$$

The triangle inequality and inequality (35) imply that

$$|EY_m - EU_m| < \varepsilon < \delta.$$

Since h depends on only finitely many coordinates, the weak law applies to sample averages of the sequence U_j, by Step 1; hence,

$$\lim_{n\to\infty} P\left\{\left|n^{-1}\sum_{k=1}^{n} U_k - EU_1\right| > \delta\right\} = 0.$$

Combining the last three displayed inequalities yields

$$\limsup_{n\to\infty} P\left\{\left|n^{-1}\sum_{k=1}^{n} Y_k - EY_1\right| > 3\delta\right\} \le \delta.$$

Since $\delta > 0$ is arbitrary, it follows that the sample averages of the sequence Y_j converge in probability to EY_1. □

3.2. **Recurrence/Transience.** G. Polya proved, in about 1920, that simple random walk on $\mathbb{Z}^d$ is recurrent in dimensions $d = 1,2$ and is transient in $d \geq 3$. The analogous result for more general random walks was proved 30 years later by K. L. Chung & W. Fuchs.

Definition 3. A random walk on $\mathbb{Z}^d$ is said to be *recurrent* if $P\{S_n = S_0$ for some $n \geq 1\} = 1$, and otherwise is said to be *transient*. Equivalently, a random walk is recurrent if $P\{$no return to $S_0\} = 0$.

Theorem 12. *Random walk with step distribution F is recurrent if $d = 1$ and F has mean 0, or if $d = 2$ and F has mean zero* and *finite second moment. Random walk in dimension $d \geq 3$ is transient unless the step distribution F is supported by a two-dimensional subspace of $\mathbb{R}^d$.*

The hypothesis of finite second moment in dimension $d = 2$ is necessary: There are mean-zero step distributions F on $\mathbb{Z}^2$ which generate *transient* random walks.

A number of different proofs of Theorem 12 are now known. The original proof of Polya for simple random walk was based on Stirling's Formula. Chung and Fuchs proved their more general version using Fourier analysis (characteristic functions). An interesting probabilistic proof was found by D. Ornstein ten years later. Yet another proof for the one-dimensional case is based on a second important theorem about random walk discovered by Kesten, Spitzer, and Whitman in the mid-1960s. This is the proof that appears below. Before getting to this, let's look at some of the ramifications of the theorem.

Corollary 13. *Any one-dimensional random walk whose step distribution has mean zero will revisit its starting point infinitely many times.*

Proof. According to the recurrence theorem, any such random walk will revisit its starting point at least once. Assume without loss of generality that the starting point is $S_0 = 0$. Let $T \geq 1$ be the first time that this happens. Then T is a stopping time, so by the strong Markov property (Proposition 3), the post-T process $S_{T+1}, S_{T+2}, \dots$ is again a random walk with the same step distribution. Hence, the recurrence theorem applies to this random walk, making it certain that it will revisit the origin at least once. Thus, the original random walk will return to the origin at least twice. Now use induction: If the random walk is certain to return at least m times, then the strong Markov property and the recurrence theorem ensure that it will return at least $m+1$ times. □

It is not necessarily the case that a recurrent random walk on $\mathbb{Z}$ will visit *every* integer, for the trivial reason that it may not be possible to reach certain states. For instance, if the step distribution puts mass $1/2$ on each of the two values ± 2 (so that the resulting random walk is just $2\times$ a simple random walk) then the only states that can be reached from the starting state $S_0 = 0$ are the even integers.

Definition 4. If $\{p_k\}_{k\in\mathbb{Z}}$ is a non-trivial probability distribution on the integers, define its *period* to be the greatest common divisor d of the set $\{k \in \mathbb{Z} : p_k > 0\}$.

Corollary 14. *Let S_n be a mean-zero random walk on the integers whose step distribution $\{p_k\}_{k\in\mathbb{Z}}$ has period d. Assume that the starting state is $S_0 = 0$. Then with probability one, the random walk will visit every integer multiple of d infinitely often.*

Proof. Say that an integer x is *accessible* from 0 if there is a positive-probability path from 0 to x, that is, if there are integers $k_1, k_2, \dots, k_r$ such that $p(k_i) > 0$ for every i, and *positive* integers

$m_1, m_2, \dots, m_r$ such that

$$m_1 k_1 + m_2 k_2 + \cdots + m_r k_r = x. \tag{38}$$

Let $\mathscr{A}$ be the set of states accessible from 0. The key to the corollary is this: If d is the period of the distribution $\{p_k\}_{k\in\mathbb{Z}}$, then $\mathscr{A} = d\mathbb{Z}$, that is, the accessible states are precisely the integer multiples of d. It is clear that $\mathscr{A} \subset d\mathbb{Z}$, because if d is the period of the step distribution then all steps of the random walk are multiples of d. The reverse inclusion follows from a basic result of elementary number theory, according to which (in our terminology) either $\pm d \in \mathscr{A}$. This implies that $d\mathbb{Z} \subset \mathscr{A}$, by the following argument:

Suppose (for definiteness) that $-d \in \mathscr{A}$, that is, there is a positive-probability path γ to $-d$. Then for every integer $k \geq 1$ there is a positive-probability path to $-kd$, to wit, γ repeated k times. Next, because the step distribution has mean zero, there must be a positive integer in its support, and this must be a multiple of d. Thus, there is *some* $m \geq 1$ such that $md \in \mathscr{A}$. But it then follows that $d \in \mathscr{A}$: take a positive-probability path to md, then attach $(m-1)$ copies of γ. Finally, if $d \in \mathscr{A}$ then every positive integer multiple of d is also in $\mathscr{A}$.

By Corollary 13, the random walk will revisit the origin infinitely often. Let $0 < T_1 < T_2 < \cdots$ be the times of these visits. Fix $x \in \mathscr{A}$, and let F_n be the event that the random walk visits x at some time between T_{n-1} and T_n. Since each T_n is a stopping time, the events F_n are mutually independent, by the strong Markov property, and all have the same probability (say) $p = P(F_n)$. This probability p cannot be zero, because if it were then there would be no positive probability path to x. Consequently, $p > 0$, and therefore infinitely many of the events F_n must occur (the indicators $\mathbf{1}_{F_n}$ are i.i.d. Bernoulli-p). $\square$

3.3. **The Kesten-Spitzer-Whitman Theorem.**

Theorem 15. *Let S_n be a random walk on $\mathbb{Z}^d$. For each $n = 0, 1, 2, \dots$ define R_n to be the number of distinct sites visited by the random walk in its first n steps, that is,*

$$R_n := \text{cardinality}\{S_0, S_1, \dots, S_n\}. \tag{39}$$

Then

$$\frac{R_n}{n} \longrightarrow P\{\textit{no return to } S_0\} \quad a.s. \tag{40}$$

Proof. To calculate R_n, run through the first $n+1$ states S_j of the random walk and for each count $+1$ if S_j is not revisited by time n, that is,

$$R_n = \sum_{j=0}^{n} \mathbf{1}\{S_j \text{ not revisited before time } n\}.$$

The event that S_j is not revisited by time n contains the event that S_j is never revisited at all; consequently,

$$R_n \geq \sum_{j=0}^{n} \mathbf{1}\{S_j \text{ never revisited}\} = \sum_{j=0}^{n} \mathbf{1}\{S_j \neq S_{n+j} \text{ for any } n \geq 1\}.$$

The sum on the right is of the type covered by the Ergodic Theorem 11, because the event that S_j is never revisited coincides with the event that the random walk $S_{n+j} - S_j$ with increments $\xi_{j+1}, \xi_{j+2}, \dots$ never revisits 0. Therefore, with probability one,

$$\liminf R_n/n \geq P^0\{S_k \neq 0 \text{ for all } k \geq 1\}. \tag{41}$$

To bound R_n above, consider again the event that S_j is not revisited by time n. Fix $M \geq 1$. If $j \leq n-M$, then this event is contained in the event that S_j is not revisited in the next M steps. Thus,

$$R_n \leq \sum_{j=0}^{n-M} \mathbf{1}\{S_j \neq S_{j+i} \text{ for any } 1 \leq i \leq M\} + M.$$

The sum on the right is once again of the type covered by the Ergodic Theorem (in fact, the summands in this case depend on only finitely many coordinates of the random walk). Fix M, divide by n, and let $n \to \infty$: then the Ergodic Theorem 11 implies that

$$\limsup_{n\to\infty} R_n/n \leq P^0\{S_k \neq 0 \text{ for all } 1 \leq k \leq M\}.$$

This holds for every finite $M \geq 1$; since the events on the right decrease with M, their probabilities decrease to a limit. By the dominated convergence theorem (alternatively, the lower continuity property of probability measures), the limit is the probability of the intersection. The intersection is the event that 0 is never revisited at all; thus,

$$\limsup_{n\to\infty} R_n/n \leq P^0\{S_k \neq 0 \text{ for all } k \geq 1\}. \tag{42}$$

Putting (42) with (41) gives (40). □

Exercise 16. Use the Kesten-Spitzer-Whitman theorem to calculate $P\{\text{no return to } 0\}$ for $p-q$ nearest-neighbor random walk on $\mathbb{Z}$ when $p > q$.

3.4. **Proof of the Recurrence Theorem in $d = 1$.** Assume that S_n is a random walk with mean $\mu = 0$. By the strong law of large numbers, for any $\varepsilon > 0$ the sample averages S_n/n will eventually stay between $\pm\varepsilon$, and so for all sufficiently large n the set of points visited by the random walk up to time n will lie entirely in the interval $[-n\varepsilon, n\varepsilon]$. Therefore, with probability one,

$$\limsup_{n\to\infty} R_n/n \leq 2\varepsilon. \tag{43}$$

Since $\varepsilon > 0$ can be chosen arbitrarily small, it follows that $R_n/n \to 0$ almost surely. The Kesten-Spitzer-Whitman Theorem now implies that

$$P\{\text{no return to } S_0\} = 0. \tag{44}$$

□

3.5. **The Ratio Limit Theorem.**

Theorem 16. *Let S_n be a recurrent random walk on the integers with aperiodic step distribution. Assume that the step distribution has finite first moment, so that $ES_1 = 0$. Let T_n be the time of the nth return to 0. Then for all $y, z \in \mathbb{Z}$, with probability one (under $P = P^0$),*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n \mathbf{1}\{S_k = y\}}{\sum_{k=1}^n \mathbf{1}\{S_k = z\}} = 1. \tag{45}$$

In addition, for any $x \neq 0$,

$$E \sum_{k=1}^{T_1} \mathbf{1}\{S_n = x\} = 1. \tag{46}$$

Thus, the long run frequency of visits to y is the same as the long run frequency of visits to z, and the expected number of visits to x before the first return to 0 is 1.

Exercise 17. With $z=0$, prove that the ratios on the left side of (45) converge almost surely to a non-random limit $m(y)$, and that $m(x)$ is the expectation in equation (46). HINT: By the strong Markov property, the counts

$$\sum_{k=T_n}^{T_{n+1}-1} \mathbf{1}\{S_k = y\}$$

are independent and identically distributed.

Exercise 18. Prove that the limiting constants $m(y)$ satisfy the system of linear equations

$$m(y) = \sum_x m(x) p(y-x)$$

where $p(z) = P\{\xi_i = z\}$ is the step distribution of the random walk.

Exercise 19. Prove that for some positive number α,

$$m(x) = \alpha^x \quad \text{for all } x \in \mathbb{Z}.$$

HINT: Begin by showing that $m(2) = m(1)^2$, using the spatial homogeneity of the random walk.

Exercise 20. Show that if a geometric sequence $m(x) = \alpha^x$ satisfies the sytem of linear equations in Exercise 18 then $\alpha = 1$. This proves equation (46). HINT: Use Jensen's inequality.

4. LADDER VARIABLES FOR 1D RANDOM WALKS

4.1. **Queueing and Inventory Models.** In the simplest of queueing systems, the so-called $G/G/1$ queue, jobs arrive at a single processor where they wait in a queue, in order of arrival, to be served. Jobs arrive one at a time, and the times $A_1, A_2, \dots$ between successive arrivals are independent, identically distributed *positive* random variables. (Thus, the random times $A_1, A_1 + A_2, \dots$ at which jobs arrive constitute a *renewal process*; more on these later in the course.) The processor times required for the jobs are random variables $V_1, V_2, \dots$; these are also independent and identically distributed, and independent of the interarrival times A_n. Of natural interest (among other things) is the waiting time W_n for job n (that is, the amount of time it spends in the queue before the processor begins work on it). This can be described inductively as follows:

$$W_{n+1} = (W_n - A_{n+1} + V_n)_+ \tag{47}$$

where the subscript $+$ indicates positive part. (Explanation: Job n spends $W_n + V_n$ time units in the system after it arrives, but job $n+1$ doesn't arrive until A_{n+1} time units after job n.)

The same model can be used to describe certain inventory systems. Imagine a warehouse with unlimited capacity that stores a particular commodity. At each time $n = 1, 2, \dots$, a random amount V_n is added to the current inventory W_n. Simultaneously, a request is made for A_{n+1} units of the commodity; this request is immediately filled from the available inventory $W_n + V_n$ unless the request exceeds inventory, in which case only $W_n + V_n$ is sent. The new inventory W_{n+1} is then given by (47).

The queueing process W_n has an equivalent description in terms of the random walk S_n with increments $\xi_j = V_j - A_{j+1}$. Observe that the process W_n makes exactly the same jumps as S_n *except* when these would take it below 0. Thus, $W_n = S_n$ until the first time T_1^- that $S_n < 0$, at which time the queueing process is reset to 0. Thus,

$$\begin{aligned} W_n &= S_n && \text{for } n < T_1^-; \\ &= S_n - S_{T_1^-} && \text{for } n = T_1^-. \end{aligned}$$

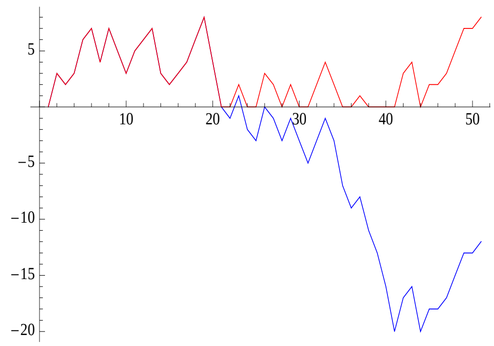

FIGURE 3. Queueing Process and Associated Random Walk

After time T_1^-, the processes W and S make the same jumps until the next time that W would fall below 0 — equivalently, the next time that S falls below $S_{T_1^-}$ — at which time W is again reset to 0. Thus, by induction,

$$
\begin{aligned}
W_n &= S_n && \text{for } n < T_1^-; \\
&= S_n - S_{T_1^-} && \text{for } T_1^- \leq n < T_2^-; \\
&= S_n - S_{T_2^-} && \text{for } T_2^- \leq n < T_3^-; \\
&\cdots && .
\end{aligned}
\tag{48}
$$

where $T_1^-, T_2^-, \dots$ are the successive times at which the random walk S_n achieves new lows. These are called the *(strong) descending ladder times* for the random walk. (The *weak* descending ladder times are the successive times at which the random walk achieves a new low *or* equals the previous low.) The relation between the queueing process and the random walk can be written in the equivalent, but more compact form

$$W_n = S_n - \min_{k \leq n} S_k \tag{49}$$

The first formula (48) has the advantage, though, that it explicitly shows the times when the queue is empty (that is, when the waiting time is 0): these are precisely the descending ladder times T_k^-. See the figure above for an illustration.

4.2. **The Duality Principle.** The *Duality Principle* for random walks on $\mathbb{Z}$ is just the simple observation that the joint distribution of the first n increments $(\xi_1, \xi_2, \dots, \xi_n)$ is the same as that of the time-reversal $(\xi_n, \xi_{n-1}, \dots, \xi_1)$. Stated this way, the Duality Principle is obvious, and the proof is a two-liner (exercise). Nevertheless, duality leads to some of the deepest and unexpected results in the theory.

The effect of reversing the increments ξ_i on the random walk can be described geometrically as follows: Plot the path $\{(k, S_k)\}_{k \leq n}$ of the random walk; then look at it while hanging upside down from the ceiling, re-setting the coordinate axes at the (old) endpoint (n, S_n). See the figure above for an example. Formally, the time-reversal replaces the path

$$
\begin{aligned}
&(0, S_1, S_2, \dots, S_n) \quad \text{by} \\
&(0, S_n - S_{n-1}, S_n - S_{n-2}, \dots, S_n - 0),
\end{aligned}
$$

and so the duality principle implies that these two paths have the same probability.

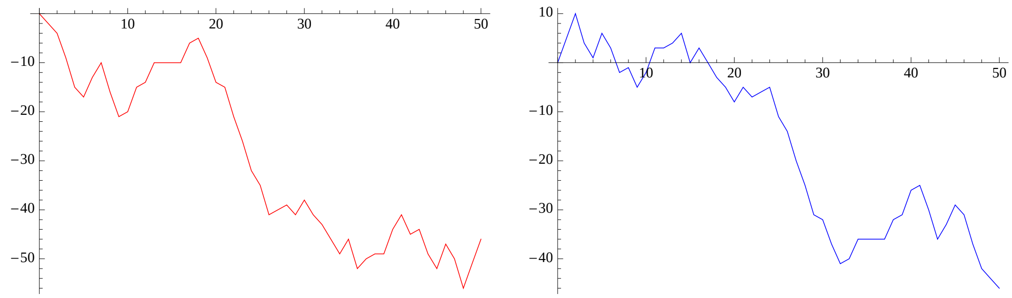

FIGURE 4. Random Walk Path and Dual Path.

Now recall the relation (49) between the queueing process W_n and the random walk S_n: the value of W_n is the amount by which S_n exceeds the minimum of the path $\{S_k\}_{k\leq n}$ up to time n. But this amount is the same as the *maximum* of the dual path! (Exercise: Prove this.) Thus:

Proposition 17. *For each $n \geq 1$, the random variables W_n and $M_n := \max_{m\leq n} S_m$ have the same distribution. Consequently, if the step distribution of the random walk S_n has* negative *mean μ then the waiting-time random variables W_n converge in distribution as $n \to \infty$ to*

$$M_\infty := \max_{n\geq 0} S_n. \tag{50}$$

Proof. You have already proved that W_n has the same distribution as M_n. Now suppose that the random walk has negative drift μ; then by the SLLN, $S_n \to -\infty$ and so M_∞ is well-defined and finite. Clearly, the random variables M_n converge monotonically to M_∞. Hence, the random variables W_n converge in distribution to M_∞. □

Remark 6. If the random walk S_n has positive or zero drift μ, then by the SLLN (in the first case) or the recurrence theorem (in the second), the maxima M_n diverge to infinity. Thus, in these cases the queueing system has no steady state: the waiting time distributions travel off to ∞ as time progresses. For an instructive if not entirely pleasant example, visit the Division of Motor Vehicles late in the afternoon.

4.3. **Duality and Ladder Variables.** The *ladder variables* are the times and heights at which record highs and lows are achieved. The *ascending* ladder variables are those associated with record highs; the *descending* ladder variables are those associated with record lows. For definiteness, we will work with *strong* ascending and *weak* descending ladder variables. The ladder indices (times) are defined as follows:

$$\begin{aligned} T^+ = T_1^+ &:= \min\{n \geq 1 : S_n > 0\}; \\ T^- = T_1^- &:= \min\{n \geq 1 : S_n \leq 0\}; \\ T_{k+1}^+ &:= \min\{n \geq 1 : S_{n+T_k^+} > S_{T_k^+}\}; \\ T_{k+1}^- &:= \min\{n \geq 1 : S_{n+T_k^-} \leq S_{T_k^-}\}. \end{aligned}$$

These may take the value $+\infty$: for instance, if the random walk has positive drift then it converges to ∞, and so there will be only finitely many record lows. If a ladder index $T_k^- = +\infty$, then all subsequent ladder indices T_{k+l}^- must also be $+\infty$. The *ladder heights* are the random variables

$$S_k^+ := S_{T_k^+} \quad \text{and} \quad S_k^- := S_{T_k^-}.$$

These are defined only on the events $T_k^{\pm} < \infty$. The first ascending and descending ladder heights will be denoted by $S_+ = S_1^+$ and $S_- = S_1^-$.

Since the ladder indices are not necessarily finite with probability 1, they are not stopping times (see Definition 2). However, they are *extended stopping times*, in the following sense:

Definition 5. An *extended stopping time* is a random variable τ taking values in the set $\mathbb{N}\cup\{\infty\}$ of extended nonnegative integers such that for any finite integer n the indicator of the event $\tau = n$ is a function only of $S_1, S_2, \dots, S_n$.

Proposition 18. *(Extended Strong Markov Property) If τ is an extended stopping time for the random walk S_n, then on the event $\tau < \infty$ the post-τ process $\{S_{n+\tau} - S_\tau\}_{n\geq 0}$ is a random walk with the same step distribution, and is independent of the path of the original random walk up to time τ. In particular, for any $m, l \in \mathbb{N}$, and any choice of states y_i and z_j,*

$$P^x\{S_k = y_k \ \forall\ k \le m;\ \ \tau = m;\ \ \textit{and}\ S_{n+m} - S_m = z_n \ \forall\ 0 \le n \le l\} = \\ P^x\{S_k = y_k \ \forall\ k \le m;\ \ \tau = m\} P^0\{S_k = z_k \ \ \forall\ k \le l\}.$$

Proof. Same as for Proposition 3. □

Corollary 19.

$$P\{T_k^+ < \infty\} = P\{T^+ < \infty\}^k \quad \textit{and} \\ P\{T_k^- < \infty\} = P\{T^- < \infty\}^k. \tag{51}$$

Proof. This is an easy consequence of the Extended Strong Markov Property. □

Recall that when the random walk has negative drift $\mu < 0$, the path attains a finite maximum $M_\infty := \max_{n\geq 0} S_n$. This maximum must be first reached at a ladder time T_k^+, for some $k \geq 0$ (with the convention $T_0^+ = 0$); and in order that no higher level is reached, it must be the case that $T_{k+1}^+ = \infty$. Therefore, the distribution of the maximum M_∞ can be obtained by summing over all possibilities k for the ladder index at which the max is attained:

$$P\{M_\infty = x\} = \sum_{k=0}^{\infty} P\{T_k^+ < \infty \text{ and } S_k^+ = x\} P\{T^+ = \infty\}. \tag{52}$$

This leads to an explicit representation of the generating function in terms of the generating function of the first ladder height S_1^+; just multiply both sides by β^x, sum over $x \geq 0$, and switch the order of summation on the right. The extended strong Markov property implies that the kth term in the outer sum on the right side is just the kth power of the first term, so the sum on the right is a geometric series with ratio $E\beta^{S_1^+}$. (NOTE: Here and in the following $E\beta^{S_-}$ means $E\beta^{S_-} = E\beta^{S_-}\mathbf{1}\{T_- < \infty\}$.) The end result:

Proposition 20.

$$E\beta^{M_\infty} = P\{T^+ = \infty\}/(1 - E\beta^{S_1^+}) \tag{53}$$

The Duality Principle has important implications for the distributions of the ladder variables, because reversing the order of the increments in a random walk has the effect of switching maxima and minima (look again at the figure above for illustration). The following four *duality relations* distill the role of time reversal in the study of the ladder variables.

Proposition 21. *(Duality Relations) For all integers $x, n \geq 1$ and $y \geq 0$,*

$$P\{S_n = -y \text{ and } T^+ > n\} = P\{S_n = -y \text{ and } \min_{1\leq m\leq n} S_m = -y\}; \tag{54}$$

$$P\{S_n = +x \text{ and } T^- > n\} = P\{S_n = +x \text{ and } \max_{1\leq m<n} S_m < +x\}; \tag{55}$$

$$P\{S_n = +x \text{ and } T^+ = n\} = P\{S_n = +x \text{ and } \min_{1\leq m\leq n} S_m \geq x\}; \quad \text{and} \tag{56}$$

$$P\{S_n = -y \text{ and } T^- = n\} = P\{S_n = -y \text{ and } \max_{1\leq m\leq n} S_m < -y\}. \tag{57}$$

Proof. Exercise. HINT: It may help to look again at the figure above depicting a random walk path and its dual, or to sketch your own. □

Corollary 22. *For all positive integers n,*

$$P\{T^+ > n\} = \sum_{k=1}^{\infty} P\{T_k^- = n\} \quad \text{and} \quad P\{T^- > n\} = \sum_{k=1}^{\infty} P\{T_k^+ = n\}. \tag{58}$$

Consequently,

$$ET^+ = 1/P\{T^- = \infty\} \quad \text{and} \tag{59}$$

$$ET^- = 1/P\{T^+ = \infty\}. \tag{60}$$

Proof. Summing the relation (54) over all $y \geq 0$ shows that $P\{T^+ > n\}$ coincides with the probability that the random walk attains its minimum value at time n. But the latter happens if and only if n is a descending ladder index. This proves the first equality; the second is similar. To evaluate ET^+, sum the probabilities in (58) and add 1 (see Lemma 8). This shows that ET^+ is the sum of a geometric series with ratio $P\{T^- < \infty\}$:

$$ET^+ = 1 + \sum_{k=1}^{\infty} P\{T_k^- < \infty\} = \sum_{k=0}^{\infty} P\{T^- < \infty\}^k.$$

□

4.4. **Step Distributions with Finite Support.** There is an analytic procedure, the so-called *Wiener-Hopf factorization technique*, that directly relates the joint distribution of the first ladder index and ladder height (T^+, S_1^+) to the characteristic function of the step distribution. In the special case where the step distribution of the random walk has *finite support*, Wiener-Hopf factorization is transparent and elementary, because the probability generating function of a step distribution with finite support is (essentially) a polynomial. We'll consider only this case — see FELLER vol. 2 or SPITZER for the general case.. Here is what we need to know about polynomials:

Proposition 23. *Every polynomial $p(x)$ of degree n has exactly n complex roots $\zeta_1, \zeta_2, \dots, \zeta_n$ (listed according to multiplicity), and*

$$p(x) = C \prod_{i=1}^{n} (x - \zeta_i) \tag{61}$$

where C is the (nonzero) coefficient of x^n in $p(x)$.

For definiteness, assume throughout this section that the distribution $F = \{p_x\}_{-L\le x\le M}$ is concentrated on the integers $-L, -L+1, \dots, +M$ where $L, M \ge 1$, that the distribution is aperiodic, and that $p_{-L} > 0$ and $p_M > 0$. Let

$$Q(\beta) := \sum_{x=-L}^{M} p_x \beta^x \tag{62}$$

be the probability generating function of F, and let $\mu = Q'(1)$ be its mean. Observe that $\beta^L Q(\beta)$ is a polynomial of exact degree $M+L$, with nonvanishing constant term p_{-L}. The generating function Q itself is a *rational function*, that is, it is the quotient of two polynomials $\beta^L Q(\beta)$ and β^L.

Lemma 24. *If β is a (possibly complex) root of the equation $Q(\beta) = 1$, then*

$$E\beta^{S_{T^+}} = 1 \quad \text{if } |\beta| > 1 \quad \text{and} \tag{63}$$

$$E\beta^{S_{T^-}} = 1 \quad \text{if } |\beta| < 1. \tag{64}$$

More generally, if β is a root of the equation $Q(\beta) = 1/t$ for some $0 < t \le 1$, then

$$Et^{T^+}\beta^{S_{T^+}} = 1 \quad \text{if } |\beta| > 1 \quad \text{and} \tag{65}$$

$$Et^{T^-}\beta^{S_{T^-}} = 1 \quad \text{if } |\beta| < 1. \tag{66}$$

Proof. The third Wald identity holds not only for real but also complex arguments of the (moment) generating function. (Re-read the proof — nowhere did we do anything that required real values of the arguments.) Thus, for each $n = 1, 2, \dots,$

$$E\beta^{S_{T^\pm\wedge n}}/Q(\beta)^{T^\pm\wedge n} = 1.$$

To deduce (63) and (64), use the dominated convergence theorem. For definiteness, consider the ascending ladder variables, and use the abbreviation $T = T^+$. First, the factor $t^{T\wedge n} = 1/Q(\beta)^{T\wedge n}$ is bounded above by 1. Second, because $|\beta| > 1$, and because $S_{n\wedge T}$ can't be larger than $+M$, the integrands β^{S_T} are bounded in absolute value by $|\beta|^M$. Now if the mean of the step distribution is ≥ 0, then $T < \infty$ (why?), and so $S_{T\wedge n} \to S_T$; but if the mean of the step distribution is < 0, then by SLLN the random walk drifts to $-\infty$, and so on the event $T = \infty$,

$$\beta^{S_{T\wedge n}} \longrightarrow 0.$$

Therefore, the dominated convergence theorem allows passage to the limit in the expectation, yielding $E\beta^{S_T} = 1$. $\square$

There is one root (possibly a double root) of the equation $Q(\beta) = 1$ that isn't covered by Lemma 24, to wit, $\beta = 1$. That there are no other roots on the unit circle follows from the aperiodicity of the step distribution. In fact:

Lemma 25. *If the step distribution F is aperiodic, then for all $\beta > 0$ and all $\theta \in \mathbb{R} - \mathbb{Z}$,*

$$|Q(\beta e^{2\pi i\theta})| \le Q(\beta) \quad \text{and} \quad Q(\beta e^{2\pi i\theta}) \ne Q(\beta) \tag{67}$$

Proof. Inequality clearly holds, by the triangle inequality, since the coefficients of all terms of Q are nonnegative. Now since all terms of the sum $Q(\beta)$ are nonnegative, the only way that equality $Q(\beta e^{2\pi i\theta}) = Q(\beta)$ can hold is if $e^{2\pi ik\theta} = 1$ for every k such that $p_k > 0$. Since $p_M > 0$, the only possibilities are $\beta_k := \beta\exp\{2\pi ik/M\}$, where k is an integer between 0 and $M-1$. But for β_k to satisfy $Q(\beta_k) = Q(\beta)$ it would have to be the case that $xk/M \in \mathbb{Z}$ for every x in the support of the step distribution. Since this distribution is aperiodic, the only possibility is $k = 0$. $\square$

Lemma 26. *The generating function $Q(\beta)$ is strictly convex for $\beta \in (0,\infty)$. It attains its minimum value uniquely at the point β_* where $Q(\beta_*) = 0$. For every value of $t < 1/Q(\beta_*)$, the equation $Q(\beta) = 1/t$ has two simple roots in the interval $(0,\infty)$, one on each side of β_*. The equation $Q(\beta) = Q(\beta_*)$ has a double root at $\beta = \beta_*$ and no other root $\beta \in (0,\infty)$.*

Proof. Exercise. This is an excellent opportunity to review what you know about probability generating functions and moment generating functions. Note while you're at it that the family of probability distributions

$$q_x^\beta := p_x \beta^x / Q(\beta) \quad \text{for} \quad \beta \in (0,\infty)$$

is an exponential family, with natural parameter $\log\beta$. □

Now the payoff.

Proposition 27. *(A) If $\mu < 0$, then the equation $Q(\beta) = 1$ has M roots $\beta_0, \beta_1, \dots \beta_{M-1}$ outside the unit circle, one root $\alpha_0 = 1$ on the unit circle, and $L-1$ roots $\alpha_1, \alpha_2, \dots, \alpha_{L-1}$ inside the unit circle. (B) If $\mu = 0$ then the equation $Q(\beta) = 1$ has $L-1$ roots $\alpha_1, \alpha_2, \dots, \alpha_{L-1}$ inside the unit circle, $M-1$ roots $\beta_1, \beta_2, \dots, \beta_{M-1}$ outside the unit circle, and a double root $\alpha_0 = \beta_0 = 1$. (C) In both cases, the ladder heights $S_+ := S_{T^+}$ and $S_- := S_{T^-}$ have probability generating functions*

$$E\beta^{S_+} = 1 + C_+ \prod_{i=1}^{M} (\beta - \beta_i) \quad \textit{and} \tag{68}$$

$$E\beta^{S_-} = 1 + C_- \prod_{i=0}^{L-1} (\beta - \alpha_i). \tag{69}$$

The normalizing constants are

$$C_- = p_{-L} \quad \textit{and} \quad C_+ = (-1)^{M-1} \Big/ \prod_{i=1}^{M-1} \beta_i \tag{70}$$

Proof. I'll prove this only for the case $\mu < 0$, as the case $\mu = 0$ is quite similar, and only in the case where the roots of $Q(\beta) = 1$ are all *simple*. (See Remark 8 below for a discussion of the case where there are multiple roots.) The key is that the generating function $\psi_+(\beta) := E\beta^{S_+}$ is a polynomial of degree M, because S_+ cannot be larger than M. Similarly, $\psi_-(\beta) := E\beta^{S_-}$ is a polynomial of degree L in β^{-1}. Consequently, the equation $\psi_+(\beta) = 1$ has exactly M roots, and the equation $\psi_-(\beta) = 1$ has L roots. Now Lemma 24 implies that every root of $Q(\beta) = 1$ outside the unit circle is a root of $\psi_+(\beta) = 1$. Hence, the equation $Q(\beta) = 1$ can have *no more than* M roots outside the unit circle. Similarly, every root of $Q(\beta) = 1$ inside the unit circle is a root of $\psi_-(\beta) = 1$, and because $T_- < \infty$ with probability one, $\alpha_0 = 1$ is also a root. Hence, the equation $Q(\beta) = 1$ can have *no more than* $L-1$ roots inside the circle. But $Q(\beta) = 1$ has $M+L$ roots in total, and (by Lemma 25) only one, $\alpha_0 = 1$, on the unit circle. Therefore, there must be *precisely* M roots outside and $L-1$ roots inside. These exhaust the roots of $\psi_+(\beta) = 1$ and $\psi_-(\beta) = 1$, so equations (68)–(69) must hold for some choice of the constants $C_\pm$. The constant C_+ is easily computed using the fact that the polynomial $E\beta^{S_+}$ has constant term 0. To see that $C_- = p_{-L}$, observe that the only way that $S_- = -L$ can occur is if the very first step is to $-L$; consequently, the coefficient of β^{-L} in the generating function must be p_{-L}. □

Remark 7. Another way to get the normalizing constant C_+ in the case $\mu < 0$ is to observe that at $\beta = 1$ the generating function $E\beta^{S_+}$ takes the value $P\{T_+ < \infty\}$. This implies that

$$C_+ = P\{T_+ = \infty\} \Big/ \prod_{i=1}^{M} (1 - \beta_i). \tag{71}$$

Combining this with the formula for C_+ in the statement of the proposition leads to the following interesting formula for $P\{T_+ = \infty\}$:

$$P\{T_+ = \infty\} = (-1) \prod_{i=1}^{M} (1 - 1/\beta_i) \tag{72}$$

Remark 8. The proposition remains true even when there are multiple roots. One way to prove this is to let $\beta_i(t)$ and $\alpha_i(t)$ be the roots of $Q(\beta) = 1/t$ for $0 < t \leq 1$. Using some elementary complex analysis (e.g., the argument principle) one can show that the roots $\alpha_i(t)$ and $\beta_i(t)$ are continuous functions of t. Furthermore, the equation $Q(\beta) = 1/t$ can have multiple roots for at most finitely many values of t, because multiple roots can only occur at points β where $Q'(\beta) = 0$, and there are at most $M + L - 1$ such points. Therefore, at all but finitely many $t < 1$ the same argument as in the proof of the proposition shows that

$$Et^{T^+} \beta^{S_+} = 1 + C_+(t) \prod_{i=1}^{M} (\beta - \beta_i(t)) \quad \text{and} \tag{73}$$

$$Et^{T^-} \beta^{S_-} = 1 + C_-(t) \prod_{i=0}^{L-1} (\beta - \alpha_i(t)). \tag{74}$$

where

$$C_-(t) = p_{-L} t \quad \text{and} \quad C_+(t) = (-1)^{M-1} \Big/ \prod_{i=1}^{M-1} \beta_i(t). \tag{75}$$

But both sides of equations (73) and (74) are continuous in t, so the formulas must remain true even at those t where the equation $Q(\beta) = 1/t$ has a multiple root. The equations (73) and (74) are interesting in their own right, as they yield simple formulas for the probability generating functions of T^+ and T^-, by setting $\beta = 1$.

Exercise 21. Use the formula (73) to give yet another derivation of equation (15). (This will be either the fifth or sixth derivation, depending on how you count.)

Corollary 28. *Assume that $\mu < 0$, and that β_i and α_j are the roots of $Q(\beta) = 1$, as in Proposition 27. Then for any integers $1 \leq m \leq M$ and $0 \leq l \leq L$,*

$$P\{S_+ = +m\} = C_+ e_{M-m}(\beta_1, \beta_2, \ldots, \beta_M) \quad \textit{and} \tag{76}$$

$$P\{S_- = -l\} = C_- e_{L-l}(\alpha_0, \alpha_1, \ldots, \alpha_{L-1}) \tag{77}$$

where $e_k(x_1, x_2, \ldots, x_n)$ is the kth elementary symmetric polynomial:

$$e_k(x_1, x_2, \ldots, x_n) = \sum_{\substack{A \subset [n] \\ |A| = k}} \prod_{i \in A} x_i.$$

Proof. The distributions of S_+ and S_- are gotten by reading off the coefficients in their probability generating functions. □

Corollary 29. *Assume that $\mu < 0$, and that β_i and α_j are the roots of $Q(\beta) = 1$. Then the probability generating function of the maximum $M_\infty = \max_{n\geq 0} S_n$ is*

$$E\beta^{M_\infty} = -\frac{P\{T^+ = \infty\}}{C_+\prod_{i=1}^M (\beta - \beta_i)}. \tag{78}$$

Consequently, for each $x = 0, 1, 2, \dots$,

$$P\{M_\infty = x\} = \sum_{i=1}^M C_i/\beta_i^x \quad \textit{where} \quad C_i = (-1)^{M-1}\frac{P\{T_+ = \infty\}}{C_+\beta_i \prod_{j\neq i}(\beta_j - \beta_i)}. \tag{79}$$

Proof. The first assertion is a direct consequence of Proposition 20 and equation (68). The second follows from the first by *partial fraction decomposition.* (NOTE: I was never any good at partial fraction decomposition, so my formula for the constants C_i is possibly wrong.) The idea is this: If the generating function in (78) is rewritten in the form

$$E\beta^{M_\infty} = \sum_{i=1}^M \frac{C_i}{(1 - \frac{\beta}{\beta_i})}$$

(as the partial fraction method guarantees that it can) then the coefficients can be recovered by expanding each term on the right as a geometric series. □

The formula (79) can be quite useful numerically, especially for large x, because by Lemma 25, when $\mu < 0$ one root (designate it β_1) is positive, and has smaller absolute value than any of the other roots $\beta_2, \beta_3, \dots, \beta_M$. Thus, when x is large, the contribution of the $i = 1$ term in the sum is large relative to those of the remaining terms $i \geq 2$, and so

$$P\{M_\infty = x\} \sim C_1/\beta_1^x \quad \text{as } x \to \infty. \tag{80}$$

5. APPENDIX: STRONG LAWS AND MAXIMAL INEQUALITIES

The proof of the weak ergodic theorem in section 3 above used the Chebyshev-Markov inequality to reduce the problem to proving the theorem for functions depending on only finitely many coordinates. This is an instance of a general strategy that works in many convergence problems: (A) First, prove the convergence theorem for a restricted, simpler class of functions or random variables. (B) Then use a suitable inequality to deduce the convergence for a larger class of functions by approximation. For most strong convergence theorems, the appropriate inequality for step (B) is a *maximal inequality.* Following is a maximal inequality (due, in essence, to N. Wiener) for stationary sequences. As in section 3, let

$$Y_n = g(X_n, X_{n+1}, \dots)$$

where g is a measurable function and $X_1, X_2, \dots$ are independent, identically distributed random variables. Denote by $S^Y_{n;m}$ and $A^Y_{n;m}$ the partial sums and sample averages:

$$S^Y_{n;m} = \sum_{k=1}^n Y_{k+m} \quad \text{and} \quad A^Y_{n;m} = S^Y_{n;m}/n,$$

and use the abbreviations $A^Y_n = A^Y_{n;0}$ and $S^Y_n = S^Y_{n;0}$.

Maximal Inequality. *Assume that g is nonnegative, and that $\mu = EY_1 < \infty$. Then for any $\alpha > \mu$,*

$$P\left\{\sup_{n\geq 1} A_n^Y > \alpha\right\} \leq \frac{\mu}{\alpha}. \tag{81}$$

Proof of the Maximal Inequality. This is optional reading; the ideas involved won't be needed again in this course. However, the argument is elementary and also quite interesting. Fix $L \geq 1$, and define events

$$B_m = B_{m,L} = \left\{\max_{1\leq n\leq L} A_{n;m}^Y > \alpha\right\}.$$

For each $\alpha > 0$ and each L the events $B_0, B_1, \dots$ all have the same probability $P(B_m) = P(B_{m,L})$, because the sequence of random variables $Y_1, Y_2, \dots$ is *stationary*. Furthermore, by the monotone convergence theorem,

$$\lim_{L\to\infty} P(B_{m,L}) = P\left\{\sup_{n\geq 1} A_{n;0}^Y > \alpha\right\}.$$

Hence, to prove the theorem it suffices to show that $P(B_{m,L}) \leq \mu/\alpha$ for every $L = 1, 2, \dots$.

The trick is to partition the set of positive integers m into two disjoint subsets, which I will call *purple* and *white*, depending on the realization of the stationary process Y_j. The partition is done as follows. First, color an integer $m \geq 1$ *red* if the event B_m occurs, and *white* otherwise. Now grab a can of purple paint, and march forward through the positive integers starting at $m = 1$. If $m = 1$ is white, leave it white and move to $m = 2$. On the other hand, if $m = 1$ is red, you can find $n \leq L$ (for definiteness, take the smallest such n) so that the average of the first n terms Y_j exceeds α. Paint all of the integers $1, 2, \dots, n$ *purple*, and move to $m = n + 1$. If m is white, leave it white and move to $m + 1$. Otherwise, select an integer $n \geq 1$ so that the average of the n terms Y_j beginning at $j = m$ exceeds α; paint the integers from m to $m + n - 1$ purple and move on to $m + n$. Continue in this fashion indefinitely. At the end of your march, all red integers — and some white ones — will have been painted purple. No purple interval will have length $\geq L$, but purple intervals may abut. By construction, in every purple interval, the average of the terms Y_j will exceed α. Equivalently, the *sum* of the terms Y_j in every purple interval will exceed α times the length of the interval.

The last is the key point, because by hypothesis the expectation of each term Y_j is only μ. Denote by $\mathscr{P}$ the purple integers and by $\mathscr{R}$ the red integers. Then for any K, since $\mathscr{R} \subset \mathscr{P}$,

$$\mu = K^{-1} E \sum_{k=1}^{K} Y_j \geq K^{-1} E \sum_{k\leq K;\, k\in\mathscr{P}} Y_j \geq K^{-1} \alpha E|\mathscr{R} \cap [1, K - L]|.$$

Now $\mathscr{R}$ is the set of integers $m \geq 1$ for which the event $B_{m,L}$ occurs; consequently, its expected cardinality is just the sum of the probabilities $P(B_{m,L})$. Thus,

$$K^{-1} E|\mathscr{R} \cap [1, K - L]| = K^{-1} \sum_{m=1}^{K-L} P(B_m);$$

since all of the events B_m have the same probability, it follows by letting $K \to \infty$ that

$$P(B_1) \leq \mu/\alpha.$$

$\square$

I'll give two applications of the Maximal Inequality. First, I'll prove the SLLN, and then I'll show how to deduce the Ergodic Theorem (Theorem 11) from the SLLN.

Proof of the SLLN. First, consider the case where the random variables X_i have finite second moment. Without loss of generality, assume that $EX_i = 0$, and let $\sigma^2 = EY_i^2$. For convenience, drop the superscript X from A_n^X and S_n^X. Then by the Cauchy-Schwartz inequality,

$$E|A_n| \leq \sqrt{EA_n^2} = \sigma/\sqrt{n} \longrightarrow 0 \quad \text{as } n \to \infty.$$

Fix $\delta > 0$ small, and choose $K \geq 1$ large enough that $E|A_K| < \delta$. For any $n \geq K$, the sample average A_n is nearly (but not quite) the average of the sample averages $A_{K;m}$ from $m = 1, 2, \ldots, n$; the error comes about at the boundaries. Precisely:

$$A_n - \frac{1}{n}\sum_{m=1}^{n} A_{K;m} = \frac{1}{nK}\sum_{j=1}^{K-1}(K-j)X_j - \frac{1}{nK}\sum_{j=n+1}^{n+K-1}(j-n)X_j := V_{n;K}$$

Lemma 30. *If $E|X_1| < \infty$ then $\lim_{n\to\infty} V_{n;K} = 0$ almost surely.*

Proof. Exercise. HINT: Use the easy half of the Borel-Cantelli Lemma. □

It follows that the lim sup and lim inf of the sample averages A_n are the same as the lim sup and lim inf of the averages of the averages:

$$\limsup_{n\to\infty} A_n = \limsup_{n\to\infty} n^{-1}\sum_{m=1}^{n} A_{K;m} \quad \text{and} \quad \liminf_{n\to\infty} A_n = \liminf_{n\to\infty} n^{-1}\sum_{m=1}^{n} A_{K;m}.$$

But the Maximal Inequality guarantees that these limsups and liminfs must be close to zero with high probability: in particular, if $\delta = \varepsilon^2$ and $E|A_K| < \delta$, then

$$P\{\sup_n n^{-1}\sum_{m=1}^{n} |A_{K;m}| > \varepsilon\} < \delta/\varepsilon = \varepsilon.$$

Therefore,

$$P\{\limsup_{n\to\infty} A_n > \varepsilon\} < \varepsilon \quad \text{and} \quad P\{\liminf_{n\to\infty} A_n < -\varepsilon\} < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that the limsup and liminf must equal 0 with probability one. This proves the SLLN under the assumption that the random variables X_i have finite second moment.

Now consider the general case, where $E|X_i| < \infty$. Assume that $EX_i = 0$. Then for any $\delta > 0$ there exist i.i.d. *bounded* random variables Y_i such that $E|X_i - Y_i| < \delta$. (This follows from the dominated convergence theorem: Truncate X_i at $\pm m$ and let $m \to \infty$.) Without loss of generality, the random variables Y_i can be chosen so that $EY_i = 0$. (Why?) Since the random variables Y_i are bounded, they have finite variance, and so

$$\lim_{n\to\infty} A_n^Y = EY_1 = 0 \quad \text{almost surely.}$$

But the Maximal Inequality guarantees that the differences $A_n^X - A_n^Y$ must all remain near zero, except with small probability. In particular, if $\delta = \varepsilon^2$, then

$$P\{\sup_{n\geq 1} |A_n^X - A_n^Y| > \varepsilon\} < \varepsilon.$$

It follows that the lim sup and lim inf of the sequence A_n^X must lie in the interval $[-\varepsilon, \varepsilon]$ with probability at least $1 - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, this implies that the liminf and limsup must be 0 almost surely. □

Proof of Theorem 11. Let $Y_n = g(X_n, X_{n+1}, \dots)$ for a measurable function g such that $E|Y_n| < \infty$. As in the proof of the *weak* ergodic theorem in section 3, there exists a function h depending on only finitely many coordinates so that inequality (35) holds, with $\varepsilon = \delta^2 < 1$. Thus, if

$$U_n = h(X_n, X_{n+1}, \dots),$$

then $E|Y_1 - U_1| < \varepsilon$. Without loss of generality, the function h can be chosen so that $EU_n = EY_n$ (for the same reason as in the proof of the SLLN). The Maximal Inequality (applied to sample averages of the *differences* $|Y_j - U_j|$) implies that

$$P\{\sup_{n\geq 1} n^{-1} \sum_{k=1}^{n} |Y_j - U_j| > \delta\} < \varepsilon/\delta = \delta.$$

But we have already seen (see Step 1 of the proof of the weak ergodic theorem) that the sample averages of the random variables U_j converge to the expectation $EU_1 = EY_1$ almost surely, and so the liminf and limsup of these sample averages are both EY_1. Thus,

$$P\{|\limsup_{n\to\infty} A_n^Y - EY_1| \geq \delta\} < \delta \quad \text{and} \quad P\{|\liminf_{n\to\infty} A_n^Y - EY_1| \geq \delta\} < \delta.$$

Since $\delta > 0$ is arbitrary, it follows that the limsup and liminf both equal EY_1 with probability one. $\square$