Random Walks on Infinite Discrete Groups

Steven P. Lalley*

July 19, 2018

Abstract

Lecture notes for a 1-week introductory course for advanced undergraduates and beginning graduate students. The course took place at the Northwestern University Summer School in Probability, July 2018. My thanks to Antonio Auffinger and Elton Hsu for inviting me to give this course.

Contents

*Department of Statistics, University of Chicago

1 Introduction

1.1 Groups and their Cayley graphs

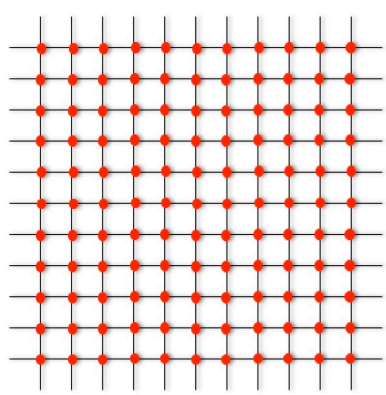
The random walks to be studied in these lectures will all live in infinite, finitely generated groups. A group Γ is said to be *finitely generated* if there is a finite set A, the set of *generators*, such that every element of Γ can be expressed as a finite product of elements of A. Because the set of finite sequences from a finite set A is countable, every finitely generated group is either finite or countable. We will assume throughout that the set A of generators is *symmetric*, that is, for every $a \in A$ it is also the case that $a^{-1} \in A$.

For any finitely generated group Γ with (symmetric) generating set A there is a homogeneous *graph* $G_\Gamma = G_{\Gamma;A}$, called the *Cayley graph*, that reflects, in a natural way, the geometry of the group. This is defined as follows:

(a) the vertex set of G_Γ is the group Γ; and
(b) the edge set of G_Γ is the set of unordered pairs $\{x, y\}$ of group elements such that $y = xa$ for some $a \in A$.

The edge structure of this graph provides a convenient way to define a metric d, called the *word metric* or just the *Cayley graph metric*, on the group Γ: for any two group elements x, y, the distance $d(x, y)$ is defined to be the length of the shortest path in G_Γ from x to y, equivalently, $d(x, y)$ is the length m of the shortest word $a_1 a_2 \cdots a_m$ in the generators that represents the group element $x^{-1}y$. Clearly, the word metric d depends on the choice of the generating set A.

Example 1.1. The additive group of $d-$dimensional integer points $\mathbb{Z}^d$ has generating set $\{\pm e_i\}_{1 \leq i \leq d}$, where e_i is the ith standard unit vector in $\mathbb{R}^d$. The Cayley graph with respect to this set of generators is the usual cubic lattice in d dimensions.



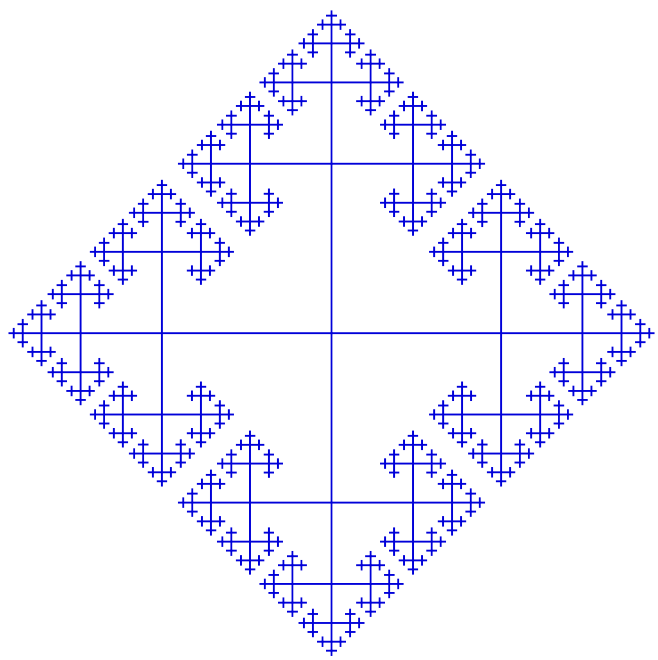
Example 1.2. The *free group* $\mathbb{F}_d$ on $d \geq 2$ generators $a_1, a_2, \cdots, a_d$ is the set of all finite words (including the empty word $\emptyset$, which represents the group identity)

$$x = b_1 b_2 \cdots b_n$$

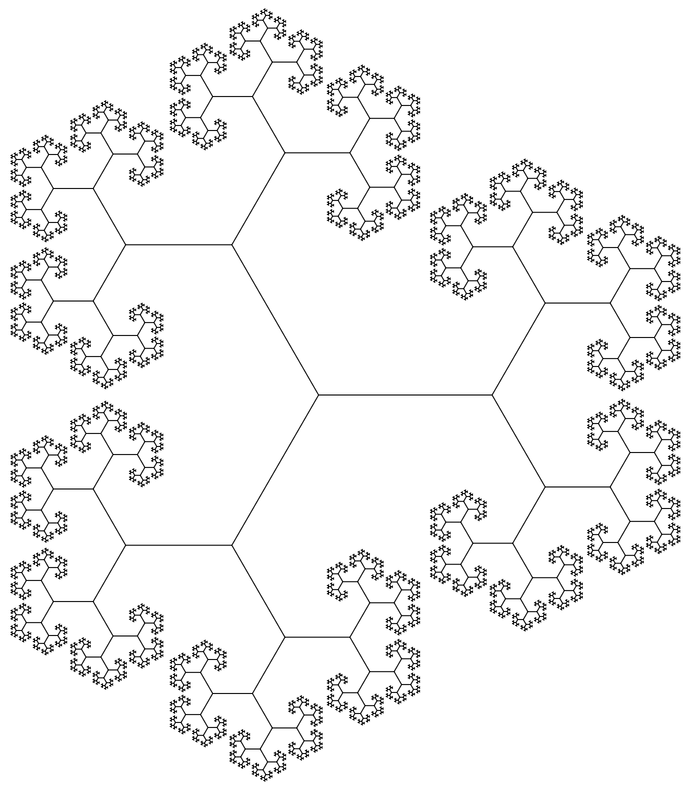
in which each b_i is one of the symmetric generators $a_i^{\pm 1}$, and in which no entry b_i is adjacent to its inverse b_i^{-1}. Such words are called *reduced*. Two such words x, y are multiplied by concatenating theeir representative words and then doing whatever cancellations of adjacent letters are possible at the juxtaposition point. For instance, if $x = de^{-1}fd$ and $y = d^{-1}f^{-1}dde$ then

$$xy = (de^{-1}fd)(d^{-1}f^{-1}dde) = de^{-1}dde.$$

The Cayley graph of the free group $\mathbb{F}_d$ relative to the standard set of generators $A = \{a_i^{\pm 1}\}_{i \leq d}$ is the infinite, homogeneous tree $\mathbb{T}_d$ of degree d: for $d = 2$,



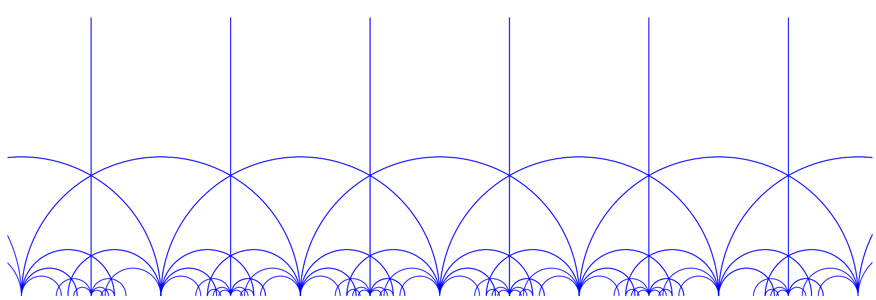
Example 1.3. The *free product* $\mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$ of three copies of the two element group consists of all finite words (once again including the empty word $\emptyset$) from the three-element alphabet $\{a, b, c\}$ in which no letter $a, b,$ or c is adjacent to itself. Group multiplication is concatenation followed by successive elimination of as many "spurs" $aa, bb,$ or cc as possible at the juxtaposition point. (Thus, each one-letter word a, b, c is its own inverse.) The Cayley graph is once again an infinite tree, this one of degree 3:



Example 1.4. The group $SL(2, \mathbb{Z})$ is the group of 2×2 matrices with integer entries and determinant 1, with matrix multiplication. This group has the two-element normal subgroup $\{\pm I\}$; the quotient group $SL(2, \mathbb{Z})/\{\pm I\}$ is known as $PSL(2, \mathbb{Z})$. Its Cayley graph relative to the generating set

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

looks like this (actually, it looks like the *dual graph* to this):



Here each *geodesic triangle* represents a vertex (group element); two vertices share an edge in the Cayley graph if the corresponding triangles meet in a side.

1.2 Random Walks: Definitions and Conventions

Henceforth, we will use multiplicative notation for the group operation (except when the group is the integer lattice $\mathbb{Z}^d$), and we will denote the group identity by the symbol 1 (but for free groups and free products, we will sometimes use $\emptyset$ for the group identity to emphasize its word representation). Assume henceforth that Γ is a finitely generated group with symmetric generating set A and corresponding Cayley graph G_Γ.

Definition 1.5. A *random walk* on Γ (or equivalently on G_Γ) is an infinite sequence $X_0, X_1, \cdots$ of $\Gamma-$valued random variables, all defined on the same probability space $(\Omega, \mathcal{F}, P^x,)$ of the form

$$X_n = X_0 \xi_1 \xi_2 \cdots \xi_n \tag{1.1}$$

where $\xi_1, \xi_2, \cdots$ are independent, identically distributed random variables that take values in Γ. When the increments ξ_i take values in the generating set A, the random walk is a *nearest-neighbor random walk*. Unless otherwise specified, the initial state is $X_0 = 1$. For any other (non-random) initial state $X_0 = x \in \Gamma$, we shall use a superscript x on the probability and expectation operators P^x and E^x; thus, $P^x\{X_n = y\} = P\{X_n = x^{-1}y\}$. The *step distribution* of the random walk X_n is the common distribution μ of the increments ξ_i, that is,

$$\mu(y) := P\{\xi_i = y\} \implies \tag{1.2}$$

$$\mu^{*n}(y) = P\{X_n = y\}. \tag{1.3}$$

Here μ^{*n} denotes the $n-$fold convolution of the probability measure μ with itself. The $n-$*step transition probabilities* are defined by

$$p_n(x, y) = P^x\{X_n = y\} = \mu^{*n}(x^{-1}y). \tag{1.4}$$

Technical Note: A $\Gamma-$valued random variable X_n is, by definition, a measurable mapping $X_n : \Omega \to \Gamma$, where Ω is a set equipped with a $\sigma-$algebra $\mathcal{F}$. What it means for a mapping $X_n : \Omega \to \Gamma$ to be *measurable* [1] is that for every element $y \in \Gamma$ the set $\{X_n = y\} := X_n^{-1}\{y\}$ is an element of the $\sigma-$algebra $\mathcal{F}$. For each $x \in \Gamma$ there is a probability measure P^x on the $\sigma-$algebra $\mathcal{F}$, which is uniquely specified by the requirement that for any finite sequence $a_1, a_2, \cdots, a_n$ in the generating set A,

$$P^x\{X_0 = x \text{ and } X_{j+1} = X_j a_{j+1}\ \forall\ 0 \le j < n\} = \prod_{j=1}^{n} \mu(a_j). \tag{1.5}$$

When $x = 1$, we shall omit the superscript on P^x; thus, $P = P^1$. See the Appendix for a brief explanation of how to build a measure space that supports an infinite sequence

[1] More generally, a mappling $T : \Omega \to \Upsilon$ from one measurable space $(\Omega, \mathcal{F})$ to another $(\Upsilon, \mathcal{G})$ is *measurable* if for every set $G \in \mathcal{G}$ the inverse image $T^{-1}(G) \in \mathcal{F}$. If the $\sigma-$algebra $\mathcal{G}$ is *generated* by a subset $\mathcal{A} \subset \mathcal{G}$ – that is, if $\mathcal{G}$ is the minimal $\sigma-$algebra containing $\mathcal{A}$ – then to check measurability it is enough to check that $T^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{A}$.

$\xi_1, \xi_2, \cdots$ of independent, identically distributed random variables with given distribution μ.

A random walk X_n is said to be *symmetric* if its step distribution is invariant by inversion (that is, if $\mu(y) = \mu(y^{-1})$ for every $y \in \Gamma$), and it is *irreducible* if every element y of the group is *accessible* from the initial point $X_0 = 1$ (that is, if $P\{X_n = y\} > 0$ for some $n = 1, 2, \cdots$). Obviously, a sufficient condition for irreducibility of a nearest-neighbor random walk is that the step distribution μ attach positive probability to every element of the generating set A. For a *symmetric* random walk, there is no loss of generality in assuming that this latter condition holds, because if a nearest-neighbor random walk is irreducible, then the (symmetric) subset of the generating set A on which μ is positive must itself be a generating set.

The *period* of an irreducible random walk is the greatest common divisor of the set

$$\{n \geq 1 \,:\, p_n(1,1) > 0\};$$

the random walk is said to be *aperiodic* if the period is 1. Every *symmetric* random walk must have period either 1 or 2, because symmetry forces $p_2(1,1) > 0$. It is sometimes a nuisance to have to account for periodicity, but fortunately there is a simple device for reducing many questions — in particular, questions about "where the random walk goes" — about periodic random walks to questions about aperiodic random walks. This works as follows.

Let X_n be a random walk with step distribution μ, and let $Y_1, Y_2, \cdots$ be an independent sequence of Bernoulli($\frac{1}{2}$) random variables (i.e., coin tosses). Define

$$X'_n = X_{S_n} \quad \text{where} \quad S_n = \sum_{i=1}^{n} Y_i;$$

then the sequence X'_n is an *aperiodic* random walk, with step distribution $(\mu + \delta_1)/2$, that follows the same trajectory as the original random walk X_n. The modified random walk X'_n is called the *lazy* version of X_n, because in effect, at each step the random walker tosses a coin to decide whether to move or to stay put. Although the lazy random walk follows the same trajectory as the original random walk, it obviously does not do so at the same speed; in fact, the step distributions are related as follows:

$$P\{X'_n = y\} = \sum_{m=0}^{n} \binom{n}{m} 2^{-n} P\{X_m = y\} = \sum_{m=0}^{n} \binom{n}{m} 2^{-n} \mu^{*m}(y).$$

1.3 Recurrence and Transience

Definition 1.6. A random walk X_n on a finitely generated group Γ is said to be *recurrent* if the event $\{X_n = 1 \text{ for some } n \geq 1\}$ that the random walk eventually returns to its initial point has probability 1; otherwise, the random walk is said to be *transient*.

The distinction between recurrent and transient random walks is of obvious interest, but will not be a principal concern of these notes. Most interesting random walks on most infinite groups are, as we will see, transient, and so most of our effort will go to the study of transient random walks. Nevertheless, we will mention in passing the following important criterion of G. POLYA for determining when a random walk is recurrent.

Proposition 1.7. *A random walk X_n on a finitely generated group is recurrent if and only if*

$$\sum_{n=0}^{\infty} P\{X_n = 1\} = \infty. \tag{1.6}$$

Proof. For each $n = 1, 2, \cdots$, let F_n be the event that the number of returns to the initial state $X_0 = 1$ is at least n. Then the total number of returns to the initial point can be written as

$$\sum_{n=1}^{\infty} \mathbf{1}\{X_n = 1\} = \sum_{n=1}^{\infty} \mathbf{1}_{F_n}.$$

Consequently,

$$\sum_{n=0}^{\infty} P\{X_n = 1\} = 1 + \sum_{n=1}^{\infty} P(F_n) = \sum_{n=0}^{\infty} P(F_1)^n = 1/(1 - P(F_1)),$$

and so $P(F_1) = 1$ if and only if the sum in (1.6) is infinite.

The missing step in this argument is the identity $P(F_n) = P(F_1)^n$. That this holds should be intuitively clear: each time the random walk returns to its starting point, it "begins afresh", so the conditional probability that it returns again is the same as the *unconditional* probability that it returns at least once. To fashion a rigorous proof, break the event F_n into elementary pieces, that is, write it as a union of cylinder events

$$C(x_0, x_1, x_2, \cdots, x_m) = C := \{X_i = x_i \ \forall\, 0 \le i \le m\} \tag{1.7}$$

where $x_1, x_2, \cdots, x_m$ is a finite sequence in Γ with exactly n entries $x_i = 1$, the last at time $n = m$. For each such cylinder C, the event $F_{n+1} \cap C$ occurs if and only if C occurs and the sequence of partial products

$$\xi_{m+1}, \ \xi_{m+1}\xi_{m+2}, \cdots$$

returns to 1. Since the random variables ξ_i are independent and identically distributed, it follows that

$$P(C \cap F_{n+1}) = P(C)P(F_1).$$

Summing over the cylinder events that constitute F_n gives $P(F_{n+1}) = P(F_{n+1} \cap F_n) = P(F_1)P(F_n)$.

□

1.4 Simple Random Walk on $\mathbb{Z}^d$

The *simple random walk* on the integer lattice $\mathbb{Z}^d$ is the nearest-neighbor random walk with the uniform step distribution on the natural set of generators, i.e.,

$$\mu(\pm e_i) = \frac{1}{2d} \quad \text{for } i = 1, 2, \cdots, d.$$

For random walks on $\mathbb{Z}^d$, it is customary to use additive rather than multiplicative notation for the group operation; thus, we denote the random walk by

$$S_n = \sum_{i=0}^{n} \xi_i$$

where $\xi_1, \xi_2, \cdots$ are independent, identically distributed with common distribution μ.

Polya's Theorem . *The simple random walk on $\mathbb{Z}^d$ is recurrent in dimensions $d = 1, 2$ and transient in dimensions $d \geq 3$.*

Proof Sketch. Simple random walk on $\mathbb{Z}^d$ can only return to 0 at even times, and the return probabilities at even times obey the "local limit law"

$$P\{S_{2n} = 0\} \sim \frac{C_d}{n^{d/2}},$$

where C_d is a positive constant depending on the dimension (see section 7.3 in the Appendix). Thus, the return probabilities are summable if and only if $d > 2$. $\square$

1.5 Random Walks on Homogeneous Trees

For ease of discussion, let's limit our discussion to random walks on the free product group $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$. Recall that the Cayley graph C_Γ (with respect to the natural generating set $A = \{a, b, c\}$) is the the homogeneous tree $\mathbb{T}_3$ of degree 3. The vertices of this tree are identified with group elements, which in turn are finite, reduced words on the alphabet A (*reduced* means that no letter appears twice in succession). The empty word $\emptyset$ is the group identity, and can be viewed as the *root vertex* of the tree.

The edges of the tree $\mathbb{T}_3$ can be assigned labels from the alphabet A in a natural way, as follows. For any two adjacent words (=vertices) w and w', one is an extension of the other by exactly one letter, e.g.,

$$\begin{aligned} w &= x_1 x_2 \cdots x_m \quad \text{and} \\ w' &= x_1 x_2 \cdots x_m x_{m+1}. \end{aligned}$$

For any such pairing, label the edge connecting w and w' by the "color" x_{m+1}. The edge-colorings and the word representations of the vertices then complement each other, in that for any word(=vertex) $w = x_1 x_2 \cdots x_m$, the letters x_i indicate the sequence of edges crossed along the unique self-avoiding path in the tree from $\emptyset$ to w.

The step distribution of a nearest-neighbor random walk on Γ is a probability distribution $\mu = \{p_a, p_b, p_c\}$ on the alphabet A. The random walk X_n with step distribution μ is irreducible if and only if each $p_i > 0$. The main result of section 4 below will imply the following.

Proposition 1.8. *Every irreducible, nearest-neighbor random walk on the group $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$ is transient.*

In the special case where the step distribution μ is the uniform distribution on A an easy proof can be given, using the strong law of large numbers. (The random walk on Γ with the uniform step distribution is said to be *isotropic*.)

Strong Law of Large Numbers . *If $\xi_1, \xi_2, \cdots$ are independent, identically distributed real random variables such that $E|\xi_1| < \infty$, then with probability 1,*

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \xi_i = E\xi_1. \tag{1.8}$$

Proof of Proposition 1.8 for Isotropic Random Walk. Let X_n be the isotropic nearest-neighbor random walk on Γ, and set $Z_n = |X_n|$ where for any element $w \in \Gamma$ we denote by $|w|$ the distance from w to the root $\emptyset$, i.e., the length of the reduced word w. It is easily checked that the sequence Z_n is a nearest-neighbor Markov chain on the set $\mathbb{Z}_+$ of nonnegative integers, with transition probabilities

$$\begin{aligned} q(x, x+1) &= 2/3 \quad \text{for } x \geq 1; \\ q(x, x-1) &= 1/3 \quad \text{for } x \geq 1; \\ q(0, 1) &= 1. \end{aligned}$$

The steps $\zeta_{n+1} = Z_{n+1} - Z_n$ of this Markov chain are independent, identically distributed random variables ζ_{n+1} *except* when $Z_n = 0$, at which time the next increment must always be $+1$. However, whenever $Z_n = 0$ we are free to generate an independent $(\frac{2}{3}, \frac{1}{3})$ coin toss ζ'_n which the process Z_n will ignore; setting $\zeta'_i = \zeta_i$ for all other times i, we obtain a sequence ζ'_n of i.i.d. random variables with common distribution

$$\begin{aligned} P\{\zeta'_n = +1\} &= 2/3, \\ P\{\zeta'_n = -1\} &= 1/3 \end{aligned}$$

such that

$$Z_n \geq \sum_{i=1}^{n} \zeta'_i.$$

The strong law of large numbers implies that the sums on the right side of this inequality converge to $+\infty$ almost surely. Consequently, the same must be true for the sequence Z_n, and so the random walk X_n will only return to the initial point 1 finitely often. $\square$

1.6 Lamplighter Random Walks

Imagine an infinitely long street with streetlamps regularly placed, one at each integer point. Each lamp can be either *on* or *off* (henceforth *on* will be denoted by 1 and *off* by 0). At time 0, all lamps are off. A World Cup fan, who has perhaps been celebrating a bit too long, moves randomly from streetlamp to streetlamp, randomly turning lamps on or off. His[2] behavior at each time $n = 1, 2, \cdots$ is governed by the following rules: he first tosses a fair coin twice, then

(i) if HH he moves one step to the right;
(ii) if HT he moves one step to the left;
(iii) if TH he flips the switch of the lamp at his current configuration;
(iv) if TT he does nothing.

At any time $n = 0, 1, 2, \cdots$, the state of the street is described by the pair $X_n = (S_n, L_n)$, where S_n is the position of the random walker, and $L_n \in \{0, 1\}^{\mathbb{Z}}$ describes the current configuration of the lamps. Note that at each time n, with probability 1, the lamp configuration L_n has all but finitely many entries 0.

The lamplighter process (S_n, L_n) just described is, in fact, itself a symmetric, nearest-neighbor random walk on a finitely generated group Γ known as the *lamplighter group* or the *wreath product* $\mathcal{G}_1 = \mathbb{Z} \wr \oplus_{\mathbb{Z}}\mathbb{Z}_2$, where $\oplus_{\mathbb{Z}}\mathbb{Z}_2$ is the additive group of $0-1$ configurations on $\mathbb{Z}$ with only finitely many 1s. Elements of the group are pairs (x, ψ), where $x \in \mathbb{Z}$ and $\psi \in \oplus_{\mathbb{Z}}\mathbb{Z}_2$; multiplication is defined by

$$(x, \psi) * (y, \varphi) = (x + y, \sigma^{-x}\psi + \varphi)$$

where the addition of configurations is entrywise modulo 2 and σ is the shft operator on configurations. A natural set of generators is the 3-element set

$$\{(1, \mathbf{0}), (-1, \mathbf{0}), (0, \delta_0)\},$$

where $\mathbf{0}$ is the configuration of all 0s and δ_y is the configuration with a single 1 at location x and 0s elsewhere. The step distribution of the random walk obeying (i)–(iv) above is the uniform distribution on the 4-element set gotten by adding the identity $(0, 0)$ to the set of generators.

The lamplighter group is of interest in large part because it (and its higher-dimensional analogues $\mathcal{G}_d := \mathbb{Z}^d \oplus_{\mathbb{Z}^d} \mathbb{Z}_2$) are *amenable* (see section 4 below) but have *exponential growth* (see section 2.2).

Proposition 1.9. *The lamplighter random walk on $\mathcal{G}_1$ is* transient.

The proof is a bit involved, but requires only elementary probability inequalities, notably the *Hoeffding bound*, which is discussed in the Appendix. There is, however, a simple modification of the lamplighter random walk for which transience is much easier to prove. In this random walk, the soccer hooligan behaves as follows. At each time $n = 1, 2, 3, \cdots$ he tosses a fair coin 3 times. The first toss tells him whether or not to flip the switch of the

[2]Nearly all soccer hooligans are male.

lamp at his current position x; the second toss tells him whether to then move right or left (to $x+1$ or $x-1$); and the third toss tells him whether or not to flip the switch of the lamp at his new location. The state of the street at time n is once again described by a pair (S_n, L_n), with S_n describing the hooligan's position and L_n the configuration of the streetlamps. This process (S_n, L_n) is no longer a nearest-neighbor random walk with respect to the natural set of generators listed above, but it is a symmetric, nearest-neighbor random walk relative to a different set of generators. (Exercise: Write out the list of generators for the modified lamplighter random walk.)

Proposition 1.10. *The modified lamplighter random walk (S_n, L_n) is transient.*

Proof Sketch. The sequence of positions S_n is itself a simple random walk on $\mathbb{Z}$. Define two related sequences:

$$S_n^+ := \max_{m \leq n} S_m \quad \text{and}$$
$$S_n^- := \min_{m \leq n} S_m.$$

Exercise 1.11. Prove that for each sufficiently small $\varepsilon > 0$ there exist $\delta > 0$ and $C < \infty$ such that

$$P\{S_n^+ - S_n^- \leq n^\varepsilon\} \leq Cn^{-1-\delta}.$$

HINT: Divide the time interval $[0, n]$ into thirds, and use the Local Central Limit Theorem (alternatively, Stirling's Formula) to estimate the probability that the change of S_j acoss any one of these three intervals is less (in absolute value) than n^ε.

The relevance of the random variables S_n^+ and S_n^- is this: at each time the hooligan leaves a site, he randomizes the state of the lamp at that site. Thus, conditional on the trajectory $(S_m)_{m \leq n}$, the distribution of the configuration in the segment $[S_n^-, S_n^+]$ will be uniform over all possible configurations, of which there are $2^{S_n^+ - S_n^- + 1}$. Therefore,

$$P\{(S_n, L_n) = (0, \mathbf{0})\} \leq P\{S_n^+ - S_n^- \leq n^\varepsilon\} + \frac{1}{2^{n^\varepsilon}} \leq Cn^{-1-\delta} + \frac{1}{2^{n^\varepsilon}}.$$

This sequence is summable, so Polya's criterion implies that the random walk is transient. □

2 Speed, Entropy, and Laws of Large Numbers

2.1 Speed

Every nearest-neighbor random walk on a finitely generated group travels at a definite *speed* (possibly 0): this is the content of our first major theorem. Furthermore, this is true relative to any metric on the group, not just the word metric, although the speed will depend on both the metric d and the step distribution μ of the random walk.

Theorem 2.1. *For any nearest-neighbor random walk X_n on a finitely generated group Γ, and for any metric d on Γ, with probability one,*

$$\lim_{n\to\infty} \frac{d(X_n, 1)}{n} = \inf_{n\geq 1} \frac{Ed(X_n, 1)}{n}. \tag{2.1}$$

In the special case when d is the word metric, we set

$$\ell := \inf_{n\geq 1} \frac{Ed(X_n, 1)}{n}, \tag{2.2}$$

and call this constant the *speed* of the random walk.

In section 2.3 we will prove a weaker statement than (2.1), specifically, that $d(X_n, 1)/n$ converges to ℓ in probability. Almost sure convergence is a bit more subtle (but can still be carried out by elementary arguments). Theorem 2.1 can be viewed as an extension of the usual strong law of large numbers for random walks on $\mathbb{Z}$ to arbitrary discrete groups.

The strong law of large numbers, applied componentwise, implies that for any random walk on $\mathbb{Z}^d$ whose step distribution is *symmetric* and has finite first moment, the speed is 0, because symmetry implies that $E\xi_1 = -E\xi_1 = 0$. It is not always true, though, that a symmetric random walk on a finitely generated group has speed 0 — the isotropic nearest-neighbor random walk on $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$, as we have seen, has speed $\frac{1}{3}$. This suggests an obvious question: for which symmetric random walks, on which groups, is the speed 0, and for which is it positive? One of our main goals in these lectures will be to answer this question, at least partially.

2.2 Subadditivity

A sequence a_n of real numbers is said to be *subadditive* if for every pair m, n of indices,

$$a_{m+n} \leq a_n + a_m.$$

Exercise 2.2. Show that for any subadditive sequence a_n,

$$\alpha := \lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\geq 1} \frac{a_n}{n} \geq -\infty.$$

Example 2.3. Let Γ be a finitely generated group with associated Cayley graph G_Γ. Define B_n to be the ball of radius n in C_Γ with center at the group identity 1; thus, B_n contains all group elements x such that $w(x) \leq n$. Because Γ is finitely generated, each ball B_n contains only finitely many group elements. Furthermore, by the triangle inequality, for any integers $m, n \geq 0$,

$$|B_{n+m}| \geq |B_n||B_m|.$$

Thus, the sequence $-\log|B_n|$ is subadditive, and so the limit

$$\beta := \lim_{n\to\infty} \frac{\log|B_n|}{n} \tag{2.3}$$

exists and is nonnegative. The quantity β is known as the *exponential growth rate* of the group relative to the generating set A. Clearly, $\beta \leq \log 2g$, where $2g$ is the number of generators (= number of nearest neighbors of 1 in the Cayley graph).

Exercise 2.4. It is easily checked that the exponential growth rate of $\mathbb{Z}^d$ is 0, while the exponential growth rate of the tree $\mathbb{T}_d$ is $\log(d-1)$. What is the exponential growth rate of the lamplighter group $\mathbb{Z} \wr \mathbb{Z}_2^{\mathbb{Z}}$? NOTE: There are two natural generating sets for the lamplighter group. The exponential growth rate will depend on which generating set is used.

Associated with any nearest-neighbor random walk X_n on a finitely generated group Γ are a number of important subadditive sequences. The limit constants α for these sequences are invariants of basic interest.

Example 2.5. For any metric d on the group Γ, the sequence $Ed(X_n, 1)$ is subadditive. Consequently, the limit

$$\ell(d) := \lim_{n\to\infty} \frac{Ed(X_n, 1)}{n} \tag{2.4}$$

exists and satisfies $0 \leq \ell(d) < \infty$. If d is the word metric and the random walk is nearest-neighbor, then $\ell = \ell(d) \leq 1$.

Example 2.6. The word metric (i.e., the usual graph distance in the Cayley graph) is the first that comes to mind, but it is not the only metric on Γ of interest. Here is another, the *Green metric*, which is defined for any *symmetric, transient* random walk. For $x \in \Gamma$, define

$$u(x) = P\{X_n = x \text{ for some } n \geq 0\}.$$

Clearly, $u(xy) \geq u(x)u(y)$ for any elements $x, y \in \Gamma$, because the random walker can reach xy by first going to x and then to xy, and so the function

$$d_G(x, y) := -\log u(x^{-1}y)$$

is a metric. (Observe that the hypothesis of *symmetry* guarantees that $d_G(x, y) = d_G(y, x)$.) The limit

$$\gamma := \lim_{n\to\infty} \frac{Ed_G(X_n, 1)}{n} \tag{2.5}$$

is a measure of the the exponential rate of "breakup" of the group as seen by the random walk.

Example 2.7. Let $p_n(1, 1) = P\{X_n = 1\}$ be the $n-$step return probability for a random walk X_n with period d. If $d \geq 2$, return to the initial point 1 will only be possible at times which are integer multiples of d, so let's consider the subsequence $p_{dn}(1, 1)$. Since the event of return at time $dm + dn$ contains the event that the random walk returns at time dm and then again at time $dn + dm$, the sequence of return probabilities is *super-multiplicative*, that is, $p_{dm+dn}(1, 1) \geq p_{dm}(1, 1)p_{dn}(1, 1)$. Consequently, the limit

$$\varrho := \lim_{n\to\infty} p_{dn}(1, 1)^{1/dn} \leq 1 \tag{2.6}$$

exists. This limit ϱ is called the *spectral radius* of the random walk; it will be studied in section 4 below.

Example 2.8. Denote by μ the step distribution of the random walk and by μ^{*n} its nth convolution power, that is, $\mu^{*n}(x) = P\{X_n = x\}$. Consider the random variable $\mu^{*n}(X_n)$ gotten by evaluating the function μ^{*n} at the random point X_n: this should be thought of as the conditional probability that a second, independent random walk X'_n, started at the group identity 1 and run for the same number n of steps, would end up at the same location as the first. Since the event $\{X'_{m+n} = X_{m+n}\}$ contains the event $\{X'_m = X_m\} \cap \{X'_{m+n} = X_{m+n}\}$, we have

$$\mu^{*(m+n)}(X_{m+n}) \geq \mu^{*m}(X_m)\mu^{*n}(X_n),$$

so the sequence $-E \log \mu^{*n}(X_n)$ is subadditive.[3] The limit

$$h := \lim_{n\to\infty} \frac{-E\log\mu^{*n}(X_n)}{n} \tag{2.7}$$

is called the *Avez entropy* of the random walk.

Exercise 2.9. (A) Show that for a *symmetric* random walk, if the spectral radius is strictly less than 1 then the Avez entropy is strictly positive. (B) Show that for the asymmetric p, q random walk on the integers $\mathbb{Z}$ (i.e., the random walk that at each step jumps $+1$ with probability p and -1 with probability q) the spectral radius is less than 1 but the Avez entropy is 0. HINT: For (B) you will want to use the *Hoeffding inequality*. See section 7.2 for the statement.

Exercise 2.10. Show that if a random walk has positive entropy h, then its lazy version also has positive entropy.

Exercise 2.11. Show that the speed, Avez entropy, and the exponential growth rate of the group satisfy the basic inequality

$$h \leq \beta\ell.$$

One of the interesting open problems in the subject is to characterize those groups for which equality can hold in this relation. HINT: Show that the *Shannon entropy* of a probability measure on a finite set $\mathcal{Y}$ is maximal for the uniform distribution. Note: The Shannon entropy of a probability measure ν on $\mathcal{Y}$ is defined (cf. section 6.1 below) by

$$H(\mu) := -\sum_{y\in\mathcal{Y}} \nu(y)\log\nu(y).$$

Exercise 2.12. Assume that $P\{X_1 = 1\} > 0$. This assumption ensures that if $\mu^{*n}(y) > 0$ then $\mu^{*(n+1)}(y) > 0$.

(A) Show that for each $\alpha \geq 1$ the limit

$$\psi(\alpha) := -\lim_{n\to\infty}(\alpha n)^{-1}E\log\mu^{*[\alpha n]}(X_n) \quad \text{exists.}$$

(B) Show that the function $\psi(\alpha)$ (called the *entropy profile*) is convex.
(C) Show that $\lim_{\alpha\to\infty}\psi(\alpha) = -\log\varrho$ where $\varrho =$ spectral radius.
(D) What is the entropy profile of the simple random walk on $\mathbb{Z}$?

[3] For any group element x such that $P\{X_n = x\} > 0$ it must be the case that $\mu^{*n}(x) \geq \eta^n$, where $\eta > 0$ is the minimum of $\mu(y)$ over all $y \in \Gamma$ such that $\mu(y) > 0$. Hence, the random variable $-\log\mu^{*n}(X_n)$ is bounded above by $-\log\eta$, and so the expectation $-E\log\mu^{*n}(X_n)$ is well-defined.

2.3 Subadditive Law of Large Numbers

In several of the examples presented in the preceding section, the entries of the subadditive sequence a_n were obtained by taking the expectation of a function w_n of the first n entries of the sequence $\xi_1, \xi_2, \cdots$ of steps of the random walk. In each case the functions w_n satisfied a natural system of inequalities that led to the subadditivity of the expectations. This pattern presents itself sufficiently often that it deserves a name.

Definition 2.13. A sequence of functions $w_n : \mathbb{Y}^n \to \mathbb{R}$ is *subadditive* if for any sequence $y_1, y_2, y_3, \cdots \in \mathbb{Y}$ and each pair m, n of nonnegative integers,

$$w_{n+m}(y_1, y_2, \cdots, y_{m+n}) \leq w_m(y_1, y_2, \cdots y_m) + w_n(y_{m+1}, y_{m+2}, \cdots, y_{m+n}). \tag{2.8}$$

Theorem 2.14. *If $Y_1, Y_2, \cdots$ is a sequence of independent, identically distributed random variables taking values in a (measurable) space $\mathbb{Y}$, and if $w_n : \mathbb{Y}^n \to \mathbb{R}$ is a subadditive sequence of (measurable) functions such that*

$$E(w_1(Y_1))_+ < \infty$$

then with probability one,

$$\lim_{n\to\infty} \frac{w_n(Y_1, Y_2, \cdots, Y_n)}{n} = \inf_{m\geq 1} \frac{Ew_m(Y_1, Y_2, \cdots, Y_m)}{m} := \alpha. \tag{2.9}$$

The conclusion holds more generally when the sequence $Y_1, Y_2, \cdots$ is a *stationary, ergodic* sequence. This generalization is known as *Kingman's subadditive ergodic theorem*.

In all of the applications below, the space $\mathbb{Y}$ will be a finite set – in most cases, $\mathbb{Y}$ can be taken to be the set of generators of the group Γ in which the random walk lives. When $\mathbb{Y}$ is finite, measurability is automatic (relative to the maximal sigma algebra on $\mathbb{Y}$), and since the random variables $w_n(Y_1, Y_2, \cdots, Y_n)$ take only finitely many values they all have finite first moments. It is still possible, though, that the limit α in (2.9) could be $-\infty$ (for instance, if $w_n \equiv -n^2$).

Proof of Theorem 2.14. We will prove only the weaker statement that the random variables

$$w_n(Y_1, Y_2, \cdots, Y_n)/n$$

converge *in probability* to the constant α, and only under the more restrictive hypothesis that the function $|w_1|$ is bounded. This hypothesis holds for the subadditive sequences exhibited in each of the examples discused in section 2.2, provided the random walk in question has step distribution supported by a finite generating set of the group Γ. Without loss of generality we can assume that $|w_1| \leq 1$, because multiplying all of the functions w_n by a scalar does not affect the validity of the result. This implies, by subadditivity, that for every $n \geq 1$ and all $y_i \in \mathbb{Y}$,

$$|w_n(y_1, y_2, \cdots, y_n)| \leq \sum_{i=1}^{n} |w_1(y_i)| \leq n.$$

For notational ease, let's use the abbreviation

$$W(m;n) = w_n(Y_{m+1}, Y_{m+2}, \cdots, Y_{m+n});$$

then the subadditivity relation reads

$$W(0;mn) \le \sum_{i=0}^{m-1} W(in;n) \quad \text{or, more generally,}$$

$$W(0;mn+k) \le \sum_{i=0}^{m-1} W(in;n) + W(mn;k).$$

The crucial fact here is that for each choice of $n \ge 1$ the random variables $(W(in;n))_{i=0,1,\cdots}$ are independent and identically distributed. Thus, the strong law of large numbers implies that for each $n \ge 1$, with probability one,

$$\limsup_{m\to\infty} \max_{0\le k<n} \frac{W(0;mn+k)}{mn} \le \lim_{m\to\infty} \left\{ (mn)^{-1} \sum_{i=0}^{m-1} W(in;n) + (mn)^{-1} n \right\} = \frac{EW(0;n)}{n}.$$

Here we have used the hypothesis that $|W(j;k)| \le k$ to bound the term $(mn)^{-1}W(mn;k)$ in the subadditivity relation. It now follows that for each $n \ge 1$, with probability one,

$$\limsup_{m\to\infty} \frac{W(0;m)}{m} \le \frac{EW(0;n)}{n},$$

and therefore

$$\limsup_{m\to\infty} \frac{W(0;m)}{m} \le \alpha := \inf_{n\ge1} \frac{EW(0;n)}{n} \quad \text{almost surely.} \tag{2.10}$$

To complete the argument we must show that the random variables $W(0;m)/m$ cannot drop much below α with any appreciable probability. For this, we will rely on the hypothesis that $|w_1| \le 1$, which by subadditivity implies that for each $m \ge 1$ we have $|W(0;m)| \le m$. Fix $\varepsilon > 0$; by relation (2.10), if m is sufficiently large then

$$P\{W(0;m)/m > \alpha + \varepsilon\} \le \varepsilon.$$

Since $|W(0;m)| \le m$, it follows that for any $\delta > 0$,

$$\begin{aligned} EW(0;m) &= EW(0;m)\mathbf{1}\{-m\delta \le W(0;m) - m\alpha \le m\varepsilon\} \\ &\quad + EW(0;m)\mathbf{1}\{W(0;m) > m\alpha + m\varepsilon\} \\ &\quad + EW(0;m)\mathbf{1}\{W(0;m) < m\alpha - m\delta\} \\ &\le m\alpha + m\varepsilon + m\varepsilon - m\delta P\{W(0;m) < m\alpha - m\delta\}. \end{aligned}$$

But $EW(0;m) \ge m\alpha$, by definition of α, so we must have, for all sufficiently large m, that

$$\delta P\{W(0;m) < m\alpha - m\delta\} < 2\varepsilon.$$

Since ε can be chosen arbitrarily small relative to δ, it follows that

$$\lim_{m\to\infty} P\{W(0;m) < m\alpha - m\delta\} = 0.$$

□

Corollary 2.15. *Let X_n be a nearest-neighbor random walk on a finitely generated group Γ, with step distribution μ. Let d be the word metric and d_G be the Green metric. Then as $n \to \infty$,*

$$d(X_n, 1)/n \xrightarrow{P} \ell := \inf_{m \geq 1} Ed(X_m, 1)/m;$$

$$d_G(X_n, 1)/n \xrightarrow{P} \gamma := \inf_{m \geq 1} Ed_G(X_m, 1)/m; \quad \textit{and}$$

$$-\log \mu^{*n}(X_n)/n \xrightarrow{P} h := \inf_{m \geq 1} (-E \log \mu^{*n}(X_n)/n)$$

2.4 The Kesten-Spitzer-Whitman Theorem

Another interesting subadditive sequence associated with a random walk $X_n = \xi_1 \xi_2 \cdots \xi_n$ on a group is the sequence R_n (for *range*) that counts the number of distinct sites visited by time n, formally,

$$R_n := |\{X_0, X_1, \cdots, X_n\}| = |\{1, \xi_1, \xi_1\xi_2, \cdots, \xi_1\xi_2 \cdots \xi_n\}| \tag{2.11}$$

The subadditive law of large numbers (Theorem 2.14) implies that for any random walk on any group, the averages R_n/n converge almost surely to $\inf ER_m/m$. Kesten, Spitzer, and Whitman showed that this limit has an interesting interpretation.

Theorem 2.16. *For any random walk X_n on any group,*

$$\lim_{n \to \infty} \frac{R_n}{n} = P\{\textit{no return to } 1\} := P\{X_n \neq 1 \; \forall \, n \geq 1\} \quad \textit{almost surely.} \tag{2.12}$$

Recall that a random walk X_n on a finitely generated group is *recurrent* if and only if the probability of return to the initial point is 1, equivalently, if $P\{\text{no return to } 1\} = 0$. Consequently, the Kesten-Spitzer-Whitman theorem implies that X_n is recurrent if and only if the range R_n grows *sublinearly* with n.

Corollary 2.17. *Any irreducible random walk $S_n = \sum_{i=1}^n \xi_i$ on the integers $\mathbb{Z}$ whose step distribution has finite first moment $E|\xi_1|$ and mean $E\xi_1 = 0$ is recurrent.*

Note: It is not required that the random walk be nearest-neighbor, or even that the step distribution have finite support. In fact, both the corollary and the KSW theorem hold not only for random walks with independent, identically distributed steps ξ_i, but more generally for random walks whose steps ξ_i form an *ergodic, stationary* sequence.

Proof of the Corollary. If $E\xi_i = 0$, then the strong law of large numbers implies that the sequence S_n/n converges to 0 almost surely. This implies that for any $\varepsilon > 0$, all but at most finitely many terms of the sequence S_n/n will fall in the interval $(-\varepsilon, \varepsilon)$; and this in turn implies that for large n the range R_n will satisfy

$$R_n \leq 2n\varepsilon.$$

Since $\varepsilon >$ is arbitrary, it follows that $R_n/n \to 0$. By the KSW theorem, the random walk must be recurrent. $\square$

Proof of the KSW Theorem. By the subadditive law of large numbers, $R_n/n \to \inf ER_m/m$, so it suffices to show that the expectations ER_n/n converge to $P\{\text{no return to } 1\}$. Now R_n is the number of distinct sites visited by time n; these are in one-to-one correspondence with the times $j \leq n$ that the random walk is at a site X_j that will not be visited again by time n. Hence,

$$\begin{aligned} R_n &= \sum_{j=0}^{n} \mathbf{1}\{X_j \text{ not revisited by time } n\} \\ &\geq \sum_{j=0}^{n} \mathbf{1}\{X_j \text{ never revisited}\} \\ &= \sum_{j=0}^{n} \mathbf{1}\{\xi_{j+1}\xi_{j+2}\cdots\xi_{j+m} \neq 1 \ \forall \mu \geq 1\}. \end{aligned}$$

Since the random variables $\xi_1, \xi_2, \cdots$ are i.i.d., the events in the last sum all have the same probability, specifically, the probability that the random walk never returns to its initial site 1. Thus, taking expectations and letting $n \to \infty$ gives

$$\liminf_{n\to\infty} ER_n/n \geq P\{\text{no return to } 1\}.$$

A similar argument proves the upper bound. Fix an integer $K \geq 1$; then for any $0 \leq j \leq n-K$, the event that X_j is not revisited by time n is contained in the event that X_j is not revisited in the K steps following time j. Consequently,

$$\begin{aligned} R_n &= \sum_{j=0}^{n} \mathbf{1}\{X_j \text{ not revisited by time } n\} \\ &\leq \sum_{j=0}^{n-K} \mathbf{1}\{X_j \neq X_{j+i} \ \forall\, 1 \leq i \leq K\} + K. \end{aligned}$$

Once again, the events in the sum all have the same probability, so taking expectations, dividing by n, and then letting $n \to \infty$ yields

$$\limsup ER_n/n \leq P\{1 \neq X_i \ \forall\, 1 \leq i \leq K\}.$$

Since K is arbitrary, it follows that

$$\limsup ER_n/n \leq \inf_{K\geq 1} P\{1 \neq X_i \ \forall\, 1 \leq i \leq K\} = P\{\text{no return to } 1\}.$$

□

3 The Carne-Varopoulos Inequality

3.1 Statement and Consequences

In 1985 N. VAROPOULOS discovered a remarkable upper bound for the $n-$step transition probabilities of a reversible Markov chain, and shortly afterward T. K. CARNE found an elegant short proof. Every *symmetric* nearest-neighbor random walk on a finitely generated group is a reversible Markov chain, so Varopoulos' inequality applies; the upshot is as follows.

Theorem 3.1. *Let X_n be a symmetric, nearest-neighbor random walk on a finitely generated group Γ with word metric d (i.e., the usual metric on the Cayley graph C_Γ). Let $\varrho = \lim P^1\{X_{2n} = 1\}^{1/2n}$ be the spectral radius of the random walk. Then for every element $x \in \Gamma$ and every $n \geq 1$,*

$$P^1\{X_n = x\} \leq 2\varrho^n \exp\{-d(1,x)^2/2n\} \tag{3.1}$$

The hypothesis of symmetry is crucial: the theorem fails badly for non-symmetric random walks. Consider, for instance the random walk on $\mathbb{Z}$ that at each step moves $+1$ to the right.

Before turning to the proof of the Carne-Varopoulos inequality, let's look at one of its implications.

Corollary 3.2. *If X_n is a symmetric, nearest-neighbor random walk on a finitely generated group Γ with speed ℓ and Avez entropy h, then*

$$h \geq \ell^2/2 - \log \varrho \tag{3.2}$$

Proof. By Corollary 2.15, for any $\varepsilon > 0$ and all sufficiently large n the probability will be nearly 1 that the distance $d(X_n, 1)$ will lie in the range $n\ell \pm n\varepsilon$. By the Carne-Varopoulos inequality, for any group element x at distance $n\ell \pm n\varepsilon$ from the group identity 1,

$$P\{X_n = x\} \leq 2\varrho^n \exp\{-d(x,1)^2/2n\} \leq 2\varrho^n \exp\{-n(\ell - \varepsilon)^2/2\}.$$

Thus, with probability approaching 1 as $n \to \infty$,

$$\mu^{*n}(X_n) \leq 2\varrho^n \exp\{-n(\ell - \varepsilon)^2/2\}.$$

Since $\varepsilon > 0$ is arbitrary, the inequality (3.2) follows. □

Corollary 3.2 implies that a random walk can have positive speed only if it also has positive Avez entropy. The next result shows that the converse is also true.

Proposition 3.3. *For any symmetric, nearest-neighbor random walk on a finitely generated group,*

$$\ell > 0 \iff h > 0.$$

Proof. Recall (Example 2.3) that any finitely generated group Γ has at most exponential growth: in particular, there is a constant $\beta \geq 0$ such that for large n the cardinality of the ball B_n of radius n satisfies $\log |B_n| \sim n\beta$. If a symmetric random walk X_n on Γ has speed $\ell = 0$, then for any $\varepsilon > 0$ and all sufficiently large n the distribution of the random variable X_n will be almost entirely concentrated in $B_{\varepsilon n}$, by Corollary 2.15. But since $B_{\varepsilon n}$ contains at most $e^{2\beta\varepsilon n}$ distinct group elements (for large n), it follows that with probability approaching 1,

$$\mu^{*n}(X_n) \geq \exp\{-2\beta\varepsilon n\}.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $h = 0$. □

An easy modification of this argument (exercise!) yields the following related result.

Proposition 3.4. *If a finitely generated group Γ has growth constant $\beta = 0$ then every nearest-neighbor random walk on Γ has entropy $h = 0$, and consequently, every* symmetric, *nearest-neighbor random walk on Γ has speed 0.*

□

3.2 Markov Operators

The $1-$step transition probabilities of a random walk on an infinite, finitely generated group Γ are the entries of an infinite matrix $M = (p(x,y))_{x,y\in\Gamma}$ called the *transition probability matrix* of the random walk. The advantage of viewing the transition probabilities $p(x,y)$ as matrix entries stems from the fact that the $n-$step transition probabilities $p_n(x,y)$, for any integer $n \geq 0$, are the entries of the matrix M^n obtained by multiplying M by itself n times. (This is easily proved by induction on n.) Henceforth, we will identify M with the corresponding linear operator on the Hilbert space $L^2(\Gamma)$ of real-valued, square-summable functions on Γ. This operator is usually called the *Markov operator* associated with the random walk.

Definition 3.5. For any finitely generated group Γ define $L^2(\Gamma)$ to be the real Hilbert space of square-summable functions $f : \Gamma \to \mathbb{R}$ with norm and inner product

$$\|f\|_2^2 := \sum_{x\in\Gamma} f(x)^2 \quad \text{and} \quad \langle f, g\rangle := \sum_{x\in\Gamma} f(x)g(x). \tag{3.3}$$

For a nearest-neighbor random walk X_n on Γ with $1-$step transition probabilities $p(x,y)$ define the associated *Markov operator* $M : L^2(\Gamma) \to L^2(\Gamma)$ by

$$Mf(x) := E^x f(X_1) = Ef(xX_1) = \sum_{y\in\Gamma} p(x,y)f(y). \tag{3.4}$$

Properties of the Markov Operator:

(M1) M is linear.
(M2) $\|f\|_2 \leq 1$ implies $\|Mf\|_2 \leq 1$.
(M3) $M^n f(x) = E^x f(X_n) = \sum_{y\in\Gamma} p_n(x,y)f(y)$.
(M4) $p_n(x,y) = \langle \delta_x, M^n \delta_y\rangle$.
(M5) If $p(x,y) = p(y,x)\ \forall\, x,y \in \Gamma$ then $\langle f, M^n g\rangle = \langle M^n f, g\rangle$.

Here δ_z denotes the function that takes the value 1 at z and 0 elsewhere. Properties (M1)–(M5) are all easily checked. Property (M2) asserts that M has *operator norm* bounded by 1; later we will show that for symmetric random walks the operator norm of M coincides with the *spectral radius* ϱ.

Property (M5), *symmetry*, is of essential importance to Carne's proof of the Carne-Varopoulos inequality (3.1), as it permits use of the *Spectral Theorem* for bounded, self-adjoint operators (see, for instance, REED & SIMON, *Functional Analysis*, vol. 1). To avoid using this theorem, or any other deep results from the theory of linear operators on infinite-dimensional Hilbert spaces, we will employ a simple projection trick that will allow us to deduce everything we need about the spectral properties of M from finite-dimensional linear algebra.

Assumption 3.6. *Assume henceforth that the random walk is symmetric.*

For each $k = 1, 2, \cdots$ let V_k be the finite-dimensional linear subspace of $L^2(\Gamma)$ consisting of those functions f which vanish outside the ball B_k of radius k centered at the group identity 1. Define P_k to be the *orthogonal projection operator* for this subspace, that is, for each $g \in L^2(\Gamma)$, set

$$\begin{aligned} P_k g(x) &= g(x) \quad \text{if} \quad x \in B_k; \\ &= 0 \qquad\ \text{if} \quad x \notin B_k. \end{aligned}$$

Next, define M_k to be the restriction of M to V_k, that is,

$$M_k = P_k M P_k.$$

Clearly, M_k maps V_k into V_k. Furthermore, since the underlying random walk is symmetric, each M_k is a symmetric linear operator on V_k. Explicitly, for any $x \in B_k$,

$$M_k g(x) = \sum_{y \in B_k} p(x,y) g(y),$$

which exhibits $M_k : V_k \to V_k$ as the symmetric linear transformation with matrix representation $(p(x,y))_{x,y \in B_k}$ relative to the natural basis δ_x for V_k.

Lemma 3.7. *The norms of the operators M_k are non-decreasing in k, and*

$$\|M\| = \lim_{k \to \infty} \|M_k\|. \tag{3.5}$$

Note: By definition, the norm of an operator $T : V \to V$ on a real inner product space V is $\|T\| = \sup_{\|v\| \leq 1} \|Tv\|$, where the sup extends over all vectors of norm ≤ 1.

Exercise 3.8. Prove Lemma 3.7. HINT: For the first assertion, use the fact that the vector spaces V_k are nested, that is, $V_1 \subset V_2 \subset \cdots$.

Since the operators M_k are finite-dimensional and symmetric, they fall under the authority of the Spectral Theorem for finite-dimensional symmetric matrices. Recall what this says:

Spectral Theorem . *For any symmetric linear operator $L : V \to V$ on a finite-dimensional real inner product space V there exist a complete orthonormal basis $\{u_i\}_{i \leq D}$ of V and a corresponding set $\{\lambda_i\}_{i \leq D}$ of real numbers such that for every $v \in V$,*

$$Lv = \sum_{i=1}^{D} \langle v, u_i \rangle \lambda_i u_i. \tag{3.6}$$

Easy Consequences:

(Sp1) $Lu_i = \lambda_i u_i$.
(Sp2) $L^n v = \sum_{i \le D} \langle v, u_i \rangle \lambda_i^n u_i$.
(Sp3) $\|L\| := \max_{\|v\| \le 1} \|Lv\| = \max_{i \le D} |\lambda_i|$.

The quantity $\max_{i \le D} |\lambda_i|$ is the *spectral radius* of the operator L. Assertion (Sp3) states that the norm and the spectral radius of a symmetric, finite-dimensional linear operator coincide. Using this together with Lemma 3.7, we can now identify the norm of the Markov operator M.

Proposition 3.9. *The Markov operator M of a* symmetric, *nearest-neighbor random walk on a finitely generated group satisfies*

$$\|M\| = \varrho := \lim_{n \to \infty} p_{2n}(1, 1)^{1/2n}. \tag{3.7}$$

Proof. By property (M4), the return probabilities satisfy

$$p_n(1, 1) = \langle \delta_1, M^n \delta_1 \rangle \le \|\delta_1\|_2^2 \, \|M^n\| = \|M\|^n.$$

Hence, the spectral radius ϱ cannot be larger than $\|M\|$, so to prove the proposition it suffices to show that $\|M\|$ cannot be larger than ϱ. In doing so, we may assume without loss of generality that the random walk has a positive holding probability $\varepsilon = p(1, 1) > 0$, because increasing the probability $p(1, 1)$ by ε has the effect of changing the Markov operator from M to $(1 - \varepsilon)M + \varepsilon I$, which changes neither the norm nor the spectral radius by more than a negligible amount when ε is small.

Exercise 3.10. Check this.

By Lemma 3.7, the norm $\|M\|$ is the increasing limit of the norms $\|M_k\|$; thus, it suffices to show that for each k,

$$\|M_k\| \le \varrho.$$

By the Spectral Theorem, the norm $\|M_k\|$ coincides with the magnitude of the largest eigenvalue $|\lambda_1|$. Denote the corresponding normalized eigenfunction by f_1; by definition of M_k, the support of f_1 must be contained in the ball B_k. Moreover, since the entries of the matrix M_k are nonnegative, the eigenfunction f_1 must also be nonnegative, because

$$\|M_k\| = |\langle f_1, M_k f_1 \rangle| \le \langle |f_1|, M_k |f_1| \rangle,$$

and if the inequality were strict then we would have a contradiction to the definition of $\|M_k\|$. Consequently, for each $n \ge 1$,

$$\begin{aligned} \|M_k\|^n &= \langle f_1, M_k^n f_1 \rangle \\ &\le \langle f_1, M^n f_1 \rangle \\ &= \sum_{x \in B_k} \sum_{y \in B_k} f_1(x) f_1(y) p_n(x, y) \end{aligned} \tag{3.8}$$

Claim: If $p(1,1) > 0$ then there exist an integer $m \geq 0$ and a constant $C < \infty$ such that for all $n \geq 1$ and all $x, y \in B_k$,

$$p_n(x,y) \leq C p_{n+m}(1,1).$$

Exercise: Prove this. HINT: The random walk is irreducible.

Given the claim, we now conclude that for any $n \geq 1$,

$$\sum_{x \in B_k} \sum_{y \in B_k} f_1(x) f_1(y) p_n(x,y) \leq C p_{n+m}(1,1) \sum_{x \in B_k} \sum_{y \in B_k} f_1(x) f_1(y).$$

Taking nth roots in (3.8) and letting $n \to \infty$, we obtain the desired inequality $\|M_k\| \leq \varrho$. □

Ultimately, our interest in the Markov operator M stems from property (M4), which gives an explicit formula for the $n-$step transition probabilities $p_n(x,y)$ in terms of the iterates M^n. The Spectral Theorem enters the picture because it provides a representation of M^n in terms of the spectral properties of M, by way of the formula (Sp2) for $L = M_k$. To complete the connection, we must establish some relation between the powers M^n of the Markov operator and the corresponding powers M_k^n of the finite-dimensional truncations.

Lemma 3.11. *If $k > n \geq 1$ and $x \in B_n$, then $\langle \delta_1, M^n \delta_x \rangle = \langle \delta_1, M_k^n \delta_x \rangle$.*

Proof Sketch. Multiplication by the nth power of a matrix (finite or infinite) can be interpreted as a *sum over paths*. In particular,

$$\langle \delta_1, M^n \delta_x \rangle = \sum_{1 = x_0, x_1, \cdots, x_n = x} \prod_{i=1}^{n} p(x_{i-1}, x_i) \quad \text{and}$$

$$\langle \delta_1, M_k^n \delta_x \rangle = \sum_{1 = x_0, x_1, \cdots, x_n = x \,:\, x_i \in B_k} \prod_{i=1}^{n} p(x_{i-1}, x_i).$$

Here the sums extend over all paths in $\Gamma$of length n beginning at $x_0 = 1$ and ending at $x_n = x$; the second sum (in the equation for $\langle \delta_1, M_k^n \delta_x \rangle$) is restricted to paths that do not exit the ball B_k. But since $p(u,v) > 0$ only for nearest-neighbor pairs $u, v \in \Gamma$, all paths in the first sum for which the product $\prod p(x_{i-1}, x_i)$ is non-zero must remain in the ball B_k. Consequently, the two sums are identical. □

3.3 Chebyshev Polynomials

Definition 3.12. The nth Chebyshev polynomial of the first kind is the unique nth degree polynomial $T_n(x)$ with real (actually, integer) coefficients such that

$$T_n(\cos \theta) = \cos(n\theta). \tag{3.9}$$

The first few are

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x. \end{aligned}$$

Note for later reference that the defining identity (3.9) implies that

$$|T_n(x)| \leq 1 \quad \text{for every} \quad x \in [-1, 1]. \tag{3.10}$$

Exercise 3.13. Take 30 seconds to convince yourself that polynomials satisfying the identities (3.9) exist and are unique. If you want to know more, check the article at mathworld.wolfram.com.

The sudden appearance of these obscure special functions in a series of lectures devoted to random walks on groups has a bit of the rabbit-out-of-a-hat character. However, the next result shows that there is a natural connection between the Chebyshev polynomials and simple random walk on $\mathbb{Z}$.

Proposition 3.14. *Let $(S_n)_{n\geq 0}$ be the simple random walk on $\mathbb{Z}$; then for every $n = 0, 1, 2, \cdots$*

$$x^n = \sum_{m \in \mathbb{Z}} P\{S_n = m\} T_m(x). \tag{3.11}$$

Proof. This is simply a reformulation of the fact that the characteristic function (i.e., Fourier transform) of the random variable S_n is the nth power of the characteristic function of S_1. In detail: first, writing $S_n = \sum_{j=1}^n \xi_j$, we have

$$\begin{aligned} E e^{i\theta S_n} &= E \exp\{i\theta \sum_{j=1}^n \xi_j\} \\ &= \prod_{j=1}^n E \exp\{i\theta \xi_j\} \\ &= (E e^{i\theta \xi_1})^n \\ &= (\cos\theta)^n \end{aligned}$$

This uses the fact that the increments ξ_j are independent and identically distributed. Next, express the same characteristic function as a sum over possible values of S_n:

$$\begin{aligned} E e^{i\theta S_n} &= \sum_{m \in \mathbb{Z}} e^{i\theta m} P\{S_n = m\} \\ &= \sum_{m \in \mathbb{Z}} \frac{1}{2}(e^{+i\theta m} + e^{-i\theta m}) P\{S_n = m\} \\ &= \sum_{m \in \mathbb{Z}} \cos(m\theta) P\{S_n = m\} \\ &= \sum_{m \in \mathbb{Z}} T_m(\cos\theta) P\{S_n = m\}. \end{aligned}$$

Here, in the second equality, we have used the symmetry of the simple random walk, which implies that $P\{S_n = +m\} = P\{S_n = -m\}$. It now follows that for every real θ,

$$(\cos\theta)^n = \sum_{m\in\mathbb{Z}} T_m(\cos\theta)P\{S_n = m\}.$$

This implies that the polynomials on the left and right sides of equation (3.11) agree at infinitely many values of x, and so they must in fact be the same polynomial. □

Since (3.11) is a polynomial identity, one can substitute for the variable x any element of a module over the reals, in particular, we can use the identity for $x = L$, where $L : V \to V$ is a symmetric linear operator on a real vector space V:

$$L^n = \sum_{m\in\mathbb{Z}} T_m(L)P\{S_n = m\}. \tag{3.12}$$

Now $T_m(L)$ is nothing more than a linear combination of powers L^k of L; moreover, since the polynomial $T_m(x)$ has real coefficients, the linear operator $T_m(L)$ is symmetric. By (Sp2), if L has spectral decomposition (3.6), then for any $v \in V$

$$L^k v = \sum_{i\le D} \langle v, u_i\rangle \lambda_i^k u_i;$$

consequently,

$$T_m(L)v = \sum_{i\le D} \langle v, u_i\rangle T_m(\lambda_i) u_i.$$

This shows that $T_m(L)$ is not only symmetric, but has spectral decomposition with the same orthonormal basis of eigenvectors u_i, and corresponding eigenvalues $T_m(\lambda_i)$. Now recall that the eigenvalues λ_i of a symmetric linear operator L are real, and that $\|L\| = \max_{i\le D}|\lambda_i|$. Thus, if L has operator norm ≤ 1, then every eigenvalue $\lambda_i \in [-1, 1]$; therefore, by inequality (3.10), for every eigenvalue λ_i and every integer $m \ge 0$,

$$|T_m(\lambda_i)| \le 1.$$

Finally, using the relation between operator norm and spectrum in the reverse direction, we conclude that for every $m \ge 0$ the operator norm of $T_m(L)$ satisfies

$$\|T_m(L)\| \le 1. \tag{3.13}$$

3.4 Proof of the Carne-Varopoulos Inequality

In view of Proposition 3.9, it suffices to prove that

$$P^1\{X_n = x\} \le 2\|M\|^n \exp\{-d(1, x)^2/n\}. \tag{3.14}$$

This we will accomplish by exploiting the fact (cf. (M4) and Lemma 3.11) that for any $x \in \Gamma$ and any $k \ge n + 1$,

$$P^1\{X_n = x\} = \langle \delta_1, M^n \delta_x\rangle = \langle \delta_1, M_k^n \delta_x\rangle.$$

Recall that the matrix entries of M^n are the $n-$step transition probabilities $p_n(x,y)$. Since the random walk X_n is, by hypothesis, *nearest-neighbor*, the $n-$step transition probability $p_n(x,y)$ must be 0 if $n < d(x,y)$. Now the matrix entries of M_k^n are dominated by those of M^n (cf. the proof of Lemma 3.11), so it follows that if $n < d(x,y)$ then $\langle \delta_x, M_k^n \delta_y \rangle = 0$. Finally, since the Chebyshev polynomial $T_n(x)$ has degree n, the operator $T_n(M_k)$ contains only powers of M_k up to n; consequently,

$$d(1,x) > n \quad \Longrightarrow \quad \langle \delta_1, T_n(M_k)\delta_x \rangle = 0.$$

Let $\widehat{M}_k = M_k / \|M\|$. By Lemma 3.7, the operator $\widehat{M}_k$ has norm ≤ 1, and so by inequality (3.13), for each $n \geq 0$ the operator norm of $T_n(\widehat{M}_k)$ is bounded by 1. Now by the inversion formula for Chebyshev polynomials (cf. Proposition 3.14 and equation (3.12)), if S_n is the simple random walk on $\mathbb{Z}$ then

$$\begin{aligned}
\|M\|^{-n}\langle \delta_1, M_k^n \delta_x \rangle &= \langle \delta_1, \widehat{M}_k^n \delta_x \rangle \\
&= \sum_{m \in \mathbb{Z}} P\{S_n = m\} \langle \delta_1, T_{|m|}(\widehat{M}_k)\delta_x \rangle \\
&= \sum_{|m| \geq d(1,x)} P\{S_n = m\} \langle \delta_1, T_{|m|}(\widehat{M}_k)\delta_x \rangle \\
&\leq \sum_{|m| \geq d(1,x)} P\{S_n = m\} \|\delta_1\|_2 \, \|\delta_x\|_2 \, \|T_{|m|}(\widehat{M}_k)\| \\
&\leq \sum_{|m| \geq d(1,x)} P\{S_n = m\} \\
&= P\{|S_n| \geq d(1,x)\}.
\end{aligned}$$

Finally, HOEFFDING's inequality implies that

$$P\{|S_n| \geq d(1,x)\} \leq 2\exp\{-d(1,x)^2/n\},$$

and so (3.14) follows. □

4 Amenability, Nonamenability, and Return Probabilities

4.1 Amenable and Nonamenable Groups

Let $G = (V, \mathcal{E})$ be a connected, locally finite graph. (*Locally finite* means that every vertex $v \in V$ is incident to at most finitely many edges; equivalently, no vertex has infinitely many nearest neighbors.) For any finite set $F \subset V$, denote by ∂F the set of all vertices $y \notin F$ such that y is a nearest neighbor of some verrtex $x \in F$.

Definition 4.1. The *isoperimetric constant* $\iota(G)$ is defined by

$$\iota(G) := \inf_{F \text{ finite}} \frac{|\partial F|}{|F|}. \tag{4.1}$$

Exercise 4.2.

(A) Show that the isoperimetric constant of $\mathbb{Z}^d$ is 0.
(B) Find the isoperimetric constant of the homogeneous tree $\mathbb{T}_d$.
(C) Show that the isoperimetric constant of the $d-$dimensional lamplighter group $\mathcal{G}_1 = \mathbb{Z}^d \wr \oplus_{\mathbb{Z}^d}\mathbb{Z}_2$ is 0, for every $d \geq 1$.

Definition 4.3. A finitely generated group Γ is *amenable* (respectively, *nonamenable*) if its Cayley graph G_Γ (relative to any finite, symmetric set of generators) has isoperimetric constant $\iota(G_\Gamma) = 0$ (respectively, $\iota(G_\Gamma) > 0$).

It is not difficult to show that if the isoperimetric constant is 0 for *some* finite set of generators then it is 0 for *every* finite set of generators. Therefore, amenability (or nonamenability) is a property of the group, not of the particular Cayley graph chosen. For a nonamenable group, of course, the isoperimetric constants might be different for different Cayley graphs; but for every Cayley graph, the isoperimetric constant must be positive.

The distinction between amenable and nonamenable groups is of fundamental importance in representation theory and geometric group theory. It also plays a basic role in random walk theory; the following celebrated theorem of H. KESTEN explains why.

Theorem 4.4. *Every symmetric, nearest-neighbor random walk on a finitely generated amenable group has spectral radius* 1*, and every* irreducible*, symmetric, nearest-neighbor random walk on a finitely generated nonamenable group has spectral radius* < 1*.*

The hypothesis of irreducibility is obviously necessary, because if the step distribution of the random walk has full support in an amenable subgroup then the random walk would have spectral radius 1, by the first assertion of the theorem.[4]

Recall that the spectral radius and the entropy of a symmetric random walk obey the inequality $h \geq -\log \varrho$. Thus, we have the following immediate corollary.

Corollary 4.5. *Every symmetric, nearest-neighbor random walk on a finitely generated nonamenable group has positive Avez entropy and hence also positive speed.*

Recall also that if a finitely generated group Γ has sub-exponential growth (i.e., $\beta = 0$ in relation (2.3)) then every nearest-neighbor random walk on Γ has entropy 0. This implies

Corollary 4.6. *Every finitely generated group* Γ *with sub-exponential growth is amenable.*

Exercise 4.7. Give a direct proof that the modified lamplighter random walk (see section 1.6) on the lamplighter group $\mathbb{Z} \wr \mathbb{Z}_2^{\mathbb{Z}}$ has spectral radius 1. (Note: In Exercise 2.4 you showed that the lamplighter group has exponential growth, and in Exercise 4.2 you showed that the lamplighter group is amenable. Consequently, Kesten's theorem already implies that the spectral radius of the lamplighter random walk is 1. The point of this exercise is that this can be proved in a far more elementary fashion.)

[4] Consider, for instance, the free group $\mathbb{F}_2$ on two generators a, b, which is nonamenable. If the step distribution of a random walk is $\mu(a) = \mu(a^{-1}) = \frac{1}{2}$, then the random walk is just the simple random walk on the subgroup $\{a^n\}_{n\in\mathbb{Z}} \cong \mathbb{Z}$.

Exercise 4.8. Show that the (modified) lamplighter random walk in dimension 3 has positive speed, and therefore also positive entropy. This shows that it is possible for a symmetric, nearest-neighbor random walk on an amenable group to have positive speed.

Sections 4.2–4.5 will be devoted to the proof of the second assertion of Kesten's theorem, that any irreducible, symmetric, nearest-neighbor random walk on a nonamenable group Γ has spectral radius < 1. Section 4.6 will discuss a useful necessary condtion for a finitely generated group to be amenable.

4.2 Standing Assumptions

Assume in sections 4.2–4.5 that Γ is a finitely generated, nonamenable group.

We begin with several easy reductions of Kesten's theorem. Recall (cf. Proposition 3.9) that the spectral radius ϱ of a random walk coincides with the norm $\|M\|$ of its Markov operator. Thus, it will suffice to prove that for any irreducible, symmetric, nearest-neighbor random walk on a nonamenable group Γ the Markov operator satisfies

$$\|M\| < 1. \tag{4.2}$$

If a symmetric random walk on a finitely generated group Γ is both irreducible and nearest-neighbor, then the set of group elements h for which $p(1,h) > 0$ is a symmetric generating set A. Without loss of generality, we will assume that this generating set A is used to construct the Cayley graph G_Γ. (Recall that nonamenability does not depend on the choice of generating set.) Therefore, there is a positive constant ε such that for every nearest neighbor y of the group identity,

$$p(1,y) = \mu(y) \geq \varepsilon. \tag{4.3}$$

Lemma 4.9. *For every symmetric, nearest-neighbor random walk whose step distribution satisfies the condition (4.3), there is a positive constant κ such that for every finite set $F \subset \Gamma$,*

$$\sum_{x\in F}\sum_{y\in\partial F} p(x,y) \geq \kappa|F|. \tag{4.4}$$

Proof. Take $\kappa = \varepsilon\iota(G_\Gamma)$; then (4.4) follows directly from (4.3) and the definition of the isoperimetric constant. $\square$

4.3 The Sobolev Inequality

Definition 4.10. For any function $f : \Gamma \to \mathbb{R}$, define the *Sobolev norm* $S(f)$ by

$$S(f) := \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma} |f(x) - f(y)|p(x,y).$$

Proposition 4.11. *For any symmetric, nearest-neighbor random walk on a nonamenable group Γ for which condition (4.3) holds, there exists a constant $\kappa > 0$ such that for every function $f \in L^1(\Gamma)$,*

$$S(f) \geq \kappa \|f\|_1. \tag{4.5}$$

Proof. Without loss of generality, assume that $f \geq 0$. For any $t > 0$, let F_t be the set of all $y \in \Gamma$ for which $f(y) > t$. Clearly, if $f \in L^1(\Gamma)$ then each F_t is a finite set. Consequently, by Lemma 4.9,

$$\begin{aligned} S(f) &= \sum_{x,y\,:\,f(x)<f(y)} (f(y) - f(x))p(x,y) \\ &= \int_0^\infty \sum_{x,y} \mathbf{1}\{f(x) < t < f(y)\} p(y,x)\, dt \\ &= \int_0^\infty \sum_{y \in F_t} \sum_{x \in \partial F_t} p(y,x)\, dt \\ &\geq \kappa \int_0^\infty |F_t|\, dt = \kappa \|f\|_1. \end{aligned}$$

□

4.4 Dirichlet Form

Recall that our objective is to prove inequality (4.2), which asserts that the norm of the Markov operator M is strictly less than 1. Because the random walk is symmetric, so is M, and because the matrix entries $p(x,y)$ of M are nonnegative, $Mf \geq 0$ for any nonnegative function f. This allows the following characterization of the norm $\|M\|$.

Lemma 4.12. $1 - \|M\| = \inf_{\|f\|_2=1} \langle f, (I-M)f \rangle$.

Proof Sketch. By Lemma 3.7, $\|M\|$ is the supremum of the sequence $\|M_k\|$, where $M_k = P_k M P_k$ is the restriction of M to the ball B_k. Moreover, for any $f \in L^2(\Gamma)$, the function Mf is the L^2-limit of the sequence $M_k f$. Hence, it suffices to prove that for each $k \geq 1$,

$$1 - \|M_k\| = \inf_{\|f\|_2=1} \langle f, (I - M_k) f \rangle,$$

where the inf is now over all f with support in the ball B_k. By the Spectral Theorem for finite-dimensional symmetric operators, $\|M_k\|$ is the maximum magnitude of its eigenvalues. Furthermore, since M_k has nonnegative matrix entries, the max must occur at the largest *nonnegative* eigenvalue λ_*, and the corresponding normalized eigenfunction f_* must be nonnegative. Therefore,

$$1 - \|M_k\| = 1 - \lambda_* = \langle f_*, f_* \rangle - \langle f_*, M_k f_* \rangle$$

Finally, since $\langle f, M_k f \rangle$ is maximized at $f = f_*$ among all f with norm 1, the inner product $\langle f, (I - M_k) f \rangle$ is minimized at $f = f_*$. □

Definition 4.13. The *Dirichlet form* on $L^2(\Gamma)$ is the quadratic form

$$\mathcal{D}(f,f) = \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}(f(y)-f(x))^2 p(x,y). \tag{4.6}$$

Proposition 4.14. *For any $f \in L^2(\Gamma)$,*

$$\mathcal{D}(f,f) = \langle f, (I-M)f\rangle. \tag{4.7}$$

Proof. Expand the square in the Dirichlet form and use the fact that $\sum_y p(x,y) = \sum_y p(x,y) = 1$ for all group elements x, y to obtain

$$\begin{aligned}
\mathcal{D}(f,f) &= \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}\left(f(x)^2 + f(y)^2 - 2f(x)f(y)\right)p(x,y) \\
&= \sum_x f(x)^2 - \sum_x\sum_y f(x)f(y)p(x,y) \\
&= \langle f, f\rangle - \langle f, Mf\rangle = \langle f, (I-M)f\rangle.
\end{aligned}$$

□

4.5 Spectral Gap: Proof of Kesten's Theorem

To show that $\|M\| < 1$ it suffices, by Lemma 4.12, to show that there exists $\varepsilon > 0$ such that for every function $f \in L^2(\Gamma)$ with norm $\|f\|_2 = 1$,

$$\langle f, (I - Mf)\rangle \geq \varepsilon.$$

Proposition 4.14 implies that this is equivalent to proving

$$\mathcal{D}(f,f) \geq \varepsilon \quad \text{provided } \|f\|_2 = 1. \tag{4.8}$$

To accomplish this, we will use the Cauchy-Schwartz inequality to bound the Dirichlet form $\mathcal{D}(f,f)$ from below by a multiple of the Sobolev norm $S(f^2)$ of f^2, and then use the Sobolev inequality to bound the Sobolev norm by the L^2-norm of f. Here's how it works: let f be any function with $\|f\|_2 = 1$; then with κ as in the Sobolev inequality,

$$\begin{aligned}
\kappa\|f\|_2^2 = \kappa\|f^2\|_1 &\leq S(f^2) \\
&= \frac{1}{2}\sum_x\sum_y |f(y)^2 - f(x)^2| p(x,y) \\
&= \frac{1}{2}\sum_x\sum_y |f(y) - f(x)||f(y) + f(x)| p(x,y) \\
&\leq \left(\frac{1}{2}\sum_x\sum_y (f(y)-f(x))^2 p(x,y)\right)^{1/2}\left(\frac{1}{2}\sum_x\sum_y (f(y)+f(x))^2 p(x,y)\right)^{1/2} \\
&\leq \mathcal{D}(f,f)^{1/2}\left(\frac{1}{2}\sum_x\sum_y 4(f(y)^2 + f(x)^2)p(x,y)\right)^{1/2} = 4\mathcal{D}(f,f)^{1/2}\|f\|_2^2.
\end{aligned}$$

Squaring both sides and using the assumption that $\|f\|_2 = 1$ now yields

$$\mathcal{D}(f, f) \geq \kappa/4,$$

which proves (4.8). □

4.6 A Necessary Condition for Amenability

Definition 4.15. Let Γ be a finitely generated group. A $\Gamma-$*action* on a compact metric space $\mathcal{Y}$ is a group homomorphism $\Phi : \Gamma \to \text{Homeo}(\mathcal{Y})$ from Γ to the group of homeomorphisms of $\mathcal{Y}$.

The homeomorphism $\Phi(g)$ assigned to a given group element $g \in \Gamma$ is usually abbreviated as g, and for a given element $y \in \mathcal{Y}$ the image $\Phi(g)(y)$ is abbreviated gy. A group action Φ on $\mathcal{Y}$ induces an action on the space of Borel probability measures on $\mathcal{Y}$, as follows: for a Borel probability measure ν, a group element g, and a continuous function $f : \mathcal{Y} \to \mathbb{R}$,

$$\int_{y \in \mathcal{Y}} f(y)\, (g * \nu)(dy) := \int_{y \in \mathcal{Y}} f(gy)\, \nu(dy). \tag{4.9}$$

Definition 4.16. Let Φ be a $\Gamma-$action on $\mathcal{Y}$. A Borel probability measure ν on $\mathcal{Y}$ is said to be $\Gamma-$*invariant* if for every $g \in \Gamma$,

$$g * \nu = \nu. \tag{4.10}$$

Proposition 4.17. *An infinite, finitely generated group Γ is amenable if and only if for any $\Gamma-$action on any compact metric space $\mathcal{Y}$ there is a $\Gamma-$invariant Borel probability measure.*

This provides a useful tool for showing that a group is *non*-amenable: one only need find a single $\Gamma-$action for which there is no $\Gamma-$invariant Borel probability measure.

Exercise 4.18. (For those of you who know the basics of Fuchsian groups.) Show that every co-compact Fuchsian group is nonamenable. HINT: A Fuchsian group Γ acts on the circle at infinity by linear fractional transformations. Show that this action has no invariant probability measure. To do this, you will need the following fact: every co-compact Fuchsian group has *hyperbolic* elements (linear fractional transformations with two fixed point on the circle at infinity), and the set of fixed point pairs of these hyperbolic elements is dense in the circle at infinity.

Proof of Proposition 4.17$\Longrightarrow$. We will prove only the forward implication, that amenability implies the existence of invariant probability measures. Suppose that Γ is amenable, and let $\Phi : \Gamma \to \text{Homeo}(\mathcal{Y})$ be a $\Gamma-$action on a compact metric space $\mathcal{Y}$. Because Γ is amenable, there exist finite sets $F_n \subset \Gamma$ such that

$$\lim_{n\to\infty} \frac{|\partial F_n|}{|F_n|} = 0. \tag{4.11}$$

Claim: Without loss of generality, the sets F_n can be chosen in such a way that $A \subset F_1 \subset F_2 \subset F_3 \subset \cdots$, where A is the generating set.

Proof of the Claim. Let F_n be a sequence of finite sets satisfying (4.11); then $|F_n| \to \infty$, and so (by taking a subsequence if necessary) we may assume that $|F_m| \geq 2^{2^m}|F_{m-1}|$. Let

$$G_m = A \cup F_1 \cup F_2 \cup \cdots \cup F_m.$$

The sets G_n are nested and contain contain the generating set A. Moreover, because the cardinality of F_m dominates the cardinality of G_{m-1}, the sequence G_n satisfies (4.11). □

Now let ν be any Borel probability measure on $\mathcal{Y}$. For each $n = 1, 2, \cdots$, define a Borel probability measure ν_n by averaging the translates of ν by elements of the set F_n, that is,

$$\nu_n = \frac{1}{|F_n|} \sum_{g \in F_n} g * \nu.$$

By the Banach-Alaoglu theorem, the sequence ν_n contains a weak$-*$ convergent subsequence ν_m; and by the Riesz representation theorem, the limit ν_* of this weak$-*$ convergent subsequence is a Borel probability measure on $\mathcal{Y}$. Furthermore, since any group element $g \in \Gamma$ acts on $\mathcal{Y}$ as a homeomorphism, it is also the case that the sequence $g * \nu_n$ converges weak$-*$ to $g * \nu_*$.

We claim that ν_* is $\Gamma-$invariant. To see this, observe that for any continuous function $f : \mathcal{Y} \to \mathbb{R}$ and any group generator $a \in A$,

$$\left| \int f \, d\nu_n - \int f \, d(a * \nu_n) \right| \leq \frac{2\|f\|_\infty |\partial F_n|}{|F_n|} \longrightarrow 0;$$

consequently, since $\nu_n \to \nu_*$ weakly,

$$\int f \, d\nu_* = \int f \, d(a * \nu_*).$$

Since f is arbitrary, this proves that $\nu_* = a * \nu_*$ for every group generator a, and it follows trivially that $\nu_* = g * \nu_*$ for every $g \in \Gamma$. □

5 Harmonic Functions

Assume throughout this section that $X_n = X_0\xi_1\xi_2 \cdots \xi_n$ is an irreducible, nearest-neighbor random walk on a finitely generated group Γ with symmetric generating set A, with transition probabilities $p(x, y)$. For any finite set $U \subset \Gamma$, denote by ∂U the (finite) set of all group elements $y \notin U$ such that y is a nearest neighbor of some $x \in U$, that is, such that $x^{-1}y$ is an element of the generating set A.

5.1 Harmonic Functions and the Dirichlet Problem

Definition 5.1. [5] A function $h : U \cup \partial U \to \mathbb{R}$ is called *harmonic* at a point $x \in U$ if

$$h(x) = E^x h(X_1) = \sum_{y \in \Gamma} p(x, y)h(y), \tag{5.1}$$

[5] In section 5, the Avez entropy will play no role, so there should be no ambiguity in the use of the letter h to denote a harmonic function on a subset of Γ.

and h is said to be *harmonic in* U if it is harmonic at every $x \in U$. A function $h : \Gamma \to \mathbb{R}$ is *harmonic* if it is harmonic at every $x \in \Gamma$, that is, if equation (5.1) is valid for every x.

For any function $h : \Gamma \to \mathbb{R}$ that is harmonic on the entire group Γ, the defining formula (5.1) can be iterated: thus, for any $n \geq 0$,

$$h(x) = E^x h(X_n) = \sum_{y \in \Gamma} p_n(x, y) h(y). \tag{5.2}$$

A routine calculation shows that if h is harmonic then the sequence $h(X_n)$ is a *martingale*, that is, for every cylinder set

$$C(x_0, x_1, \cdots, x_n) = \cap_{i=0}^n \{X_i = x_i\}, \tag{5.3}$$

the martingale identity

$$E^x(h(X_{n+1}) \,|\, C(x_0, x_1, \cdots, x_n)) := \frac{E^x h(X_{n+1}) \mathbf{1}_{C(x_0, x_1, \cdots, x_n)}}{P^x(C(x_0, x_1, \cdots, x_n))} = h(x_n) \tag{5.4}$$

holds. We will avoid the use of martingale theory in the following development. The reader who is familiar with the basics of this theory will recognize that the Poisson formula (5.6) below, on which we will base our exposition, is a direct consequence of Doob's "Optional Sampling Formula" for martingales.

Obviously, linear combinations of harmonic functions are harmonic, and every constant function is harmonic. Since constant functions do not tell us anything about random walks, our interest will be in non-constant harmonic functions.

Example 5.2. The function $h : \mathbb{Z} \to \mathbb{R}$ defined by $h(x) = x$ is harmonic for the simple random walk on $\mathbb{Z}$. The function $h(x) = ((1-p)/p)^x$ is harmonic for the asymmetric random walk on $\mathbb{Z}$ with step distribution $P\{X_1 = +1\} = p = 1 - P\{X_1 = -1\}$.

Exercise 5.3. Prove the *Maximum Principle:* If $h : U \cup \partial U \to \mathbb{R}$ is harmonic in a finite set U then h does not have a local max or min in U. More precisely, there is no $x \in U$ such that

$$h(x) > h(y) \quad \text{for all neighbors } y \text{ of } x.$$

Exercise 5.4. Prove the *Uniqueness Theorem:* Given a finite set U and a function $f : \partial U \to \mathbb{R}$ defined on its boundary, there is at most one harmonic function h in U such that

$$h(x) = f(x) \quad \text{for all } \; x \in \partial U.$$

Definition 5.5. For any finite set $U \subset \Gamma$ with boundary ∂U, define the *first exit time* from the region U to be the random variable

$$\tau_U := \min\{n \geq 0 : X_n \notin U\}. \tag{5.5}$$

When there is no danger of ambiguity, we will sometimes abbreviate τ_U by τ.

Proposition 5.6. *Let $U \subset \Gamma$ be a finite set with boundary ∂U, and let $f : \partial U \to \mathbb{R}$ be a specified function on the boundary. Then the function*

$$h(x) = E^x f(X_{\tau_U}) \tag{5.6}$$

is harmonic in U and satisfies $h = f$ on the boundary ∂U.

Proof. If the starting point $X_0 = x$ of the random walk is an element of ∂U then the first exit time must be $\tau_U = 0$, which implies $E^x f(X_\tau) = E^x f(x) = f(x)$. Thus, $h = f$ on ∂U.

If the starting point $X_0 = x$ of the random walk is an element of U, then the first exit time must satisfy $\tau_U \geq 1$. Thus, we can compute the expectation $E^x f(X_\tau)$ by "conditioning on the first step":

$$E^x f(X_\tau) = \sum_{y \sim x} E^x f(X_\tau) \mathbf{1}\{X_1 = y\} = \sum_{y \sim x} p(x,y) E^x (f(X_\tau) \,|\, X_1 = y).$$

On the event $X_1 = y$, the first exit time must be $\tau = 1 + \tau'$, where τ' is the time of first exit from U for the random walk $X'_n = y\xi_2\xi_3 \cdots \xi_{n+1}$. Since the increments ξ_n of the random walk are independent and identically distributed, the distribution of the exit site $X'_{\tau'}$ is the same as that of the random walk $X_n = y\xi_1\xi_2 \cdots \xi_n$ when started at y; consequently,

$$E^x (f(X_\tau) \,|\, X_1 = y) = E^y f(X_\tau) = h(y).$$

This shows that h satisfies $h(x) = \sum_y p(x,y)h(y)$, and therefore proves that h is harmonic in U. □

Exercise 5.7. Assume that the random walk is *symmetric*. Prove the *Dirichlet Principle:* Among all functions $u : U \cup \partial U \to \mathbb{R}$ that satisfy the boundary condition $u(x) = f(x)$ for $x \in \partial U$, the harmonic function $h(x) = E^x f(X_\tau)$ is the one that uniquely minimizes the *Dirichlet form*

$$\mathcal{D}_U(u,u) := \frac{1}{2} \sum \sum_{x,y} (u(y) - u(x))^2 p(x,y)$$

where the sum is over all nearest-neighbor pairs (x,y) such that at least one of the vertices x, y is an element of U.

5.2 Convergence Along Random Walk Paths

Together with the Uniqueness Theorem, formula (5.6), the so-called *Poisson formula*, shows that any harmonic function in a region U is uniquely determined by its values on the boundary ∂U. Our next objective is to find an analogous principle for bounded harmonic functions on the entire group Γ. Obviously, there is no "finite" analogue of the Poisson formula, because there is no first exit time for Γ. However, the following theorem shows that there is, nevertheless, a limiting analogue of the Poisson formula.

Assumption 5.8. *Assume for the remainder of section 5.2 that $h : \Gamma \to [0, \infty)$ is a nonnegative, harmonic function on the entire group Γ.*

Theorem 5.9. *For any initial point $x \in \Gamma$, with P^x−probability 1, the limit $Y := \lim_{n\to\infty} h(X_n)$ exists and is finite. Furthermore, if the function h is bounded then*

$$h(x) = E^x Y. \tag{5.7}$$

The first assertion of the theorem is a direct consequence of the *martingale convergence theorem*, which implies that every nonnegative martingale converges almost surely to a finite limit. The second asertion of the theorem, that for bounded h the integral formula (5.7) holds, then follows from the bounded convergence theorem, because harmonicity implies that $h(x) = E^x f(X_n)$ for every $n \geq 0$. Because the martingale convergence theorem relies on some rather heavy machinery, we will devote the remainder of this section 5.2 to an elementary proof[6] of Theorem 5.9.

Proposition 5.10. *(Maximal Inequality) For any $\alpha > 0$ and any $x \in \Gamma$,*

$$P^x\{\sup_{n\geq 0} h(X_n) \geq \alpha\} \leq h(x)/\alpha. \tag{5.8}$$

Proof. The inequality is trivial unless $\alpha > h(x)$, so let's assume this is the case. Define $G = G(\alpha)$ to be the set of all group elements y such that $h(y) \geq \alpha$, and let T be the first time n that $X_n \in G$, or $T = \infty$ if there is no such n. Our goal is to show that

$$P^x\{T < \infty\} \leq h(x)/\alpha. \tag{5.9}$$

Let $m \geq 1$ be large enough that the initial point x is contained in the ball B_m of radius m centered at the group identity 1. Define $U_m = B_m \setminus G$ to be the set of points y in the ball B_m such that $h(y) < \alpha$. By the Poisson formula (Proposition 5.6),

$$h(x) = E^x h(X_{\tau_m})$$

where τ_m is the first exit time from the set U_m. At the exit time τ_m, the random walk will either have reached a point of G or it will have reached the boundary of the ball B_m, so $\tau_m = T \wedge \nu_m$, where ν_m is the first exit time from B_m. If $T = \tau_m \leq \nu_m$ then $h(X_{\tau_m}) \geq \alpha$; but if $\tau_m = \nu_m < T$, then $h(X_{\tau_m})$ will take some value between 0 and α. Consequently, by the Poisson formula,

$$\alpha P^x\{T = \tau_m\} = \alpha P^x\{T \leq \nu_m\} \leq E^x h(\tau_m) = h(x).$$

As $m \to \infty$, the random variables $\nu_m \to \infty$; this implies that the event $\{T < \infty\}$ is the increasing limit of the events $\{T \leq \nu_m\}$. Therefore, monotone convergence implies the inequality (5.9). □

The maximal inequality (5.8) clearly implies that with, P^x−probability 1, the sequence $h(X_n)$ has finite supremum. Hence, the sequence $h(X_n)$ will fail to converge to a finite limit only if there is a nonempty interval $(\alpha, \beta) \subset (0, \infty)$ with rational endpoints such that

$$\liminf_{n\to\infty} h(X_n) \leq \alpha \quad \text{and}$$
$$\limsup_{n\to\infty} h(X_n) \geq \beta.$$

[6] This is, of course, modeled on the standard approach to the martingale convergence theorem.

This will occur only if the sequence makes infinitely many crossings from $[0, \alpha]$ to $[\beta, \infty)$. The following result shows that this cannot happen with positive probability.

Proposition 5.11. *(Upcrossings Inequality) Fix $0 \leq \alpha < \beta < \infty$, and let N be the number of* upcrossings, *that is, the number of times that the sequence $h(X_n)$ crosses from below α to above β. Then for every $m = 0, 1, 2, \cdots$ and every $x \in \Gamma$,*

$$P^x\{N \geq m\} \leq \left(\frac{\alpha}{\beta}\right)^m. \tag{5.10}$$

Proof. This is by induction on m. The inequality is trivial if $m = 0$, so it suffices to show that if it is true for m then it is true for $m + 1$.

Obviously, the event $N \geq m + 1$ can only occur if $N \geq m$ and if the sequence $h(X_n)$ returns to the interval $[0, \alpha]$ following the completion of its mth upcrossing. Partition this latter event into elementary cylinder events $C(x_0, x_1, \cdots, x_n)$ in such a way that for each such cylinder in the partition, the sequence $h(X_j)$ will have completed m upcrossings and returned to the interval $[0, \alpha]$ for the first time following these m upcrossings at time n. In order that the sequence then makes at least one more upcrossing, it must be the case that the "post-n" random walk

$$x_n, x_n\xi_{n+1}, x_n\xi_{n+1}\xi_{n+2}, \cdots$$

will eventually reach a point y where $h(y) \geq \beta$. Since the steps ξ_{n+i} of this random walk are independent of the cylinder event $C(x_0, \cdots, x_n)$, and since $h(x_n) \leq \alpha$, it now follows by the maximal inequality that the (conditional) probability of this (given the cylinder event) is no larger than α/β, i.e.,

$$P^x(N \geq m + 1 \,|\, C(x_0, x_1, \cdots, x_n) \leq \alpha/\beta.$$

Using the multiplication rule for conditional probability and then summing over all the cylinder events that make up $\{N \geq m\}$, we obtain

$$P^x\{N \geq m + 1\} \leq (\alpha/\beta) P^x\{N \geq m\}.$$

□

Proof of Theorem 5.9. As we have already remarked, the equality (5.7) will follow by the bounded convergence theorem once we establish the almost sure convergence of the sequence $h(X_n)$. The maximal inequality implies that $\sup_{n \geq 0} h(X_n)$ is finite with probability 1, and the upcrossings inequality implies that for any two real numbers $\alpha < \beta$, the number of crossings of the interval (α, β) is almost surely finite. Because the set of *rational* pairs $\alpha < \beta$ is countable, and because the union of countably many sets of probability 0 is a set of probability 0, it follows that with probability 1 there is no rational pair $\alpha < \beta$ such that the sequence $h(X_n)$ makes infinitely many upcrossings of the interval (α, β). This implies that with probability 1,

$$\liminf_{n \to \infty} h(X_n) = \limsup_{n \to \infty} h(X_n),$$

so the sequence $h(X_n)$ must converge to a finite limit. □

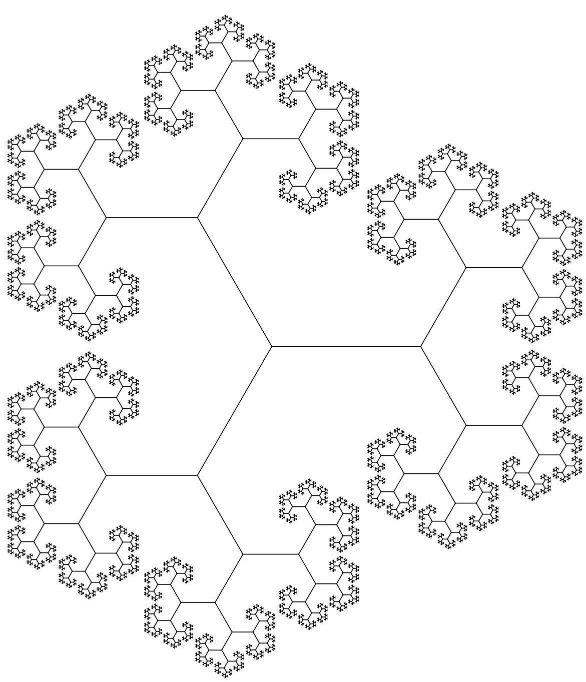
5.3 Example: Random Walk on $\mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$

A transient random walk eventually leaves every finite subset of its state space Γ. Where does it go? The convergence theorem 5.9 suggests that the answer (to the extent that a satisfactory answer can be given) has something to do with harmonic functions. The goal of this section is to study in detail an interesting, concrete example where the connection between the long-time behavior of a transient random walk and the space of harmonic functions is relatively transparent.

Recall that the group $\Gamma = \mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$ has as its Cayley graph the homogeneous tree $\mathbb{T}_3$ of degree 3, which in this section we will abbreviate $\mathbb{T}$. Elements of Γ are finite reduced words in the letters a, b, c; these are represented by vertices of the tree. Define $\partial\mathbb{T}$, the space of *ends* of the tree, to be the set of all *infinite* reduced words

$$\omega = \alpha_1\alpha_2\cdots .$$

Let d be the metric on $\mathbb{T} \cup \partial\mathbb{T}$ defined by $d(\omega, \omega') = 2^{-n}$, where $n \geq 0$ is the maximal integer such that the words ω and ω' (whether finite or infinite) agree in their first n coordinates. The topology induced by this metric is the same as the induced Euclidean topology on the tree by the embedding shown here:



In the following exercises, let X_n be the nearest-neighbor random walk on Γ with step distribution $P\{\xi_i = j\} = \mu(j) > 0$, where $j \in \{a, b, c\}$. Define the *hitting probability function*

u by

$$\begin{aligned} u(x) &= P^1\{X_n = x \text{ for some } n \geq 0\} \\ &= P^x\{X_n = 1 \text{ for some } n \geq 0\} \end{aligned}$$

Exercise 5.12. Prove that

$$\lim_{n\to\infty} X_n = X_\infty \in \partial\mathbb{T}$$

exists with P^x-probability one for any initial point $x \in \Gamma$. (Here the convergence is with respect to the metric d.) The distribution (under P^x) of the exit point X_∞ is, sensibly enough, called the *exit distribution*. Denote this by ν_x.

Exercise 5.13. Show that

(A) If x has word representation $x = a_1 a_2 \cdots a_m$ then $u(x) = \prod_{i=1}^m u(a_i)$.

(B) Show that for each generator $i = a, b, c$,

$$u(i) = \mu(i) + \sum_{j\neq i} \mu(j)u(j)u(i).$$

Exercise 5.14. For any finite reduced word $w = a_1 a_2 \cdots a_m$, define $\Sigma(w)$ to be the subset of $\partial\mathbb{T}$ consisting of all infinite reduced words whose first m letters are $a_1 a_2 \cdots a_m$.

(A) Show that

$$P^1\{X_\infty \in \Sigma(w)\} = \nu_1(\Sigma(w)) = \frac{u(w)}{1 + u(a_m)}.$$

(B) Conclude that

$$\sum_{i=a,b,c} \frac{u(i)}{1 + u(i)} = 1.$$

(C) Let X_∞ have reduced word representation $X_\infty = A_1 A_2 A_3 \cdots$. Show that under P^1 the sequence $A_1, A_2, A_3, \cdots$ is a *Markov chain* on the set $\{a, b, c\}$. What are the transition probabilities and intial distribution?

Exercise 5.15. Let x_n be a sequence of group elements that converge (in the metric d) to a point $\omega \in \partial\mathbb{T}$ of the space of ends. Prove that the exit measures ν_{x_n} converge weakly to the unit point mass at ω, that is, show that for any open set $U \subset \partial\mathbb{T}$ containing ω,

$$\lim_{n\to\infty} \nu_{x_n}(U) = 1.$$

Exercise 5.16. Let $f : \partial\mathbb{T} \to \mathbb{R}$ be any bounded, Borel measurable function. Define $h : \Gamma \to \mathbb{R}$ by

$$h(x) := E^x f(X_\infty) = \int_{\omega\in\partial\mathbb{T}} f(\omega)\, d\nu_x(\omega).$$

(A) Show that h is harmonic on Γ.

(B) Show that if f is continuous then $\lim_{n\to\infty} h(X_n) = f(X_\infty)$ almost surely (for any P^x).

Exercise 5.17. This exercise outlines a proof of the converse to Exercise 5.16. Let $h : \Gamma \to \mathbb{R}$ be any bounded, harmonic function. For each $n \geq 1$ define a function $f_n : \Gamma \cup \partial\mathbb{T} \to$ by

$$\begin{aligned} f_n(a_1 a_2 \cdots a_m) &= h(a_1 a_2 \cdots a_m) \quad \text{if } m \leq n; \\ f_n(a_1 a_2 \cdots a_m) &= h(a_1 a_2 \cdots a_n) \quad \text{if } m > n; \\ f_n(a_1 a_2 \cdots) &= h(a_1 a_2 \cdots a_n). \end{aligned}$$

(A) Use Theorem 5.9 to show that for any $x \in \Gamma$,

$$\nu_x\{\omega : \lim_{n\to\infty} f_n(\omega) := f(\omega) \text{ exists}\} = 1.$$

(B) Let G be the set of all $\omega \in \partial\mathbb{T}$ such that $\lim f_n(\omega) = f(\omega)$ exists. For $\omega \in \Gamma \setminus G$, define $f(\omega) = 0$. Prove that for every $x \in \Gamma$,

$$h(x) = E^x f(X_\infty) = \int_{\omega\in\partial\mathbb{T}} f(\omega)\, d\nu_x(\omega).$$

Exercise 5.18. For any $\omega = a_1 a_2 a_3 \cdots \in \partial\mathbb{T}$ let $w_n = a_1 a_2 \cdots a_n$ be the sequence of group elements along the geodesic ray from 1 to ω. For any $x = b_1 b_2 \cdots b_m \in \Gamma$, let $n(x,\omega) \leq m$ be the maximal nonnegative integer such that the words $a_1 a_2 \cdots$ and $b_1 b_2 \cdots b_m$ agree in the first n coordinates.

(A) Show that the sequence $u(x^{-1}w_n)/u(w_n)$ stabilizes for $n \geq n(x,\omega)$.
(B) Show that for any finite word w that has $w_{n(x,\omega)}$ as a prefix (i.e., the group element w lies on the geodesic ray from $w_{n(x,\omega)}$ to ω),

$$\frac{\nu_x(\Sigma(w))}{\nu_1(\Sigma(w))} = \frac{u(x^{-1}w_{n(x,\omega)})}{u(w_{n(x,\omega)})}.$$

(C) Conclude from (B) that the measures ν_x and ν_1 are mutually absolutely continuous, and that the Radon-Nikodym derivative (likelihood ratio) $d\nu_x/d\nu_1$ is given by

$$\frac{d\nu_x}{d\nu_1}(\omega) = \frac{u(x^{-1}x_{n(x,\omega))})}{u(x_{n(x,\omega))})} = \lim_{n\to\infty} \frac{u(x^{-1}w_n)}{u(w_n)} := K(x,\omega).$$

NOTE: The Radon-Nikodym derivative is by definition the unique Borel measurable function on $\partial\mathbb{T}$ such that for every Borel set $F \subset \partial\mathbb{T}$,

$$\nu_x(F) = \int_{\omega\in F} \frac{d\nu_x}{d\nu_1}(\omega)\, d\nu_1(\omega)$$

(D) Show that for each $\omega \in \partial\mathbb{T}$ the function $x \mapsto K(x,\omega)$ is harmonic.
(E) Show that for each $x \in \Gamma$ the function $\omega \mapsto K(x,\omega)$ is (Hölder) continuous on $\partial\mathbb{T}$.

The function $K(x,\omega)$ defined in Exercise 5.18 is called the *Martin kernel* of the random walk. It extends to a Hölder continuous function $K : \mathbb{T} \times (\mathbb{T} \cup \partial\mathbb{T}) \to (0,\infty)$ by setting

$$K(x,y) = \frac{u(x^{-1}y)}{u(y)} = \frac{P^x\{X_n = y \text{ for some } n\}}{P^1\{X_n = y \text{ for some } n\}}.$$

Since linear combinations (even infinite ones) of harmonic functions are harmonic, it follows from (E) that for any finite probability measure λ on $\partial\mathbb{T}$, the integral

$$h(x) := \int_{\omega\in\partial\mathbb{T}} K(x,\omega)\, d\lambda(\omega) \tag{5.11}$$

is wel-defined and finite, and from (D) that h is harmonic, with value $h(1) = 1$ at the identity. This is called the *Martin representation* of the harmonic function. It can be shown (cf., for example, E. B. DYNKIN, Boundary Theory of Markov Processes (The Discrete Case)) that every nonnegative harmonic function has a Martin representation, and that the representation is unique.

5.4 Harmonic Functions and the Invariant $\sigma-$Algebra

Let's return to the study of arbitrary nearest-neighbor random walks $X_n = X_0\xi_1\xi_2\cdots\xi_n$ on arbitrary finitely generated groups Γ. We have shown that for any nonnegative harmonic function h the sequence $h(X_n)$ converges almost surely to a finite limit Y. Clearly, this limit random variable Y is a function of the sequence X_n, but it doesn't depend on the entire sequence: in fact, for any $m \geq 1$ the value of Y is determined by the truncated sequence $X_{m+1}X_{m+2},\cdots$. The essential information needed to determine limits of functions along the sequence is contained in the *invariant $\sigma-$algebra*.

Recall that a $\sigma-$algebra on a set $\mathcal{Y}$ is a collection of subsets that contains the empty set $\emptyset$ and is closed under complements and countable unions. If $\mathcal{Y}$ is a topological space, the *Borel $\sigma-$algebra* $\mathcal{B} = \mathcal{B}_{\mathcal{Y}}$ is the smallest $\sigma-$algebra containing all the open sets. If $\mathcal{Y} = \Gamma^\infty$ is the set of all sequences with entries in Γ, the Borel $\sigma-$algebra $\mathcal{B}_\infty$ is the smallest $\sigma-$algebra containing all sets of the form $\{Y_n = y\}$, where $Y_n : \mathcal{Y} \to \Gamma$ is the nth coordinate projection, n is any nonnegative integer, and y is any element of Γ.

Definition 5.19. The *invariant $\sigma-$algebra* $\mathcal{I}$ on Γ^∞ is the set of all Borel sets $F \in \mathcal{B}_\infty$ whose indicators are invariant under the shift, that is,

$$(y_0, y_1, y_2, \cdots) \in F \quad\iff\quad (y_1, y_2, y_3, \cdots) \in F. \tag{5.12}$$

By definition (see the Technical Note in section 1.2), the random variables X_n that constitute a random walk are measurable mappings $X_n : \Omega \to \Gamma$ defined on some measurable space $(\Omega, \mathcal{F})$. These fit together to give a measurable mapping $\mathbf{X} : \Omega \to \Gamma^\infty$ by

$$\mathbf{X} = (X_0, X_1, X_2, \cdots).$$

Let P^x denote the probability measure on $(\Omega, \mathcal{F})$ under which the sequence $\mathbf{X}$ is a random walk with initial point $X_0 = x$ and step distribution μ. The mapping $\mathbf{X} : \Omega \to \Gamma^\infty$ naturally induces a Borel probability measure on $(\Gamma^\infty, \mathcal{B}_\infty)$, called the *distribution* of the sequence $\mathbf{X}$ under P^x, by

$$\nu_x(F) = P^x\{\mathbf{X} \in F\}. \tag{5.13}$$

The restriction of ν_x to the invariant $\sigma-$algebra $\mathcal{I}$ is called the *exit measure* of the random walk. The triple $(\Gamma^\infty, \mathcal{I}, \nu_1)$ is called the *Poisson boundary* of the random walk.

Proposition 5.20. *The exit measures are mutually absolutely continuous, i.e., for any $x, y \in \Gamma$ and any $F \in \mathcal{I}$,*

$$\nu_x(F) = 0 \quad \Longleftrightarrow \quad \nu_y(F) = 0.$$

Proof. This is an easy consequence of the irreducibility of the random walk. Suppose that $\nu_x(F) = 0$. Let y be any nearest-neighbor of x such that $P^x\{X_1 = y\} > 0$; then

$$\begin{aligned}
\nu_x(F) &= P^x\{(X_0, X_1, X_2, \cdots) \in F\} \\
&= P^x\{(X_1, X_2, X_3, \cdots) \in F\} \\
&\geq P^x\{X_1 = y \ \text{ and } \ (X_1, X_2, X_3, \cdots) \in F\} \\
&= P^x\{X_1 = y\}\nu_y(F),
\end{aligned}$$

and so $\nu_y(F) = 0$. Irreducibility now implies that $\nu_z(F) = 0$ for every $z \in \Gamma$. $\square$

Clearly, if $h : \Gamma \to \mathcal{R}$ is Borel measurable, then the random variable $\limsup h(X_n)$ is measurable with respect to the $\sigma-$algebra $\mathbf{X}^{-1}(\mathcal{I})$, because the value of the limsup is the same for any sequence and its right-shift. Therefore, the random variables Y in Theorem 5.9 are (shift-)invariant, that is, they are measurable with respect to $\mathbf{X}^{-1}(\mathcal{I})$. The following proposition shows that, conversely, every bounded, invariant random variable Y corresponds to a bounded harmonic function on Γ.

Proposition 5.21. *For any bounded random variable Y that is measurable with respect to the invariant $\sigma-$algebra $\mathbf{X}^{-1}(\mathcal{I})$, the function $h : \Gamma \to \mathbb{R}$ defined by $h(x) = E^x Y$ is harmonic, and for every $x \in \Gamma$, with P^x-probability 1,*

$$Y = \lim_{n\to\infty} h(X_n). \tag{5.14}$$

Proof that h is harmonic. This is relatively straightforward; the main idea, as in the proof of Proposition 5.6, is to condition on the first step of the random walk. In carrying out the calculation, we will rely on the fact that if the random variable Y is measurable relative to the $\sigma-$algebra $\mathbf{X}^{-1}(\mathcal{I})$ then there exists a function $g : \Gamma^\infty \to \mathcal{R}$ such that for any $n \geq 1$,

$$Y = g(X_0, X_1, X_2, \cdots) = g(X_n, X_{n+1}, X_{n+2}, \cdots).$$

Consequently,

$$\begin{aligned}
E^x Y &= \sum_{y\in\Gamma} p(x,y) E^x(Y \mid X_1 = y) \\
&= \sum_{y\in\Gamma} p(x,y) E^x(g(X_1, X_2, X_3, \cdots) \mid X_1 = y) \\
&= \sum_{y\in\Gamma} p(x,y) E^x(g(y, y\xi_2, y\xi_2\xi_3, \cdots) \mid X_1 = y) \\
&= \sum_{y\in\Gamma} p(x,y) E^y(g(y, y\xi_2, y\xi_2\xi_3, \cdots) \\
&= \sum_{y\in\Gamma} p(x,y) E^y(g(X_0, X_1, X_2, \cdots) \\
&= \sum_{y\in\Gamma} p(x,y) h(y).
\end{aligned}$$

(Note: In the second to last equality, we have used the fact that the random variables $\xi_1, \xi_2, \xi_3, \cdots$ have the same joint distribution under P^y as under P^x.) □

The proof of (5.14) will require the following elementary fact from integration theory.

Lemma 5.22. *Let $Z = f(X_0, X_1, X_2, \cdots)$ be a bounded, Borel-measurable function. If $E^x(Z\mathbf{1}_C) = 0$ for every cylinder event $C = C(x_0, x_1, \cdots, x_m)$, then $P^x\{Z = 0\} = 1$.*

Proof of the limit relation (5.14). Because $h(x) = E^x Y$ is a bounded harmonic function, Theorem 5.9 implies that $Z := \lim_{n\to\infty} h(X_n)$ exists P^x- almost surely, for every $x \in \Gamma$, and $h(x) = E^x Z = E^x Y$. Consequently, to prove (5.14) it suffices to show that for any bounded, shift-invariant random variable $Y = f(X_0, X_1, X_2, \cdots)$,

$$\begin{aligned}
&E^x Y = 0 \quad \forall\, x \in \Gamma \quad \Longrightarrow \\
&P^x\{Y = 0\} = 1 \quad \forall\, x \in \Gamma.
\end{aligned} \tag{5.15}$$

Let $C = C(x_0, x_1, \cdots, x_m)$ be the cylinder event that $X_i = x_i$ for every $i = 0, 1, \cdots, m$. Since f is a shift-invariant function,

$$\begin{aligned}
E^x(Y\mathbf{1}_C) &= E^x f(X_m, X_{m+1}, \cdots)\mathbf{1}_C \\
&= E^x f(x_m, x_m\xi_{m+1}, x_m\xi_{m+1}\xi_{m+2}, \cdots)\mathbf{1}_C \\
&= E^x f(x_m, x_m\xi_{m+1}, x_m\xi_{m+1}\xi_{m+2}, \cdots) E^x \mathbf{1}_C,
\end{aligned}$$

the last equality because the steps $\xi_{m+1}, \xi_{m+2}, \cdots$ are independent (under P^x) of the random variables $X_1, X_2, \cdots, X_m$ that determine $\mathbf{1}_C$. Now since the joint distribution of the sequence $\xi_{m+1}, \xi_{m+2}, \cdots$ is the same as that of $\xi_1, \xi_2, \cdots$,

$$E^x f(x_m, x_m\xi_{m+1}, x_m\xi_{m+1}\xi_{m+2}, \cdots) = E^x f(x_m, x_m\xi_1, x_m\xi_1\xi_2, \cdots) = E^{x_m} Y = 0.$$

This proves that Y integrates to 0 on every cylinder event, and so the lemma implies that $Y = 0$ almost surely (P^x). □

5.5 The Tail $\sigma-$Algebra

For certain purposes (see section 6.3 below), the invariant $\sigma-$algebra is a rather inconvenient object to work with. For this reason, we introduce the following larger $\sigma-$algebra.

Definition 5.23. The *tail $\sigma-$algebra* $\mathcal{T}$ on Γ^∞ is the set of all Borel sets $F \in \mathcal{B}_\infty$ whose indicator functions $\mathbf{1}_F : \Gamma^\infty \to \{0,1\}$ do not depend on the first m coordinate projections, for any m, that is, for any two sequences $\mathbf{y} = (y_0, y_1, \cdots)$ and $\mathbf{y}' = (y'_0, y'_1, \cdots)$ that agree in all but finitely many entries,

$$\mathbf{y} \in F \iff \mathbf{y}' \in F. \tag{5.16}$$

Exercise 5.24. Show that the invariant $\sigma-$algebra $\mathcal{I}$ is contained in the tail $\sigma-$algebra $\mathcal{T}$.

In fact, the tail $\sigma-$algebra is, in general, much larger than the invariant $\sigma-$algebra. [7] Nevertheless, for aperiodic random walks the difference between the tail and invariant $\sigma-$algebras is not important. The following proposition explains why.

Proposition 5.25. *For any aperiodic random walk, the tail $\sigma-$algebra and the invariant $\sigma-$algebra are P^x-equivalent, for any $x \in \Gamma$. In particular, for any bounded, $\mathcal{T}-$measurable function f there exists a bounded, $\mathcal{I}-$measurable function g such that for every $x \in \Gamma$,*

$$\nu_x\{f = g\} = 1. \tag{5.17}$$

This we will deduce from the following *shift-coupling* construction.

Lemma 5.26. *If the step distribution is aperiodic, then on some probability space there exist two versions X_n and X''_n of the random walk, both with initial point $X_0 = X''_0 = x$, such that with probability 1,*

$$X''_n = X_{n+1} \quad \textit{eventually}. \tag{5.18}$$

Similarly, for any $m = 0, 1, 2, \cdots$, there exist versions X_n and X''_n of the random walk, both with initial point $X_0 = X''_0 = x$, such that with probability 1,

$$\begin{aligned} X''_n &= X_n \quad \textit{for all } n \le m, \quad \textit{and} \\ X''_n &= X_{n+1} \quad \textit{eventually}. \end{aligned} \tag{5.19}$$

Proof. We will only prove this for the special case where the random walk is *lazy* (cf. section 1.2); the reader is invited to show how to extend the construction to cover the general case. Recall that a lazy random walk is one whose step distribution is of the form $(\delta_1 + \mu)/2$; thus, at each time $n = 1, 2, \cdots$ it either stays put or makes a step with distribution μ, depending on the result of an independent fair coin toss.

Assume that the underlying probability space supports independent sequences $\{\xi_n\}_{n\ge1}$, $\{U_n\}_{n\ge1}$, and $\{V_n\}_{n\ge1}$ of i.i.d. random variables, with $\xi_n \sim \mu$ and U_n, V_n both i.i.d.

[7] Consider, for example, the case $\Gamma = \mathbb{Z}$; the event F consisting of all sequences $(x_0, x_1, x_2, \cdots)$ such that x_{2n} is even and x_{2n+1} is odd for all large n. This event is a tail event, because membership does not depend on any finite set of coordinates, but it is not shift-invariant.

Bernoulli-$(1/2)$. (The standard Lebesgue space $([0,1], \mathcal{B}_{[0,1]}, \text{Leb})$ can always be used – cf. section 7.1 in the Appendix.) Set

$$S_n^U = \sum_{i=1}^{n} U_i \quad \text{and} \quad S_n^V = \sum_{i=1}^{n} V_i;$$

then the sequences

$$\begin{aligned} X_n &= x\xi_1\xi_2\cdots\xi_{S_n^U} \quad \text{and} \\ X_n' &= x\xi_1\xi_2\cdots\xi_{S_n^V} \end{aligned}$$

are both random walks with step distribution $(\delta_1 + \mu)/2$. Now the sequence $S_n^U - S_n^V$ is an aperiodic, symmetric, nearest-neighbor random walk on $\mathbb{Z}$, so by the recurrence theorem for 1D random walks (section 2.4) it visits every integer, with probability 1. Let T be the first time that $S_n^V = S_n^U + 1$, and define

$$\begin{aligned} X_n'' &= X_n' \quad \text{for all } n \leq T, \\ &= X_{n+1} \quad \text{for all } n > T \end{aligned}$$

This is a version of the random walk with the desired property (5.18).

This construction can be easily modified so that (5.19) also holds. The idea is rather obvious: just toss the same coin for both versions of the random walk until time m, that is, replace the definition of S_n^V above by

$$S_n^V = S_{n\wedge m}^U + \sum_{i=m+1}^{n\wedge m} V_i.$$

□

Proof of Proposition 5.25. Let $Y = f(X_0, X_1, X_2, \cdots)$ be a bounded, tail-measurable random variable (that is, $f : \Gamma^\infty \to \mathbb{R}$ is measurable with respect to the tail σ−algebra $\mathcal{T}$), and let $\sigma : \Gamma \to \Gamma$ be the shift mapping

$$\sigma(x_0, x_1, x_2, \cdots) = (x_1, x_2, x_3, \cdots).$$

Our aim is to show that for every $x \in \Gamma$,

$$f(X_0, X_1, X_2, \cdots) = f(X_1, X_2, X_3, \cdots) = (f \circ \sigma)(X_0, X_1, X_2, \cdots) \quad P^x - \text{almost surely}.$$

To accomplish this we will show that the difference $g := f - f \circ \sigma$ integrates to 0 on every cylinder set, and call once again on Lemma 5.22.

Fix an arbitrary initial state x, and let X_n, X_n'' be shift-coupled versions of the random walk, as in Lemma 5.26, satisfying (5.18) and (5.19), for some fixed integer $m \geq 1$. Because f is $\mathcal{T}$−measurable, it does not depend on any finite set of coordinates; therefore, since $X_{n+1} = X_n''$ eventually with P^x−probability 1,

$$f(X_0'', X_1'', X_2'', \cdots) = f(X_1, X_2, X_3, \cdots) \quad P^x\text{−almost surely}.$$

Furthermore, since the trajectories of X''_n and X_n coincide until time $n = m$, for any cylinder set

$$C = \bigcap_{i=0}^{m} \{X_i = x_i\} = \bigcap_{i=0}^{m} \{X''_i = x_i\}$$

involving only the first $m + 1$ coordinates,

$$f(X''_0, X''_1, X''_2, \cdots)\mathbf{1}_C = f(X_1, X_2, X_3, \cdots)\mathbf{1}_C \quad P^x\text{−almost surely.}$$

Thus, for any $x \in \Gamma$,

$$\begin{aligned} E^x Y \mathbf{1}_C &= E^x f(X_0, X_1, X_2, \cdots)\mathbf{1}_C \\ &= E^x f(X''_0, X''_1, X''_2, \cdots)\mathbf{1}_C \\ &= E^x f(X_1, X_2, X_3, \cdots)\mathbf{1}_C \\ &= E^x (f \circ \sigma)(X_0, X_1, X_2, \cdots)\mathbf{1}_C. \end{aligned}$$

Since this holds for every cylinder set C, it follows from Lemma 5.22 that $f - f \circ \sigma = 0$ with ν_x−probability 1, for every initial point x. □

6 Entropy and the Liouville Property

The *Avez entropy* h of a random walk X_n with step distribution μ was defined in section 2.2 as the limit of the sequence $-E \log \mu^{*n}(X_n)/n$. In section 3 we proved, using the Carne-Varopoulos inequality, that positive entropy is equivalent to positive speed, and in section 4 we showed that irreducible random walks on nonamenable groups always have positive entropy. In this section, we show that positivity of entropy is also the determining factor for the existence of bounded harmonic functions.

Definition 6.1. A random walk has the *Liouville property* if every bounded harmonic function is constant.

Theorem 6.2. *(Avez; Derrienic; Rosenblatt; Kaimanovich & Vershik) A random walk has the Liouville property if and only if it has Avez entropy $h = 0$.*

The proof of the implication $h = 0 \implies$ Liouville property will be given in sections 6.1, 6.2, and 6.3; it uses only elementary properties of entropy. The converse implication requires in addition a non-trivial result from martingale theory, the *reverse martingale convergence theorem*; the argument is sketched in section 6.4.2.

Together with results we have already proved in sections 2, 3, and 4, Theorem 6.2 leads to a number of interesting conclusions.

Corollary 6.3. *A nearest-neighbor random walk on a group of sub-exponential growth (i.e., a group for which $\beta = 0$, where β is defined by equation (2.3)) has no non-constant bounded harmonic functions.*

Special cases of this were proved much earlier. The special case where $\Gamma = \mathbb{Z}^d$ was first established by Choquet & Deny. Doob, Snell, and Williamson subsequently showed that the Choquet-Deny theorem extends to random walk in any abelian group.

Corollary 6.4. *An irreducible, symmetric, nearest-neighbor random walk on a nonamenable group always has non-constant bounded harmonic functions.*

This follows directly from Kesten's theorem, which implies that any irreducible, symmetric random walk on a nonamenable group has positive entropy.

Exercise 6.5. Prove that the modified lamplighter random walk in dimensions $d \geq 3$ has non-constant bounded harmonic functions, and conclude that it has positive speed and entropy. HINT: The projection to $\mathbb{Z}^d$ is a (lazy) simple random walk; since this is transient, the lamplighter random walk must leave a "trail" of permanently reset lamps.

6.1 Entropy and Conditional Entropy

Definition 6.6. Let $(\Omega, \mathcal{B}, \nu)$ be a probability space and let $\mathcal{F} = \{F_i\}_{1\leq i\leq I}$ and $\mathcal{G} = \{G_j\}_{1\leq j\leq J}$ be two finite measurable partitions of Ω (that is, partitions of Ω into measurable sets, which here are defined to be sets [events] in the sigma algebra $\mathcal{B}$) Define the *entropy* (or *Shannon entropy*) $H(\mathcal{F})$ of the partition $\mathcal{F}$ and the *conditional entropy* of $\mathcal{F}$ given $\mathcal{G}$ by

$$H(\mathcal{F}) : = -\sum_{i=1}^{I} \nu(F_i)\log\nu(F_i) \quad \text{and} \tag{6.1}$$

$$H(\mathcal{F}\,|\,\mathcal{G}) := -\sum_{i=1}^{I}\sum_{j=1}^{J} \nu(F_i\cap G_j)\log\nu(F_i\,|\,G_j). \tag{6.2}$$

Note: Here we use the convention that $0\log 0 = 0$. In the definition of conditional entropy, $\nu(F\,|\,G) = \nu(F\cap G)/\nu(G)$ is the usual naive conditional probability of F given G; if $\nu(G) = 0$ then our convention implies that

$$\nu(F\cap G)\log\nu(F\,|\,G) = \nu(F\cap G)\log\nu(F\cap G) - \nu(F\cap G)\log\nu(G) = 0.$$

Substituting the expression $\nu(F\,|\,G) = \nu(F\cap G)/\nu(G)$ for the conditional probability in (6.2) gives an equivalent definition:

$$H(\mathcal{F}\,|\,\mathcal{G}) = H(\mathcal{F}\vee\mathcal{G}) - H(\mathcal{G}), \tag{6.3}$$

where $\mathcal{F}\vee\mathcal{G}$ is the *join* of the partitions $\mathcal{F}$ and $\mathcal{G}$, that is, the partition consisting of the non-empty sets in the list $\{F_i\cap G_j\}_{i\leq I, j\leq J}$. The alternative definition (6.3) is the natural analogue for entropy of the multiplication law $P(A\,|\,B)P(B) = P(A\cap B)$ for conditional probability.

A partition $\mathcal{H}$ is a *refinement* of $\mathcal{G}$, written $\mathcal{G}\preceq\mathcal{H}$, if every element of $\mathcal{G}$ is a union of elements of $\mathcal{H}$, i.e., if the elements of $\mathcal{H}$ are obtained by partitioning the elements of $\mathcal{G}$.

Proposition 6.7. *$H(\mathcal{F}\,|\,\mathcal{G}) = 0$ if and only if $\mathcal{G}$ is a refinement of $\mathcal{F}$ (up to changes by sets of probability zero).*

Proof. Since $-\log x \geq 0$ for every $x \in (0,1]$, with equality only at $x = 1$, the definition (6.2) implies that $H(\mathcal{F}\,|\,\mathcal{G}) > 0$ unless $\nu(F_i\,|\,G_j) = 1$ for every pair i, j such that $\nu(F_i \cap G_j) > 0$. This implies that every G_j is (up to a change by a set of measure 0) contained in one of the sets F_i of the partition $\mathcal{F}$. □

Proposition 6.8. *Let $\mathcal{F}, \mathcal{G}, \mathcal{H}$ be measurable partitions such that $\mathcal{H}$ is a refinement of $\mathcal{G}$. Then*

$$H(\mathcal{F}) \geq H(\mathcal{F}\,|\,\mathcal{G}) \geq H(\mathcal{F}\,|\,\mathcal{H}). \tag{6.4}$$

Furthermore, the addition law

$$H(\mathcal{F}) = H(\mathcal{F}\,|\,\mathcal{G}) \quad \Longleftrightarrow \quad H(\mathcal{F} \vee \mathcal{G}) = H(\mathcal{F}) + H(\mathcal{G}) \tag{6.5}$$

holds if and only if the partitions $\mathcal{F}$ and $\mathcal{G}$ are independent.

Proof. Both inequalities are consequences of Jensen's inequality and the fact that the function $x \mapsto -x\log x$ is strictly concave on the unit interval. In particular, Jensen implies that for each element $F_i \in \mathcal{F}$,

$$\begin{aligned}
-\sum_j \nu(F_i \cap G_j)\log\nu(F_i\,|\,G_j) &= -\sum_j \nu(G_j)\nu(F_i\,|\,G_j)\log\nu(F_i\,|\,G_j) \\
&\leq -\left(\sum_j \nu(G_j)\nu(F_i\,|\,G_j)\right)\log\left(\sum_j \nu(G_j)\nu(F_i\,|\,G_j)\right) \\
&= -\nu(F_i)\log\nu(F_i).
\end{aligned}$$

Summing over j gives the inequality $H(\mathcal{F}) \geq H(\mathcal{F}\,|\,\mathcal{G})$. The second inequality is similar.

Next, recall that Jensen's inequality $E\varphi(X) \leq \varphi(EX)$ for a strictly concave function φ is strict unless the random variable X is constant. In the application above, the random variable is the function $j \mapsto \nu(F_i\,|\,G_j)$, with each j given probability $\nu(G_j)$. Thus, strict inequality holds unless for each i the conditional probabilities $\nu(F_i\,|\,G_j)$, where j ranges over the partition $\mathcal{G}$, are all the same. But this will be the case only when $\nu(F_i\,|\,G_j) = \nu(F_i)$ for every pair i, j, that is, if the partitions are independent. □

Proposition 6.9. *(Continuity Lemma) Let $\mathcal{G}^n = \{G_j^n\}_{1\leq j\leq J}$ be a sequence of measurable partitions that converge to a measurable partition $\mathcal{G} = \{G_j\}_{1\leq j\leq J}$ in L^1-norm, that is,*

$$\lim_{n\to\infty}\sum_{j=1}^{J}\nu(G_j^n \Delta G_j) = 0.$$

Then for any measurable partition $\mathcal{F}$,

$$\lim_{n\to\infty} H(\mathcal{F}\,|\,\mathcal{G}^n) = H(\mathcal{F}\,|\,\mathcal{G}).$$

Proof. This is an easy consequence of the continuity of the function $x \mapsto x\log x$. □

6.2 Avez Entropy

Now let's return to the world of random walks. Assume that under the probability measure P^x the sequence $X_n = x\xi_1\xi_2\cdots\xi_n$ is an irreducible, nearest-neighbor random walk on a finitely generated group Γ with step distribution μ and initial state x Assume that μ gives positive probability to each element of the generating set A.

Assumption 6.10. *In this section, all entropies and conditional entropies are computed under P^x, for some arbitrary but fixed initial state $x \in \Gamma$. For notational ease, we will suppress the dependence on x will be suppressed; thus, we will write $H(\mathcal{F})$ and $H(\mathcal{F}\,|\,\mathcal{G})$ rather than $H^x(\mathcal{F})$ and $H^x(\mathcal{F}\,|\,\mathcal{G})$.*

Notational Convention: For any discrete random variable Y taking values in a finite set $\{y_i\}_{i\leq I}$, let $\pi(Y)$ be the measurable partition $\{\{Y = y_i\}\}_{i\leq I}$. For any finite collection of discrete random variables $Y_1, Y_2, \cdots, Y_n$, all taking values in finite sets, let

$$\pi(Y_1, Y_2, \cdots, Y_n) = \bigvee_{i=1}^{n} \pi(Y_i)$$

be the measurable partition obtained by taking the join of the partitions $\pi(Y_j)$.

Lemma 6.11. *For any integers $n, m \geq 1$,*

$$H(\pi(X_1)\,|\bigvee_{i=n}^{n+m}\pi(X_i)) = H(\pi(X_1)\,|\,\pi(X_n)). \tag{6.6}$$

Proof. The key is that the partition $\pi(X_n, X_{n+1}, \cdots, X_{n+m})$ is identical to the partition $\pi(X_n) \vee \pi(\xi_{n+1}, \xi_{n+2}, \cdots, \xi_{n+m})$; this is useful because the partitions

$$\pi(X_1, X_n) \quad \text{and} \quad \pi(\xi_{n+1}, \xi_{n+2}, \cdots, \xi_{n+m})$$

are independent. Thus, by the "addition law" (6.5) for independent partitions,

$$\begin{aligned}
H(\pi(X_1)\,|\bigvee_{i=n}^{n+m}\pi(X_i)) &= H(\pi(X_1) \vee \bigvee_{i=n}^{n+m}\pi(X_i)) - H(\bigvee_{i=n}^{n+m}\pi(X_i)) \\
&= H(\pi(X_1) \vee \pi(X_n, \xi_{n+1}, \cdots, \xi_{n+m})) - H(\pi(X_n, \xi_{n+1}, \cdots, \xi_{n+m}) \\
&= (H(\pi(X_1) \vee \pi(X_n)) + H(\pi(\xi_{n+1}, \xi_{n+2}, \cdots, \xi_{n+m}))) \\
&\quad - (H(\pi(X_n)) + H(\pi(\xi_{n+1}, \xi_{n+2}, \cdots, \xi_{n+m}))) \\
&= H(\pi(X_1) \vee \pi(X_n)) - H(\pi(X_n) \\
&= H(\pi(X_1)\,|\,\pi(X_n)).
\end{aligned}$$

□

The same calculation shows that for any $1 \leq k \leq n$ and any $m \geq 1$,

$$H(\bigvee_{i=1}^{k}\pi(X_i)\,|\bigvee_{i=n}^{n+m}\pi(X_i)) = H(\bigvee_{i=1}^{k}\pi(X_i)\,|\,\pi(X_n)). \tag{6.7}$$

Corollary 6.12. $H(\bigvee_{i=1}^k \pi(X_i) \,|\, \pi(X_n)) \leq H(\bigvee_{i=1}^k \pi(X_i) \,|\, \pi(X_{n+1}))$.

Proof. This is a direect consequence of equation (6.7) and the refinement inequality (6.4). $\square$

Corollary 6.13. *For any integer $k \geq 1$, the Avez entropy satisfies*

$$kh = H(\bigvee_{i=1}^k \pi(X_i)) - \lim_{n\to\infty} H(\bigvee_{i=1}^k \pi(X_i) \,|\, \pi(X_n)). \tag{6.8}$$

Proof. For ease of exposition, consider the case $k = 1$. The previous corollary implies that the sequence $H(\pi(X_1) \,|\, \pi(X_n))$ is non-decreasing, so the limit in (6.8) exists. Now rewrite the conditional entropy using equation (6.3):

$$H(\pi(X_1) \,|\, \pi(X_n)) = H(\pi(X_1) \vee \pi(X_n)) - H(\pi(X_n)).$$

Clearly, the partition $\pi(X_1) \vee \pi(X_n)$ coincides with $\pi(X_1 \vee \pi(\xi_2\xi_3 \cdots \xi_n))$, so by the addition law for independent partitions,

$$\begin{aligned} H(\pi(X_1) \vee \pi(X_n)) &= H(\pi(X_1)) + H(\pi(\xi_2\xi_3 \cdots \xi_n)) \\ &= H(\pi(X_1)) + H(\pi(X_{n-1})), \end{aligned}$$

the latter equality because the distribution of the product $\xi_2\xi_3 \cdots \xi_n$ is the same as that of $\xi_1\xi_2 \cdots \xi_{n-1}$. Consequently,

$$H(\pi(X_1) \,|\, \pi(X_n)) = H(\pi(X_1)) + H(\pi(X_{n-1})) - H(\pi(X_n)),$$

which in view of Corollary 6.12 implies that the sequence $H(\pi(X_n)) - H(\pi(X_{n-1}))$ is non-increasing in n. But for any non-increasing sequence a_n of real numbers,

$$\inf_{n\geq 1} a_n = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^n a_n;$$

since

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^n (H(\pi(X_m)) - H(\pi(X_{m-1}))) = \lim_{n\to\infty} \frac{1}{n} H(\pi(X_n)) = h,$$

the first relation follows. The proof of the general case $k \geq 1$ is similar. $\square$

6.3 Entropy Zero Implies Trivial Tail

Corollary 6.13 shows that the Avez entropy of a random walk measures the dependence between its long-time behavior and its first few steps. When the entropy is 0, long-time behavior is independent of the initial behavior, as we will see, and this leads to the following conclusion.

Proposition 6.14. *If an irreducible, nearest-neighbor random walk on Γ has Avez entropy $h = 0$ then its tail field is trivial, that is, for every tail event $F \in \mathcal{T}$, either*

$$\nu_x(F) = P^x\{(X_0, X_1, X_2, \cdots) \in F\} = 0 \text{ for all } x \in \Gamma, \text{ or}$$
$$\nu_x(F) = P^x\{(X_0, X_1, X_2, \cdots) \in F\} = 1 \text{ for all } x \in \Gamma.$$

Corollary 6.15. *If $h = 0$ then the only harmonic functions are constant.*

Proof. Since the invariant $\sigma-$algebra $\mathcal{I}$ is contained in the tail $\sigma-$algebra $\mathcal{T}$, Proposition 6.14 implies that every invariant random variable is constant; in particular, for every invariant random variable Y there is a constant C such that $P^x\{Y = C\} = 1$ for all $x \in \Gamma$. But Theorem 5.9 implies that every bounded harmonic function f has a representation $f(x) = E^x Y$ for some bounded, invariant random variable Y; consequently, f must be constant. □

The remainder of section 6.3 will be devoted to the proof of Proposition 6.14. The key step is the following lemma, which shows that entropy $h = 0$ forces independence of the tail $\sigma-$algebra.

Lemma 6.16. *If $h = 0$ then for any $k \geq 1$ the partition $\bigvee_{i=1}^k \pi(X_i)$ is independent of the tail field $\mathbf{X}^{-1}(\mathcal{T})$ under each probability measure P^x.*

Proof. Let F be an event in the $\sigma-$algebra $\mathbf{X}^{-1}(\mathcal{T})$; we must show that the partitions

$$\mathcal{G} = \{F, F^c\} \quad \text{and} \quad \bigvee_{i=1}^k \pi(X_i)$$

are independent under P^x. By Proposition 6.8, it will suffice to show that

$$H(\bigvee_{i=1}^k \pi(X_i)) = H(\bigvee_{i=1}^k \pi(X_i) \,|\, \mathcal{G}). \tag{6.9}$$

For any $n \geq 1$, the tail $\sigma-$algebra $\mathbf{X}^{-1}(\mathcal{T})$ is contained in the $\sigma-$algebra generated by the coordinate random variables $X_n, X_{n+1}, X_{n+2}, \cdots$. Consequently, any event in $\mathbf{X}^{-1}(\mathcal{T})$ can be arbitrarily well-approximated by events in the union over $m \geq 1$ of the $\sigma-$algebras $\sigma(\bigvee_{i=n}^{n+m} \pi(X_i))$; thus, in particular, for each n there exists $m(n) \geq 1$ and an event $F_n \in \sigma(\bigvee_{i=n}^{n+m(n)} \pi(X_i))$ such that $P^x(F \Delta F_n) \leq 2^{-n}$. Let $\mathcal{G}_n = \{F_n, F_n^c\}$; by construction, the partitions $\mathcal{G}_n$ converge to the partition $\mathcal{G}$ in L^1-norm, so by the Continuity Lemma (Proposition 6.9),

$$H(\bigvee_{i=1}^k \pi(X_i) \,|\, \mathcal{G}) = \lim_{n \to \infty} H(\bigvee_{i=1}^k \pi(X_i) \,|\, \mathcal{G}^n).$$

Now for each $n \geq 1$ the partition $\bigvee_{i=n}^{n+m(n)} \pi(X_i)$ is a refinement of $\mathcal{G}_n$, since the set F_n is a union of sets in the partition $\bigvee_{i=n}^{n+m(n)} \pi(X_i)$, so the Monontonicity Principle (Proposition 6.8) implies that

$$\begin{aligned} H(\bigvee_{i=1}^{k} \pi(X_i)) &\geq H(\bigvee_{i=1}^{k} \pi(X_i) \,|\, \mathcal{G}_n) \\ &\geq H(\bigvee_{i=1}^{k} \pi(X_i) \,|\, \bigvee_{i=n}^{n+m(n)} \pi(X_i)) \\ &= H(\bigvee_{i=1}^{k} \pi(X_i) \,|\, \pi(X_n)), \end{aligned}$$

the last by Lemma 6.11. But Corollary 6.13 implies that if $h = 0$ then

$$H(\bigvee_{i=1}^{k} \pi(X_i)) = \lim_{n\to\infty} H(\bigvee_{i=1}^{k} \pi(X_i) \,|\, \pi(X_n));$$

hence,

$$H(\bigvee_{i=1}^{k} \pi(X_i)) = \lim_{n\to\infty} H(\bigvee_{i=1}^{k} \pi(X_i) \,|\, \mathcal{G}_n).$$

This proves (6.9). □

Proof of Proposition 6.14. By Lemma 6.16, if $h = 0$ then, under any P^x, the tail $\sigma-$algebra $\mathbf{X}^{-1}(\mathcal{T})$ is independent of the partition $\bigvee_{i=1}^{k} \pi(X_i)$, that is, every event $F \in \mathbf{X}^{-1}(\mathcal{T})$ is independent of every event $G \in \bigvee_{i=1}^{k} \pi(X_i)$. These partitions generate the $\sigma-$algebra

$$\mathcal{F}_\infty := \sigma(X_1, X_2, X_3. \cdots);$$

therefore, the $\sigma-$algebras $\mathbf{X}^{-1}(\mathcal{T})$ and $\mathcal{F}_\infty$ are independent. But $\mathbf{X}^{-1}\mathcal{T}$ is contained in $\mathcal{F}_\infty$; thus, it follows that the tail $\sigma-$algebra *is independent of itself*, that is, for any two events $F, G \in \mathbf{X}^{-1}(\mathcal{T})$,

$$P^x(F \cap G) = P^x(F)P^x(G).$$

This holds, in particular, for $F = G$, and so for every event $F \in \mathbf{X}^{-1}(\mathcal{T})$ we must have $P^x(F)^2 = P^x(F)$, which implies that $P^x(F)$ is either 0 or 1. Finally, the mutual absolute continuity of the exit measures ν_x (Proposition 5.20) guarantees that the value of $P^x(F)$ must be the same for all tail events F. □

6.4 Trivial Tail Implies Entropy Zero

Half of Theorem 6.2 is now proved. In this section, we tackle the other half: we will prove that if a random walk has *positive* Avez entropy, then it has non-constant bounded harmonic functions.

Several simplifications can be made at the outset. First, to prove that a random walk has non-constant bounded harmonic functions, it suffices, by Proposition 5.21, to show that there exists an invariant event $F \in \mathbf{X}^{-1}(\mathcal{I})$ whose P^x-probability is neither 0 nor 1, for some $x \in \Gamma$. For if such an event exists, then the function $f(x) := E^x \mathbf{1}_F$ is bounded and harmonic, and by Proposition 5.21

$$\lim_{n\to\infty} f(X_n) = \mathbf{1}_F$$

almost surely, so f cannot be constant. Second, we can, without loss of generality, assume that the random walk is *lazy*, because (i) the lazy version of a random walk has the same harmonic functions as the original, and (ii) if a random walk has positive Avez entropy, then so does its lazy version (cf. Exercise 2.10). Third, since for a lazy random walk the tail $\sigma-$algebra and the invariant $\sigma-$algebra are equivalent (cf. Proposition 5.25), it will suffice to show that the tail $\sigma-$algebra is non-trivial, that is, that there is a tail event $F \in \mathbf{X}^{-1}(\mathcal{T})$ whose P^x-probability is neither 0 nor 1, for some x.

6.4.1 Conditional Distributions and their Radon-Nikodym Derivatives

For each $n = 0, 1, 2, \cdots$ denote by $\mathcal{G}_n$ the $\sigma-$algebra generated by the random variables $X_n, X_{n+1}, \cdots$. These $\sigma-$algebras are nested:

$$\mathcal{G}_0 \supset \mathcal{G}_1 \supset \mathcal{G}_2 \supset \cdots,$$

and their intersection is the tail $\sigma-$algebra $\mathbf{X}^{-1}(\mathcal{T})$. For each initial point $x \in \Gamma$ and each $n \geq 1$, the probability measure P^x on $\mathcal{G}_0$ restricts to a probability measure on $\mathcal{G}_n$, which we shall also denote by P^x. Similarly, for each $x \in \Gamma$ such that $P^1\{X_1 = x\} = \mu(x) > 0$, the conditional law of the the random walk given the event $X_1 = x$ induces a probability measure on $\mathcal{G}_n$, which we denote by Q^x. Formally, for every event $G \in \mathcal{G}_n$, define

$$Q^x(G) = P^x(G \,|\, X_1 = x) = \frac{P^x(G \cap \{X_1 = x\})}{P^x\{X_1 = x\}}.$$

By definition, the measure Q^x is absolutely continuous with respect to P^x on every $\mathcal{G}_\backslash$, and hence also on $\mathbf{X}^{-1}(\mathcal{T})$. Denote the Radon-Nikodym derivatives by

$$L_n^x := \left(\frac{dQ^x}{dP^1}\right)_{\mathcal{G}_n} \quad \text{and} \quad L_\infty^x := \left(\frac{dQ^x}{dP^1}\right)_{\mathbf{X}^{-1}(\mathcal{T})}.$$

Lemma 6.17. *The Radon-Nikodym derivative L_n^x is a function only of X_n, in particular,*

$$L_n^x = \sum_{y\in\Gamma} \frac{P^x(X_n = y \,|\, X_1 = x)}{P^1(X_n = y)} \mathbf{1}_{\{X_n = y\}}$$

Proof. Exercise. □

Lemma 6.18. *As $n \to \infty$,*

$$L_n^x \longrightarrow L_\infty^x \quad P^1-\textit{almost surely and in } L^1.$$

Lemma 6.18 a consequence of the *Reverse Martingale Convergence Theorem*. See my Lecture Notes on Discrete-Time Martingales for a complete discussion of this important theorem.

6.4.2 Non-Trivial Tail Events

Proposition 6.19. *If the Avez entropy of the random walk is positive, then for at least one $x \in \Gamma$ in the support of the step distribution μ and some $\varepsilon > 0$, the event $\{L^x_\infty \geq 1 + \varepsilon\}$ is a non-trivial tail event, that is,*

$$0 < P^1\{L^x_\infty \geq 1 + \varepsilon\} < 1. \tag{6.10}$$

Proof. That $\{L^x_\infty \geq 1 + \varepsilon\}$ is a tail event is an immediate consequence of the definition of L^x_∞ as a Radon-Nikodym derivative. The upper inequality in (6.10) follows easily from the fact that Q^x is a probability measure:

$$(1 + \varepsilon)P^1\{L^x_\infty \geq 1 + \varepsilon\} \leq Q^x\{L^x_\infty \geq 1 + \varepsilon\} \leq 1.$$

To prove the lower inequality in (6.10), it suffices to show that for some x and some $\delta > 0$,

$$P^1\{L^x_n \geq 1 + \delta\} \geq \delta \tag{6.11}$$

for all large n. This is where we will use the hypothesis that the Avez entropy is positive. By Corollary 6.13 and the equivalent definition (6.3) of conditional entropy,

$$\begin{aligned} h &= \lim_{n\to\infty} H(\pi(X_1)) - H(\pi(X_1) \,|\, \pi(X_n)) \\ &= \lim_{n\to\infty} H(\pi(X_n)) - H(\pi(X_n) \,|\, \pi(X_1)). \end{aligned}$$

Now by Lemma 6.17,

$$\begin{aligned} &H(\pi(X_n)) - H(\pi(X_n) \,|\, \pi(X_1)) \\ &\qquad = \sum_x \sum_y P^1(\{X_1 = x\} \cap \{X_n = y\}) \log \frac{P^1(X_n = y \,|\, X_1 = x)}{P^1(X_n = y)} \\ &\qquad = \sum_x \sum_y P^1(\{X_1 = x\} \cap \{X_n = y\}) \log L^x_n \mathbf{1}_{\{X_n = y\}}; \end{aligned}$$

in order that this remain bounded away from 0 as $n \to \infty$ it is necessary that, for at least some x with $\mu(x) > 0$, the Radon-Nikodym derivative L^x_n remain bounded away from 1 with substantial probability. This proves (6.11).

□

7 Appendix

7.1 A 30-Second Course in Measure–Theoretic Probability

If you have only had an undergraduate course in probability, you will know how to build *finite* probability spaces that support finite sequences of independent, discrete random variables with prescribed distributions. Where does one find a probability space that supports an *infinite* sequence of independent, identically distributed random variables, as are needed to specify the steps of a random walk? One approach is to build everything on the unit interval $[0, 1]$, equipped with *Lebesgue measure*[8], as follows. Suppose, for instance, that you want an infinite sequence $\xi_1, \xi_2, \cdots$ of random variables taking values in the set $\{a, b, c\}$ with probabilities p_a, p_b, p_c, respectively. Begin by partitioning the interval $[0, 1]$ into three non-overlapping intervals J_a, J_b, J_c of lengths p_a, p_b, p_c, and define $\xi_1 = i$ on J_i. Next, partition each of the intervals J_i into three sub-intervals J_{ia}, J_{ib}, J_{ic} in the same proportions p_a, p_b, p_c, and for each of these nine sub-intervals, set $\xi_2 = j$ on J_{ij}. By construction,

$$P\{\xi_1 = i \text{ and } \xi_2 = j\} = p_i p_j,$$

so the random variables ξ_1, ξ_2 are independent, each with the same distribution. Now continue the partitioning inductively, and use the sub-intervals of the nth generation to define the random variable ξ_n. The result will be an infinite sequence $\xi_1, \xi_2, \xi_2, \cdots$ of independent, identically distributed distributions with the desired distribution.

7.2 Hoeffding's Inequality and the SLLN

Proposition 7.1. *Let $Y_1, Y_2, \cdots$ be independent real-valued random variables such that $EY_n = 0$ and $|Y_n| \leq 1$. Let $S_n = \sum_{i \leq n} Y_i$ be the nth partial sum. Then for any $\alpha > 0$ and every $n = 1, 2, \cdots$,*

$$P\{|S_n| \geq \alpha\} \leq 2\exp\{-\alpha^2/2n\}.$$

This is a standard result in elementary probability, and its (relatively easy) proof can be found in most textbooks (and in this WIKIPEDIA article). Observe that the choice $\alpha = n\varepsilon$ gives exponential decay (in n) of the large deviation probability $P\{|S_n| \geq n\varepsilon\}$. Therefore, in particular, if $Y_1, Y_2, \cdots$ are independent, identically distributed random variables with mean $EY_i = 0$ and satisfying $|Y_i| \leq 1$ then

$$\sum_{n=1}^{\infty} P\{|S_n| \geq n\varepsilon\} = E\left(\sum_{n=1}^{\infty} \mathbf{1}\{|S_n| \geq n\varepsilon\}\right) < \infty.$$

Thus, the *expected* number of times n that $|S_n| \geq n\varepsilon$ is finite, and so this number is also *almost surely* finite. Since this is true for every (rational) $\varepsilon > 0$, it follows that with probability one the limsup of the sequence $|S_n|/n$ is 0. This proves the strong law of large numbers in the special case where the summands Y_i are bounded. $\square$

7.3 Stirling's Formula

Proposition 7.2. *(Stirling) As $n \to \infty$,*

$$n! \sim n^n e^{-n}\sqrt{2\pi n}. \tag{7.1}$$

[8] For a brief introduction to the measure-theoretic underpinnings of probability theory, see my Lecture Notes on Measure Theory, or for a more thorough exposition, see H. L. ROYDEN, *Real Analysis*.

Corollary 7.3. *As* $\min(m, n-m) \to \infty$,

$$\binom{n}{m} \sim \frac{\exp\{nH(\frac{m}{n})\}}{\sqrt{2\pi n(\frac{m}{n})(\frac{n-m}{n})}} \tag{7.2}$$

where for all $p \in (0,1)$ *and* $q = 1-p$

$$H(p) = -p\log p - q\log q. \tag{7.3}$$

In both statements, the relation $\sim$ means that the ratio of the two sides converges to 1. The corollary holds *uniformly* in the range

$$\min(m, n-m) \longrightarrow \infty;$$

more precisely, for every $\varepsilon > 0$ there exists $m(\varepsilon) > 0$ such that if $\min(m, n-m) > m(\varepsilon)$ then

$$\left|\binom{n}{m}\Bigg/\left(\frac{\exp\{nH(\frac{m}{n})\}}{\sqrt{2\pi n(\frac{m}{n})(\frac{n-m}{n})}}\right) - 1\right| < \varepsilon.$$

Exercise 7.4. Use the corollary to prove that the simple random walk in $d \geq 3$ dimensions satisfies

$$P\{S_{2n} = 0\} \leq Cn^{-d/2}.$$