Lecture Notes on Random Walks

By

Omer Tamuz

California Institute of Technology

2022

Contents

Acknowledgments

These lecture note have adapted ideas from a number of expositional texts, including work by Steve Lalley, Russell Lyons, Yuval Peres and Terry Tao. I am indebted to Kimberly Golubeva, Michael Wolman and especially Fan Wu for their help in finding and correcting many errors. Any comments or suggestions are welcome.

Disclaimer

This a not a textbook. These are lecture notes.

1 Random walks on $\mathbb{Z}$

1.1 Definitions

Let μ be a probability measure on $\mathbb{Z}$. Since $\mathbb{Z}$ is countable we can think of μ as a function $\mu\colon \mathbb{Z} \to \mathbb{R}_+$ with $\sum_{x\in\mathbb{Z}} \mu(x) = 1$.

Let $(X_1, X_2, \dots)$ be a sequence of independent random variables each having distribution μ. Denote $Z_n = X_1 + \cdots + X_n$, and set $Z_0 = 0$. We call the process $(Z_0, Z_1, Z_2, \dots)$ the *μ-random walk on $\mathbb{Z}$*. For notational convenience we denote $X = X_1$.

If you prefer a measure-theoretic perspective, Let $\Omega = \mathbb{Z}^{\mathbb{N}}$, and equip it with the product topology. Thus an element of Ω is a sequence $\omega = (\omega_1, \omega_2, \dots)$ of integers, and a sequence of sequences converges if each coordinate eventually stabilizes. Let $\mathscr{F}$ be the Borel sigma-algebra. Let $\mathbb{P}$ be the product measure $\mu^{\mathbb{N}}$. Define $X_n : \Omega \to \mathbb{Z}$ by $X_n(\omega) = \omega_n$, and $Z_n(\omega) = \omega_1 + \cdots + \omega_n$.

A μ-random walk on $\mathbb{Z}$ is a *Markov chain* with state space $\mathbb{Z}$. The transition probabilities are $P(x, y) = \mu(y - x)$. We will assume that the random walk is *non-degenerate*: for every $z \in \mathbb{Z}$ there is an n such that $\mathbb{P}[Z_n = z] > 0$. Equivalently, the Markov chain is *irreducible*.

A good example to keep in mind is the *simple random walk*: this is the case that $\mu(-1) = \mu(+1) = 1/2$. Another good example is a *lazy simple random walk*, given by $\mu(-1) = \mu(1) = 1/2 - c$, $\mu(0) = 2c$ for some $0 < c < 1/2$. Unless otherwise indicated, we will assume that μ has finite support, i.e., the set $\{x : \mu(x) > 0\}$ is finite. In other cases it will be useful to consider random walks on $\mathbb{R}$, so that μ is a probability measure on the reals. Later in the course we will consider random walks on additional objects.

Denote

$$\alpha = \mathbb{E}[X] = \sum_{x\in\mathbb{Z}} x\mu(x).$$

We call α the *drift* of the random walk. Denote

$$\sigma^2 = \mathrm{Var}(X) := \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \sum_{x\in Z} x^2\mu(x) - \alpha^2.$$

Note that

$$\mathbb{E}[Z_n] = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = n\alpha$$

and that

$$\mathrm{Var}(Z_n) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\sigma^2,$$

since the variance of a sum of independent random variables is the sum of their variances. Hence

$$\mathrm{Std}(Z_n) := \sqrt{\mathrm{Var}(Z_n)} = \sqrt{n}\sigma.$$

1.2 The weak law of large numbers

Theorem 1.1 (The weak law of large numbers)**.** *For all $n \geq 1$ and $M > 0$,*

$$\mathbb{P}\left[\alpha n - M\sigma\sqrt{n} < Z_n < \alpha n + M\sigma\sqrt{n}\right] \geq 1 - \frac{1}{M^2}.$$

In particular, when $\mathbb{E}[X] = 0$, $\mathbb{P}\left[|Z_n| < M\sigma\sqrt{n}\right] \geq 1 - 1/M^2$.

To prove this theorem we will need Markov's inequality, which states that for every non-negative random variable W with $\mathbb{E}[W] = w$ it holds that

$$\mathbb{P}[W \geq Mw] \leq \frac{1}{M}.$$

Proof of Theorem 1.1. Note that

$$\mathbb{E}\left[(Z_n - \alpha n)^2\right] = \mathbb{E}\left[Z_n^2 - 2Z_n\alpha n + \alpha^2 n^2\right] = \mathbb{E}\left[Z_n^2\right] - \mathbb{E}[Z_n]^2 = \mathrm{Var}(Z_n) = n\sigma^2.$$

Therefore, by Markov's inequality applied to the random variable $(Z_n - \alpha n)^2$,

$$\mathbb{P}\left[(Z_n - \alpha n)^2 \geq M^2 n\sigma^2\right] \leq \frac{1}{M^2}.$$

The event $\{(Z_n - \alpha n)^2 \geq M^2 n\sigma^2\}$ is the same as the event $\{|Z_n - \alpha n| \geq M\sqrt{n}\sigma\}$, which is the complement of the event we are interested in, and thus we have proved the claim. $\square$

In fact, the Central Limit Theorem gives us a much more precise version of this claim, telling not only where Z_n concentrates, but also what its distribution looks like. Denote by $\Phi(x)$ the cdf (cumulative distribution function) of a standard Gaussian:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t.$$

Theorem 1.2 (Central Limit Theorem)**.** *For all $M \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left[Z_n \leq \alpha n + M\sigma\sqrt{n}\right] = \Phi(M).$$

We will not prove this theorem in this course.

The Central Limit Theorem gives us a handle for what the cdf of Z_n looks like, for large n, within distance $O(\sqrt{n})$ from the expectation αn. What about what happens within distance $O(n)$ from αn? For for $\beta > \alpha$ what can we say about $\mathbb{P}\left[Z_n > \beta n\right]$?

Suppose $\alpha = 0$ and $\sigma = 1$. If the Central Limit Theorem held beyond the $\sqrt{n}$ regime then it would imply that $\mathbb{P}\left[Z_n > \beta n\right] \approx 1 - \Phi(\beta\sqrt{n})$. Since $\Phi(x) \approx 1 - \exp(-x^2)$ for large x, this would mean that $\mathbb{P}\left[Z_n > \beta n\right] \approx \exp(-\beta^2 n)$. As we will show, the exponential dependence on n is correct, but the coefficient β^2 is not.

1.3 The moment and cumulant generating functions

For the next results we will need to define the *moment generating function of X*:

$$M_X(t) := \mathbb{E}\left[e^{tX}\right] = \sum_{x \in Z} e^{tx} \mu(x).$$

The name comes from the fact that

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}\left[X^n\right]. \tag{1.1}$$

Note that this means that $M'_X(0) = \mathbb{E}[X]$, and more generally $M_X^{(k)}(0) = \mathbb{E}\left[X^k\right]$. The *cumulant generating function of X* is given by $K_X(t) := \log M_X(t)$. As it turns out (but we will not prove), K_X is a convex function. Under our assumption of finitely supported μ, it is clear that K_X is furthermore analytic, since

$$K_X(t) = \log \sum_{x \in \mathbb{Z}} e^{tx} \mu(x),$$

and the sum has finitely many terms.

The most important property of K_X is its additivity with respect to sums of independent random variables. That is, if X and Y are independent then $K_{X+Y} = K_X + K_Y$, since

$$M_{X+Y}(t) = \mathbb{E}\left[e^{t(X+Y)}\right] = \mathbb{E}\left[e^{tX} e^{tY}\right] = \mathbb{E}\left[e^{tX}\right] \mathbb{E}\left[e^{tY}\right] = M_X(t) \cdot M_Y(t).$$

In particular this implies that $K_{Z_n} = nK_X$. In comparison, there is a much more complicated relationship between the cumulative distribution functions of X and Z_n.

1.4 The Chernoff bound

Theorem 1.3 (Chernoff bound)**.** *Let $\alpha = \mathbb{E}[X]$. Then for every $\beta > \alpha$*

$$\mathbb{P}\left[Z_n \geq \beta n\right] \leq e^{-r \cdot n}$$

where

$$r := \sup_{t \geq 0} \{t \cdot \beta - K_X(t)\} > 0.$$

Proof of Theorem 1.3. Denote $p_n = \mathbb{P}\left[Z_n \geq \beta n\right]$; we want to show that $p_n \leq e^{-r \cdot n}$.

Note that the event $\{Z_n \geq \beta n\}$ is identical to the event $\{e^{t \cdot Z_n} \geq e^{t \cdot \beta n}\}$, for any $t > 0$. Since $e^{t \cdot Z_n}$ is a positive random variable with expectation $M_{Z_n}(t)$, by the Markov inequality we have that

$$p_n = \mathbb{P}\left[e^{t \cdot Z_n} \geq e^{t \cdot \beta n}\right] \leq \frac{M_{Z_n}(t)}{e^{t \cdot \beta n}}.$$

Since $M_{Z_n}(t) = M_X(t)^n = \exp(nK_X(t))$ we have that

$$p_n \leq \exp\left(-(t \cdot \beta - K_X(t)) \cdot n\right).$$

Since $K_X'(0) = M_X'(0)/M_X(0) = \mathbb{E}[X]$, and since K_X is smooth, it follows that for $t > 0$ small enough,

$$t \cdot \beta - K_X(t) = t \cdot \beta - t \cdot \alpha - O(t^2) > 0.$$

Hence

$$p_n \leq \mathrm{e}^{-r \cdot n}.$$

for

$$r = \sup_{t \geq 0}\{t \cdot \beta - K_X(t)\} > 0.$$

□

It turns out that the Chernoff bound is asymptotically tight, in the sense that $\mathbb{P}\left[Z_n \geq \beta n\right] = \mathrm{e}^{-rn+o(\log n)}$, for all β less than the maximum of the support of X. We will prove this later.

1.5 The Legendre transform

Let the *Legendre transform* of K be given by

$$K^\star(\beta) = \sup_{t>0}(t\beta - K(t)).$$

It turns out that the fact that K is smooth and convex implies that $K^\star$ is also smooth and convex. Therefore, if the supremum in this definition is obtained at some t, then $K'(t) = \beta$. Conversely, if $K'(t) = \beta$ for some t, then this t is unique and $K^\star(\beta) = t\beta - K(t)$. Using this notation we can write the Chernoff bound as

$$\mathbb{P}\left[Z_n \geq \beta n\right] \leq \mathrm{e}^{-K^\star(\beta)n}.$$

1.6 The Hoeffding bound

The Chernoff bound implies a simpler bound, when combined with the following lemma, which we will not prove.

Lemma 1.4 (Hoeffding Lemma)**.** *If Y is a random variable with $\mathbb{E}[Y] = 0$ and $|Y| \leq M$ almost surely then $K_Y(t) \leq \frac{1}{2}M^2t^2$.*

Note that $\frac{1}{2}M^2t^2$ is equal to $K_W(t)$, where W is a Gaussian random variable with mean 0 and variance M^2.

Theorem 1.5 (The Hoeffding bound)**.** *Suppose* $|X| \le M$ *almost surely and* $\mathbb{E}[X] = 0$*. Then for every* $\beta > 0$

$$\mathbb{P}\left[Z_n \ge \beta n\right] \le \mathrm{e}^{-\frac{\beta^2}{2M^2}\cdot n}.$$

Proof. By Hoeffding's Lemma

$$\sup_{t\ge 0} t\beta - K_x(t) \ge t\beta - \frac{1}{2}M^2t^2.$$

Hence by choosing $t = \beta/M^2$ we get that

$$\sup_{t\ge 0} t\beta - K_x(t) \ge \beta^2/M^2 - \frac{1}{2}\beta^2/M^2 = \frac{1}{2}\beta^2/M^2.$$

Hence the claim follows by the Chernoff bound. □

1.7 The strong law of large numbers

The weak law of large numbers implies that

$$\lim_n \mathbb{P}\left[\left|\frac{1}{n}Z_n - \alpha\right| > \varepsilon\right] = 0$$

for all $\varepsilon > 0$. In fact, this is the usual statement of the weak law of large numbers. This does not immediately imply that $\frac{1}{n}Z_n$ converges almost surely to α (in fact, this is not true for some infinitely supported μ). It does for the finitely supported μ that we consider here, which is the content of the *strong law of large numbers*.

Theorem 1.6 (The strong law of large numbers)**.** $\lim_n \frac{1}{n}Z_n = \alpha$ *almost surely.*

To prove this theorem we will need the *Borel-Cantelli Lemma*. Let $(A_1, A_2, \ldots)$ be a sequence of events. The event

$$(A_n)_n \text{ i.o.} := \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} A_n$$

is the event that infinitely many of these events occur.

Lemma 1.7 (Borel-Cantelli Lemma)**.** *Let* $(A_1, A_2, \ldots)$ *be a sequence of events. If* $\sum_n \mathbb{P}[A_n] < \infty$ *then*

$$\mathbb{P}[(A_n)_n \textit{ i.o.}] = 0.$$

Proof of Theorem 1.6. Let

$$A_{n,m} = \left\{\frac{1}{n}Z_n > \alpha + \frac{1}{m}\right\}$$

be the event that $\frac{1}{n}Z_n$ exceeds α by more than $1/m$.

By the Chernoff bound, for each m there is some $r > 0$ such that $\mathbb{P}\left[A_{n,m}\right] \leq \mathrm{e}^{-rn}$ for all n. Since $\sum_n \mathrm{e}^{-Rx} < \infty$, it follows from Borel-Cantelli that $\mathbb{P}\left[(A_{n,m})_n \text{ i.o.}\right] = 0$. Thus, almost surely, $\frac{1}{n}Z_n > \alpha + \frac{1}{m}$ only finitely many times, and so

$$\limsup_n \frac{1}{n}Z_n \leq \alpha + \frac{1}{m}$$

almost surely. Since this holds for every m, $\limsup_n \frac{1}{n}Z_n \leq \alpha$. By a symmetric argument $\liminf_n \frac{1}{n}Z_n \geq \alpha$, and so $\lim_n \frac{1}{n}Z_n = \alpha$ almost surely. $\square$

Remark 1.8. *All of the results in this section generalize far beyond finitely supported μ, but none of them apply to every infinitely supported μ. Exploring when these results do and do not hold will not be our focus.*

2 Large deviations

By the law of large numbers we expect that a μ-random walk Z_n should be close to its drift $\alpha = \mathbb{E}[X]$ for large n. What is the probability that it is larger than some $\beta > \alpha$? We already proved the Chernoff lower bound. We here prove an asymptotically matching upper bound.

2.1 The cumulant generating function

In this section we simplify notation and denote $M := M_X$ and $K = K_X$ so that the *moment generating function* of X is

$$M(t) = \mathbb{E}\left[e^{tX}\right],$$

and that its *cumulant generating function* is

$$K(t) = \log M(t) = \log \mathbb{E}\left[e^{tX}\right].$$

Claim 2.1. *K is convex.*

For the proof of this claim we will need Hölder's inequality. For $p \in [1,\infty]$ and a real r.v. Y denote

$$|Y|_p = \mathbb{E}\left[|Y|^p\right]^{1/p}.$$

Lemma 2.2 (Hölder's inequality)**.** *For any $p,q \in [1,\infty]$ with $1/p + 1/q = 1$ and r.v.s X,Y it holds that*

$$|X \cdot Y|_1 \le |X|_p \cdot |Y|_q\,.$$

Proof of Claim 2.1. Choose $a,b \in \mathbb{R}$. Then for any $r \in (0,1)$

$$K(ra+(1-r)b) = \log \mathbb{E}\left[e^{(ra+(1-r)b)X}\right] = \log \mathbb{E}\left[\left(e^{aX}\right)^r \left(e^{bX}\right)^{1-r}\right].$$

By Hölder's inequality

$$\begin{aligned} K(ra+(1-r)b) &\le \log \mathbb{E}\left[e^{aX}\right]^r + \log \mathbb{E}\left[e^{bX}\right]^{1-r} \\ &= r \log \mathbb{E}\left[e^{aX}\right] + (1-r)\log \mathbb{E}\left[e^{bX}\right] \\ &= rK(a) + (1-r)K(b). \end{aligned}$$

□

2.2 Convolution

The probability that $Z_2 = x$ is

$$\mathbb{P}[Z_2 = x] = \sum_y \mathbb{P}[Z_2 = x, X_1 = y] = \sum_y \mathbb{P}[X_2 = x - y, X_1 = y] = \sum_y \mu(x-y)\mu(y).$$

More generally, if X has distribution μ and X' is independent with distribution ν, and we denote the distribution of $X + X'$ by ζ, then

$$\zeta(x) = \sum_y \mu(x-y)\nu(y) = \sum_y \nu(x-y)\mu(y).$$

The operation $(\mu,\nu) \mapsto \zeta$ is called *convolution*, and we denote $\zeta = \mu * \nu$. We denote the n-fold convolution of μ with itself by $\mu^{(n)}$, so that for a μ-random walk the distribution of Z_n is $\mu^{(n)}$.

2.3 Large deviations

Denote $\operatorname{supp}\mu = \{x \in \mathbb{Z} : \mu(x) > 0\}$.

Theorem 2.3. *For any* $\beta \in [\alpha, \max\operatorname{supp}\mu)$

$$\mathbb{P}\left[Z_n \geq \beta n\right] = \mathrm{e}^{-K^\star(\beta)n + o(n)}.$$

Proof. One side is given by the Chernoff bound. It thus remains to prove the lower bound. We want to prove that

$$\limsup_n -\frac{1}{n}\log\mathbb{P}\left[Z_n \geq \beta n\right] \leq K^\star(\beta).$$

As we noted above, $K'(0) = \alpha$. It can be shown that

$$\lim_{t\to\infty} K'(t) = \max\operatorname{supp}\mu.$$

Hence for every β such that $\alpha \leq \beta < \max\operatorname{supp}\mu$ there is a t^* such that $\beta = K'(t^*)$. Since K is convex and and smooth its derivative is increasing almost everywhere, and hence such a t^* exists and is unique if and only if $\alpha \leq \beta < M$.

Fix $\overline{\beta} \in (\beta, \max\operatorname{supp}\mu)$, let $\bar{t}$ be given by $K'(\bar{t}) = \overline{\beta}$, and fix $t \in (t^*, \bar{t})$. Define the measure $\tilde{\mu}$ by

$$\tilde{\mu}(x) = \frac{\mathrm{e}^{tx}}{\sum_y \mathrm{e}^{ty}\mu(y)}\mu(x) = \mathrm{e}^{tx - K(t)}\mu(x),$$

and let $(\tilde{X}_1, \tilde{X}_2, \ldots,)$ be the steps of $\tilde{\mu}$-random walk on $\mathbb{Z}$. Denote $\tilde{Z}_n = \tilde{X}_1 + \cdots + \tilde{X}_n$.

Note that

$$\mathbb{P}\left[\tilde{Z}_2 = z\right] = \tilde{\mu}^{(2)}(z) = \sum_y \tilde{\mu}(z-y)\tilde{\mu}(y)$$

by the definition of Z_2 and of convolution. Hence by the definition of $\tilde{\mu}$

$$\mathbb{P}\left[\tilde{Z}_2 = z\right] = \sum_y \mathrm{e}^{t(z-y)-K(t)}\mu(z-y)\mathrm{e}^{ty-K(t)}\mu(y) = \mathrm{e}^{tz-2K(t)}\sum_y \mu(z-y)\mu(y) = \mathrm{e}^{tz-2K(t)}\mathbb{P}[Z_2 = z].$$

Likewise,

$$\mathbb{P}\left[\tilde{Z}_n = z\right] = \mathrm{e}^{tz-nK(t)}\mathbb{P}[Z_n = z].$$

Remark 2.4. *More generally, if we denote by $\Delta_f(\mathbb{Z})$ the finitely supported probability measures on $\mathbb{Z}$, then the "tilting" operation $T_t \colon \Delta_f(\mathbb{Z}) \to \Delta_f(\mathbb{Z})$ given by $\mu \mapsto \tilde{\mu}$ commutes with the convolution operation:*

$$(T_t\mu) * (T_t\nu) = T_t(\mu * \nu).$$

I.e., T_t is an automorphism of the semigroup $(\Delta(\mathbb{Z})_f, *)$.

Using the fact that the expectation of a random variable is equal to the derivative at zero of its cumulant generating function, a simple calculation shows that

$$\mathbb{E}\left[\tilde{X}_1\right] = K'(t) \in \left(\beta, \overline{\beta}\right).$$

It follows that

$$\begin{aligned}
\mathbb{P}\left[\beta n \le Z_n\right] &\ge \mathbb{P}\left[\beta n \le Z_n \le \overline{\beta} n\right] \\
&= \sum_{z=\lceil \beta n\rceil}^{\lfloor \overline{\beta} n\rfloor} \mathbb{P}[Z_n = z] \\
&= \sum_{z=\lceil \beta n\rceil}^{\lfloor \overline{\beta} n\rfloor} \mathbb{P}\left[\tilde{Z}_n = z\right]\mathrm{e}^{-(tz-nK(t))} \\
&\ge \mathrm{e}^{-(t\overline{\beta}n-nK(t))} \sum_{z=\lceil \beta n\rceil}^{\lfloor \overline{\beta} n\rfloor} \mathbb{P}\left[\tilde{Z}_n = z\right] \\
&= \mathrm{e}^{-(t\overline{\beta}-K(t))n}\mathbb{P}\left[\beta n \le \tilde{Z}_n \le \overline{\beta} n\right].
\end{aligned}$$

Since $\mathbb{E}\left[\tilde{Z}_n\right] \in (\beta n, \overline{\beta} n)$, and since $\tilde{Z}_n$ is a $\tilde{\mu}$-random walk, by the law of large numbers

$$\lim_{n\to\infty} \mathbb{P}\left[\beta n \le \tilde{Z}_n \le \overline{\beta} n\right] = 1,$$

and so

$$\lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}\left[\beta n \le Z_n\right] \ge -(t\overline{\beta} - K(t)).$$

Since this holds for any $\overline{\beta} > \beta$ and $\overline{\beta} > K'(t) > \beta$, it also holds for $\overline{\beta} = \beta$ and t^* such that $K'(t^*) = \beta$. So

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}\left[\beta n \le Z_n\right] \le t^* \beta - K(t^*).$$

Finally, since K is convex and smooth, and since $K'(t^*) = \beta$, then t^* is the maximizer of $t\beta - K(t)$, and thus $t^*\beta - K(t^*) = K^\star(\beta)$. We have thus shown that

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}\left[\beta n \le Z_n\right] \le K^\star(\beta).$$

□

3 Recurrence and transience

3.1 Definitions and basic observations

Given μ, we say that the μ-random walk is *recurrent* if $(A_n)_n$ i.o. occurs almost surely, where $A_n = \{Z_n = 0\}$. That is, if the random walk almost surely returns to zero infinitely many times.

We say that the μ-random walk is *transient* if the probability of $(A_n)_n$i.o. is zero, i.e., the random walk almost surely visits zero a finite number of times.

Claim 3.1. *Every random walk is either transient or recurrent.*

The proof of this claim will use the fact that a random walk on $\mathbb{Z}$ is a *Markov chain*.

Proof of Claim 3.1. Denote by H_0 the event that there exists some $n > 0$ such that $Z_n = 0$. I.e., that the random walk returns to 0. Let $p = \mathbb{P}[H_0]$.

By the Markov property, conditioned on $Z_k = 0$, the probability that there is some $n > k$ such that $Z_n = 0$ is also p. It follows that if $p = 1$ the random walk is recurrent. And if $p < 1$ then the number of visits to 0 has geometric distribution with parameter p, in which case the number of visits is almost surely finite, and the random walk is transient. □

The next lemma gives useful equivalent conditions to recurrence.

Lemma 3.2. *Consider any μ-random walk. The following are equivalent.*

1. *The random walk is recurrent.*
2. *There is some $x \in \mathbb{Z}$ that the random walk almost surely hits infinitely many times.*
3. *The random walk hits every $x \in \mathbb{Z}$ almost surely.*

Note that this lemma holds much more generally, for irreducible Markov chains on countably infinite state spaces.

3.2 Random walks with a drift

As in the previous section, denote $\alpha := \mathbb{E}[X] = \sum_{x \in \mathbb{Z}} x\mu(x)$.

Claim 3.3. *A random walk on $\mathbb{Z}$ with non-zero drift is transient.*

Proof. Suppose w.l.o.g. that $\alpha > 0$. By the strong law of large numbers, $\lim_n \frac{1}{n} Z_n = \alpha > 0$. Hence $\lim_n Z_n = \infty$, and it is impossible that $Z_n = 0$ infinitely often. □

3.3 Recurrence of the simple random walk on $\mathbb{Z}$

Recall that the simple μ-random walk is given by $\mu(-1) = \mu(1) = 1/2$.

Theorem 3.4 (Pólya)**.** *The simple random walk on $\mathbb{Z}$ is recurrent.*

We will prove this in a number of ways.

First proof of Theorem 3.4. Note that $\mathbb{P}[Z_{2n+1} = 0] = 0$ and that

$$\mathbb{P}[Z_{2n} = 0] = 2^{-2n}\binom{2n}{n}.$$

By Stirling

$$\binom{2n}{n} \geq \frac{2^{2n-1}}{\sqrt{n}},$$

and so

$$\mathbb{P}[Z_{2n} = 0] \geq \frac{1}{2\sqrt{n}}.$$

The expected number of visits to 0 is thus

$$\sum_n \mathbb{P}[Z_{2n} = 0] \geq \sum_{n=1}^{\infty} \frac{1}{2\sqrt{n}} = \infty.$$

As noted in the proof of Claim 3.1, the number of returns is geometric if the random walk is transient, and hence has finite expectation. Thus this random walk is recurrent. $\square$

3.4 Superharmonic functions

For the second proof of Theorem 3.4, we introduce the notion of a *μ-superharmonic function*. A function $\varphi \colon \mathbb{Z} \to \mathbb{R}$ is mu-superharmonic if for every $x \in \mathbb{Z}$

$$\varphi(x) \geq \sum_{y \in \mathbb{Z}} \varphi(x+y)\mu(y). \tag{3.1}$$

That is, $\varphi(x)$ is larger than the average of φ around x, where we take averages using μ.

Given $x \in \mathbb{Z}$, the process $(x + Z_1, x + Z_2, \ldots)$ is the μ-random walk starting at x. We define $Z_0 = 0$. Denote by H_x the event that there exists some $n \geq 0$ such that $x + Z_n = 0$. I.e., that the random walk that starts at x eventually hits 0:

$$H_x = \{\exists n \geq 0 \text{ s.t. } x + Z_n = 0\} = \bigcup_{n=0}^{\infty} \{x + Z_n = 0\}.$$

Obviously, this is the same event as $Z_n = -x$ for some $n \geq 0$.

Define $\varphi\colon \mathbb{Z} \to \mathbb{R}$ by $\varphi(x) = \mathbb{P}[H_x]$, so that $\varphi(x)$ is the probability that the random walk starting at x eventually hits 0. We claim that φ is μ-superharmonic. Indeed,

$$\begin{aligned}
\varphi(x) &= \mathbb{P}[H_x] \\
&= \mathbb{P}[x + Z_1 = 0, H_x] + \mathbb{P}[x + Z_1 \neq 0, H_x] \\
&= \mathbb{P}[x + Z_1 = 0] + \sum_{y \neq 0} \mathbb{P}[x + Z_1 = y, H_x] \\
&= \mathbb{P}[x + Z_1 = 0] + \sum_{y \neq 0} \mathbb{P}[H_x | x + Z_1 = y]\mathbb{P}[x + Z_1 = y].
\end{aligned}$$

Now, $Z_1 = X_1$, and the distribution of X_1 is μ, so

$$\varphi(x) = \mu(-x) + \sum_{y \neq 0} \mathbb{P}[H_x | x + X_1 = y]\mu(y - x).$$

Writing out the definition of H_x we get

$$\mathbb{P}[H_x | x + X_1 = y] = \mathbb{P}[\exists n \geq 0 \text{ s.t. } x + X_1 + X_2 + \cdots + X_n = 0 | x + X_1 = y].$$

Since we are conditioning on $X_1 = y - x$ we can substitute that to arrive at

$$\mathbb{P}[H_x | x + X_1 = y] = \mathbb{P}[\exists n \geq 0 \text{ s.t. } y + X_2 + \cdots + X_n = 0 | x + X_1 = y].$$

But X_1 is independent of $(X_2, X_3, \ldots)$, so we can remove the conditioning. And we can replace $X_2 + \cdots + X_n$ by $X_1 + \cdots + X_n$, since $(X_1, X_2, \ldots)$ and $(X_2, X_3, \ldots)$ have the same distribution; we are in essence using the Markov property of the random walk here. So we get

$$\mathbb{P}[H_x | x + X_1 = y] = \mathbb{P}\left[H_y\right].$$

Hence

$$\begin{aligned}
\varphi(x) &= \mu(-x) + \sum_{y \neq 0} \mathbb{P}\left[H_y\right]\mu(y - x) \\
&\geq \sum_{y \in \mathbb{Z}} \mathbb{P}\left[H_y\right]\mu(y - x) \\
&= \sum_{y \in \mathbb{Z}} \varphi(y)\mu(y - x).
\end{aligned}$$

Finally, a change of variables gives

$$\varphi(x) \geq \sum_{y \in \mathbb{Z}} \varphi(x + y)\mu(y).$$

We have thus shown that φ is μ-superharmonic. Note that it is also non-negative.

Lemma 3.5. *Let $\mu(-1) = \mu(1) = 1/2$. Then every non-negative μ-superharmonic $\varphi\colon \mathbb{Z} \to \mathbb{R}$ is constant.*

Proof. Since φ is μ-superharmonic,

$$\varphi(x) \geq \frac{1}{2}\varphi(x-1) + \frac{1}{2}\varphi(x+1).$$

Rearranging, we get that

$$\varphi(x) - \varphi(x-1) \geq \varphi(x+1) - \varphi(x).$$

Denote $\varphi'(x) = \varphi(x) - \varphi(x-1)$. Then we have shown that

$$\varphi'(x) \geq \varphi'(x+1),$$

so that φ' is non-increasing.

If $\varphi' = 0$ then φ is constant and we are done. Otherwise, suppose $\varphi'(x) < -\varepsilon$ for some x. Then $\varphi'(x+n) \leq -\varepsilon$ for all $n \geq 0$. Hence $\varphi(x+n) \leq \varphi(x) + n\varepsilon$, and $\varphi(x)$ is negative for x large enough. An analogues argument shows that $\varphi(-x)$ is negative for x large enough if $\varphi'(x) > 0$ for some x.

□

Second proof of Theorem 3.4. Define $\varphi(x) = \mathbb{P}[H_x]$ as above. We have shown that $\varphi = p$. Since $\varphi(0) = 1$ by definition, it follows that $p = 1$. Applying the Markov property again, we conclude that $\mathbb{P}[\exists n \geq k \text{ s.t. } Z_n = 0] = 1$ for all k, and thus the random walk is recurrent. □

The argument above in fact is one direction of a more general fact relating superharmonic functions and recurrence.

Theorem 3.6. *For any μ-random walk on $\mathbb{Z}$ the following are equivalent.*

1. *The walk is transient.*
2. *There exist non-constant non-negative μ-superharmonic functions on $\mathbb{Z}$.*

Indeed, this again holds much more generally, for irreducible Markov chains on countably infinite state spaces.

To prove this theorem we will need to recall the notions of a *supermartingale* and a *stopping time*. Let $(Y_1, Y_2, \ldots)$ be a sequence of random variables, let $\mathscr{F}_n = \sigma(Y_1, \ldots, Y_n)$ and let $\mathscr{F}_\infty = \sigma(Y_1, Y_2, \ldots)$. A sequence of real random variables $(W_0, W_1, W_2, \ldots)$ is a supermartingale with respect to $(\mathscr{F}_n)_n$ if

1. W_n is $\mathscr{F}_n$-measurable.
2. $\mathbb{E}[W_{n+1} | \mathscr{F}_n] \leq W_n$.

A natural example is when Y_n is the outcome of the roulette at time n, and W_n is the amount of money gained by a gambler who plays this roulette using some fixed deterministic strategy (e.g., a dollar on red at even n and three dollars on black at odd n). The first condition states that the amount of money the gambler has is determined by the outcomes of the roulette, and the second states that given what the gambler has at time n, she expects to have (weakly) less at time $n+1$.

The key observation relating supermartingales to random walks is the following observation.

Claim 3.7. *Let φ be μ-superharmonic. Then $W_n = \varphi(Z_n)$ is a supermartingale with respect to $(\sigma(Z_1,\ldots,Z_n))_n$.*

A stopping time T is a $\mathscr{F}_\infty$-measurable random variable taking values in $\{1,2,\ldots,\infty\}$ such that for each n the event $\{T = n\}$ is $\mathscr{F}_n$-measurable. An example is the first time n such that the gambler has $17n$ dollars in their balance. More generally, T is a stopping time if it is equal to the minimum time n is which the condition A_n is met (formally, the event A_n occurs), where each A_n is $\mathscr{F}_n$-measurable, i.e., determined by $(Y_1,\ldots,Y_n)$. An important result due to Doob is the optional stopping time theorem:

Theorem 3.8 (Doob)**.** *Suppose $(W_0,W_1,W_2,\ldots)$ is a non-negative supermartingale, and let T be a finite stopping time. Then $\mathbb{E}[W_T] \le \mathbb{E}[W_0]$.*

For our gambler, this means that if she walks in with 100 dollars and has some stopping rule for leaving (and cannot go into debt), the expected amount of money she will have at the time of leaving is at most 100.

Proof of Theorem 3.6. The direction 1 implies 2 is proved using $\varphi(x) = \mathbb{P}[H_x]$ as above. For the other direction, suppose the μ-random walk is recurrent, and let φ be non-negative and μ-superharmonic. For $x, y \in \mathbb{Z}$ let T be the stopping time given by the first hitting time to y of the μ-random walk starting at x:

$$T = \min\{n : x + Z_n = y\}.$$

By recurrence and Lemma 3.2 T is finite almost surely. Let $W_n = \varphi(x + Z_n)$. By the optional stopping time theorem, $\mathbb{E}[W_T] \le \mathbb{E}[W_0]$. Since the l.h.s. of the equality is $\varphi(y)$ and the r.h.s. is $\varphi(x)$ we have that $\varphi(x) = \varphi(y)$. □

3.5 Harmonic functions

Claim 3.9. *For any random walk on $\mathbb{Z}$, the probability that $\{Z_0,Z_1,Z_2,\ldots\}$ is a finite subset of $\mathbb{Z}$ is zero.*

This claim likewise holds much more generally, for irreducible Markov chains on countably infinite state spaces.

Let μ be the simple random walk on $\mathbb{Z}$. Fix some $M \in \mathbb{Z}$, $M > 0$. Note that $\mathbb{P}[\exists n \text{ s.t. } Z_n \in \{-1, M\}] = 1$, by Claim 3.9, since otherwise the random walk would be confined in $\{0,\ldots,M-1\}$.

Let A_x be the event that $x + Z_n = -1$ before $x + Z_n = M$. Let $\varphi\colon \{-1,\ldots,M\} \to \mathbb{R}$ be given by $\varphi(x) = \mathbb{P}[A_x]$ for $x \in \{0, M-1\}$, $\varphi(-1) = 1$ and $\varphi(M) = 0$. Then for $x \in \{0,\ldots,M-1\}$

$$\begin{aligned}\varphi(x) &= \mathbb{P}[A_x] \\ &= \mathbb{P}[A_x | x + Z_1 = x+1]\mathbb{P}[x + Z_1 = x+1] + \mathbb{P}[A_x | x + Z_1 = x-1]\mathbb{P}[x + Z_1 = x-1] \\ &= \varphi(x+1)\mu(1) + \varphi(x-1)\mu(-1),\end{aligned}$$

where the penultimate equality uses the Markov property, as in the previous section, and our definitions at $x = -1$ and $x = M$. We thus have that for $x \in \{0, M-1\}$

$$\varphi(x) = \sum_y \varphi(x+y)\mu(y).$$

We say that φ is *harmonic* on $\{0, M-1\}$.

It is easy to see that the only function that satisfies this equality is linear on $\{-1, M\}$, and hence we have shown that

$$\varphi(x) = \frac{M-x}{M+1}.$$

In particular, the probability that Z_n hits -1 before it hits M is $M/(M+1)$. Now, the event that Z_n *never* reaches -1 is the same as the event that it reaches every $M > 0$ before it reaches -1, by Claim 3.9. Hence this occurs with probability at most $1/(M+1)$ for any 0, and the random walk hits -1 almost surely. By symmetry, the random walk also hits $+1$ almost surely. Hence it visits 0 again almost surely (since it has to travel either from -1 to $+1$ or from $+1$ to -1), and so it is recurrent.

3.6 Recurrence of symmetric random walks on $\mathbb{Z}$

We say that μ is *symmetric* if $\mu(x) = \mu(-x)$ for all $x \in \mathbb{Z}$.

Theorem 3.10. *The μ-random walk on $\mathbb{Z}$ is recurrent for all symmetric μ.*

To prove this theorem we will recall the tail sigma-algebra and the Kolmogorov 0-1 law. Let $(Y_1, Y_2, \ldots)$ be a sequence of random variables. Denote $\mathcal{T}_n = \sigma(Y_n, Y_{n+1}, \ldots)$. That is, a random variable W is $\mathcal{T}_n$-measurable if there is some f such that $W = f(Y_n, Y_{n+1}, \ldots)$. The tail sigma-algebra $\mathcal{T}$ is $\mathcal{T} = \cap_n \mathcal{T}_n$. That is, W is $\mathcal{T}$-measurable—in which case we call it a tail random variable—if for every n there is an f_n such that $W = f_n(Y_{n+1}, Y_{n+2}, \ldots)$. An example is $W = \limsup_n Y_n$. Kolmogorov's 0-1 law states that if $(Y_1, Y_2, \ldots)$ are independent then $\mathcal{T}$ is trivial: every tail random variable is constant.

Proof of Theorem 3.10. Let μ be symmetric and suppose $(Z_1, Z_2, \ldots)$ is transient. Then by Lemma 3.2 Z_n only visits each interval $[-M, M]$ finitely many times, and so $\lim_n |Z_n| = \infty$. If we consider M such that μ is supported on $[-M, M]$, it follows that $\lim_n \operatorname{sgn}(Z_n)$ exists, or that Z_n is eventually either even or odd. Hence $W := \lim_n Z_n$ exists and is in $\{+\infty, -\infty\}$.

Since μ is symmetric $\mathbb{P}[W = +\infty] = \mathbb{P}[W = -\infty] = 1/2$. The formal proof of this is via a *coupling argument*. Let $\check{X}_n = -X_n$. Then, by the symmetry of μ, $(\check{X}_1, \check{X}_2, \ldots)$ is also i.i.d. μ. Hence, if we define $\check{Z}_n = \check{X}_1 + \cdots + \check{X}_n = -Z_n$, $(\check{Z}_1, \check{Z}_2, \ldots)$ has the same distribution as $(Z_1, Z_2, \ldots)$. But $\lim \check{Z}_n = -\lim Z_n$, and so

$$\mathbb{P}\left[\lim_n Z_n = -\infty\right] = \mathbb{P}\left[\lim_n \check{Z}_n = +\infty\right] = \mathbb{P}\left[\lim_n Z_n = +\infty\right],$$

and we have that $\mathbb{P}[\lim_n Z_n = \infty] = 1/2$.

Finally, W is a tail event of $(X_1, X_2, \ldots)$, since

$$W_n = \sum_{k=n}^{\infty} X_k$$

is $\mathcal{T}_n$-measurable and equal to W. Since $(X_1, X_2, \ldots)$ is i.i.d., W must be constant by Kolmogorov's 0-1 law, and we have reached a contradiction. □

Corollary 3.11. *Let $(Z_1, Z_2, \ldots)$ be a symmetric random walk on $\mathbb{Z}$. Then $\sum_n \mathbb{P}[Z_n = 0] = \infty$.*

Proof. If $\sum_n \mathbb{P}[Z_n = 0] < \infty$ then by Borel-Cantelli the random walk is transient, in contradiction to Theorem 3.10. □

3.7 Recurrence of zero drift random walks on $\mathbb{Z}$

Given a transient random walk $(Z_1, Z_2, \ldots)$ on $\mathbb{Z}$, denote by V_x the number of visits to x

$$V_x = |\{n \geq 0 : Z_n = x\}|,$$

and let

$$v(x) = \mathbb{E}[V_x] = \sum_{n=0}^{\infty} \mathbb{P}[Z_n = x]$$

denote the expected number of visits to x. As discussed above, transitivity guarantees that $v(x)$ is finite for all x.

Claim 3.12. *The maximum of $v(x)$ is attained at* 0.

Proof. Let $H_x = \{\exists n \geq 0 \text{ s.t. } Z_n = x\}$ be the event that the random walk hits x. Then

$$v(x) = \mathbb{E}[V_x] = \mathbb{E}[V_x | H_x]\mathbb{P}[H_x] + \mathbb{E}\left[V_x \middle| H_x^c\right](1 - \mathbb{P}[H_x]).$$

We know that $\mathbb{P}[H_x] \geq 1$. Since $V_x = 0$ conditioned on H_x^c, we have that for $x \neq 0$

$$v(x) \leq \mathbb{E}[V_x | H_x].$$

But by the Markov property the r.h.s. is exactly equal to $v(0)$. □

Theorem 3.13. *A random walk on $\mathbb{Z}$ with zero drift is recurrent.*

Proof. Suppose $(Z_1, Z_2, \ldots)$ is random walk on $\mathbb{Z}$ with zero drift and $|X| \le M$ almost surely. Hence $\mathbb{E}\left[X^2 \le nM^2\right]$ and by Markov's inequality

$$\mathbb{P}[|X| \ge x] \le \frac{nM^2}{x^2}.$$

In particular, if we choose $x = n^{0.6}$ we get

$$\mathbb{P}\left[|Z_n| \ge Mn^{0.6}\right] \le \frac{nM^2}{M^2 n^{1.2}} = \frac{1}{n^{0.2}}.$$

Since $\mathbb{P}\left[|Z_n| \le Mn^{0.6}\right] > 0$ for all n, it follows that there exists some $c > 0$ such that

$$\mathbb{P}\left[|Z_n| \le Mn^{0.6}\right] \ge c \tag{3.2}$$

for all n.

Denote $N(n) := Mn^{0.6}$. Then for all n

$$\sum_{x=-N(n)}^{N(n)} \mathbb{P}[Z_n = x] \ge c.$$

We claim that this implies that there is some $x \in \mathbb{Z}$ such that $\sum_n \mathbb{P}[Z_n = x] = \infty$, which implies that the random walk is recurrent. Suppose not, and recall the notation $v(x) = \sum_{n \ge 0} \mathbb{P}[Z_n = x]$. Then for every $n \ge 0$,

$$\begin{aligned}\sum_{x=-N(n)}^{N(n)} v(x) &\ge \sum_{x=-N(n)}^{N(n)} \sum_{k=0}^{n} \mathbb{P}[Z_k = x] \\ &\ge \sum_{k=0}^{n} \sum_{x=-N(n)}^{N(n)} \mathbb{P}[Z_k = x] \\ &\ge cn.\end{aligned}$$

By Claim 3.12 $v(x) \le v(0)$, and so we have that

$$\sum_{x=-N(n)}^{N(n)} v(0) \ge cn$$

for all n, which is impossible, since the l.h.s. is at most $(2Mn^{0.6} + 1)v(0)$. □

4 Random walks on $\mathbb{Z}^d$

Let μ be a probability measures on $\mathbb{Z}^d$ for some $d \geq 1$, let $(X_1, X_2, \ldots)$ be i.i.d. with law μ, and let $Z_n = X_1 + \cdots + X_n$. As before, we assume that it is finitely supported and that it is *non-degenerate*: for every $z \in \mathbb{Z}^d$ there exists $n \geq 1$ such that $\mathbb{P}[Z_n = z] > 0$.

4.1 Recurrence and transience

We say that μ is symmetric if $\mu(-x) = \mu(x)$ for all $x \in \mathbb{Z}^d$. We say that μ is a *product measure* if there exists $\mu_1, \ldots, \mu_d$, all probability measures on $\mathbb{Z}$, such that $\mu(z_1, \ldots, z_d) = \mu_1(z_1) \cdots \mu_d(z_d)$. We then write $\mu = \mu_1 \times \cdots \times \mu_d$.

Theorem 4.1 (Pólya)**.** *Let $\mu_1 = \mu_2 = \cdots = \mu_d$ all equal the simple random walk on $\mathbb{Z}$, and let $\mu = \mu_1 \times \cdots \times \mu_d$. Then*

1. *If $d \leq 2$ then the μ-random walk is recurrent.*
2. *If $d \geq 3$ then the μ-random walk is transient.*

Proof. A standard bound on $\binom{2n}{n}$ is

$$\frac{4^n}{\sqrt{\pi(n+\frac{1}{2})}} \leq \binom{2n}{n} \leq \frac{4^n}{\sqrt{\pi n}}.$$

Hence, as in the first proof of Theorem 3.4,

$$\left(\frac{1}{\sqrt{\pi(n+\frac{1}{2})}}\right)^d \leq \mathbb{P}[Z_{2n} = 0] \leq \left(\frac{1}{\sqrt{\pi n}}\right)^d.$$

For odd n, $\mathbb{P}[Z_n = 0] = 0$. Hence, for $d \leq 2$, $\sum_n \mathbb{P}[Z_n = 0]$ diverges and the random walk is recurrent, while for $d \geq 3$ it converges and the random walk is transient □

4.2 A Hoeffding bound for $\mathbb{Z}^d$

Recall that the Hoeffding bound (Theorem 1.5) says that on $\mathbb{Z}$, if $|X| \leq M$ almost surely and $\beta > \mathbb{E}[X]$ then

$$\mathbb{P}\left[Z_n \geq \beta n\right] \leq \mathrm{e}^{-\frac{\beta^2}{2M^2}\cdot n}.$$

Suppose $\mathbb{E}[X] = 0$. Then for any $x \in \mathbb{Z}$ it follows that (by a change of variable $x = \beta n$)

$$\mathbb{P}[Z_n \geq x] \leq \mathrm{e}^{-\frac{1}{2M^2}\frac{|x|^2}{n}}.$$

In particular, we will be interested in the weaker form

$$\mathbb{P}[Z_n = x] \le \mathrm{e}^{-\frac{1}{2M^2}\frac{|x|^2}{n}}. \tag{4.1}$$

Now let $(Z_1, Z_2, \dots)$ be a μ-random walk on $\mathbb{Z}^d$ with $\mathbb{E}[Z_1] = 0$. We will denote the L^2-norm on $\mathbb{Z}^d$ by $|\cdot|$, and assume that the support of μ is contained in the ball of radius M. Choose $x \in \mathbb{Z}^d$. We would like to prove an inequality of the form (4.1).

Let $\pi\colon \mathbb{Z}^d \to \mathbb{Z}$ be the inner product with x: $\pi(z) = \sum_{i=1}^d x_i z_i$. Let $\hat{X}_n = \pi(X_n)$ and $\hat{Z}_n = \pi(Z_n)$. Note that $\hat{Z}_n = \hat{X}_1 + \cdots + \hat{X}_n$, and so $(\hat{Z}_1, \hat{Z}_2, \dots)$ is a random walk on $\mathbb{Z}$. The step distribution of this random walk is denoted $\pi_*\mu$ and called the *push-forward measure*:

$$[\pi_*\mu](z) = \mu(\pi^{-1}(z)) = \mu\Big(\Big\{x \in \mathbb{Z}^d \,:\, \pi(x) = z\Big\}\Big).$$

Note that $\pi_*\mu$ might not be non-degenerate, as its support might be contained in some subgroup $m\mathbb{Z}$ (e.g., if $x = (2,0)$ and $m = 2$). But on this subgroup it will be non-degenerate, and so everything we know will still go through (formally, we can define $\pi(z) = \frac{1}{m}\sum_{i=1}^d x_i z_i$). Note also that since μ has zero expectation then so does $\pi_*\mu$.

Since $|X_n| \le M$, and since $|\pi(z)| \le |x||z|$, $|\hat{X}_n| \le M|x|$. Hence, by (4.1) we have that

$$\mathbb{P}\left[\hat{Z}_n = \pi(x)\right] \le \mathrm{e}^{-\frac{1}{2M^2|x|^2}\frac{|\pi(x)|^2}{n}}.$$

Since $\pi(x) = |x|^2$ this becomes

$$\mathbb{P}\left[\hat{Z}_n = \pi(x)\right] \le \mathrm{e}^{-\frac{1}{2M^2}\frac{|x|^2}{n}}.$$

Finally, since the event $Z_n = x$ implies $\hat{Z}_n = \pi(x)$, this in implies the following Hoeffding bound for $\mathbb{Z}^d$.

Theorem 4.2. *Let $(Z_1, Z_2, \dots)$ be a μ-random walk on $\mathbb{Z}^d$ where μ is symmetric and supported on the ball of radius M. Then*

$$\mathbb{P}[Z_n = x] \le \mathrm{e}^{-\frac{1}{2M^2}\frac{|x|^2}{n}}.$$

5 Random walks on the free group

5.1 The free group

Let $S = \{a, b, a^{-1}, b^{-1}\}$ be abstract "symbols". A *(reduced) word* is a finite sequence of symbols $s_1 s_2 \cdots s_n$, with each $s_i \in S$ (e.g., $g = a^{-1}bbab^{-1}a^{-1}$) that does not include adjacent occurrences of a and a^{-1}, or of b and b^{-1}. We denote the empty word by e. We can define a concatenation operation $(g,h) \mapsto gh$ on reduced words by concatenating them, and then iteratively removing any disallowed occurrences.

The *free group with two generators* $\mathbb{F}_2$ is the set of reduced words, together with the concatenation operation. Note that our notation for the symbols is consistent with inverses in the group: a^{-1} is the inverse of a, since their product results in the empty word, which is the identity element. More generally, given a word $g = s_1 \cdots s_n$, its inverse is given by $g^{-1} = s_n^{-1} \cdots s_1^{-1}$.

An important way to think of the free group is via its *Cayley graph*. The nodes of the graph are the elements of the group. Its directed edges are labeled, and there is an edge (g,h) with label $s \in S$ if $h = gs$ (in which case there is an edge (h,g) with label s^{-1}). This graph is the 4-regular tree: the (unique up to isomorphism) graph in which all nodes have degree 4 and there are no cycles.

This graph is *vertex transitive*. Informally, it looks the same from the point of view of each vertex. Formally, the balls of radius r around each vertex are all isomorphic graphs. Note that the number of elements within distance r of a given point in this graph is $4 \cdot 3^{r-1}$, and in particular is exponential in r. In $\mathbb{Z}^d$, balls only grow polynomially.

We define a norm on $\mathbb{F}_2$ by setting $|g|$ to be the minimal number of generators whose product is equal to g. Equivalently, this is the distance between e and g in the Cayley graph. The ball of radius r in the Cayley graph is $\{g \in \mathbb{F}_2 : |g| \le r\}$.

Let μ be a probability measure on $\mathbb{F}_2$. The μ random walk on $\mathbb{F}_2$ is defined as follows: $(X_1, X_2, \ldots)$ are i.i.d. μ, and $Z_n = X_1 X_2 \cdots X_n$. We set $Z_0 = e$. As on $\mathbb{Z}^d$, we will restrict ourselves to finitely supported μ, and will assume that μ is non-degenerate, so that for all $g \in \mathbb{F}_2$ there is an n such that $\mathbb{P}[Z_n = g] > 0$.

5.2 Transience of the simple random walk

The simple random walk on $\mathbb{F}_2$ is given by $\mu(a) = \mu(a^{-1}) = \mu(b) = \mu(b^{-1}) = 1/4$. It will be useful to think of this random walk as a random walk on the 4-regular tree.

A function $\varphi \colon \mathbb{F}_2 \to \mathbb{R}$ is μ-superharmonic if for all $g \in \mathbb{F}_2$

$$\varphi(g) \ge \sum_{h \in \mathbb{F}_2} \varphi(gh)\mu(h).$$

As on $\mathbb{Z}$, this implies that $\varphi(Z_n)$ is a supermartingale. Thus the same proof as for $\mathbb{Z}$ yields the following claim.

Theorem 5.1. *For any μ-random walk on $\mathbb{F}_2$ the following are equivalent.*

1. *The walk is transient.*

2. *There exist non-constant non-negative μ-superharmonic functions on $\mathbb{Z}$.*

Corollary 5.2. *The simple random walk on $\mathbb{F}_2$ is transient.*

Proof. Let $\varphi(g) = 3^{-|g|}$. Then clearly the superhamonicity condition is satisfied at e, since that is where φ attains its maximum. Elsewhere, for $|g| = r$,

$$\sum_{h \in \mathbb{F}_2} \varphi(gh)\mu(h) = 3^{-(r-1)}\frac{1}{4} + 3 \cdot 3^{-(r+1)}\frac{1}{4} = 3^r\left(3^{+1}\frac{1}{4} + 3 \cdot 3^{-1}\frac{1}{4}\right) = 3^{-r} = \varphi(g).$$

□

5.3 Hitting probabilities of the simple random walk

Given $g \in \mathbb{F}_2$, denote by $H_g = \{\exists n \geq 0 : Z_n = g\}$ the event that the random walk eventually hits g. By the symmetry of the random walk, there is some p so that $p = \mathbb{P}[H_s]$ for all $s \in S$.

$$\begin{aligned} p &= \mathbb{P}[H_a] \\ &= \sum_{s \in S} \mathbb{P}[H_a|Z_1 = s]\mathbb{P}[Z_1 = s] \\ &= \frac{1}{4}\sum_{s \in S} \mathbb{P}[H_a|Z_1 = s] \\ &= \frac{1}{4} + \frac{1}{4}\sum_{s \in S \setminus \{a\}} \mathbb{P}[H_a|Z_1 = s]. \end{aligned}$$

By the Markov property, for $s \neq a$,

$$\begin{aligned} \mathbb{P}[H_a|Z_1 = s] &= \mathbb{P}[\exists n \geq 0 : X_1 \cdots X_n = a|X_1 = s] \\ &= \mathbb{P}[\exists n \geq 0 : sX_2 \cdots X_n = a|X_1 = s] \\ &= \mathbb{P}\left[\exists n \geq 0 : X_2 \cdots X_n = s^{-1}a\middle|X_1 = s\right] \\ &= \mathbb{P}\left[H_{s^{-1}a}\right]. \end{aligned}$$

Now, because the Cayley graph is a tree, the random walk must visit s^{-1} before visiting a. So

$$\mathbb{P}\left[H_{s^{-1}a}\right] = \mathbb{P}\left[H_{s^{-1}a}, H_{s^{-1}}\right] = \mathbb{P}\left[H_{s^{-1}a}\middle|H_{s^{-1}}\right]\mathbb{P}\left[H_{s^{-1}}\right] = \mathbb{P}\left[H_{s^{-1}a}\middle|H_{s^{-1}}\right] \cdot p.$$

Again by the Markov property and symmetry, $\mathbb{P}\left[H_{s^{-1}a}\middle|H_{s^{-1}}\right] = p$. Hence we have that

$$p = \frac{1}{4} + \frac{3}{4}p^2,$$

so that $p = 1/3$, since by transience $p \neq 1$. Indeed, a similar calculation shows more generally that that $\mathbb{P}\left[H_g\right] = 3^{-|g|}$.

5.4 Tail events of the simple random walk

Since the random walk is transient, There is a.s. a finite random N such that $Z_N \in S$ and $Z_{N+n} \neq e$ for all $n \geq 0$. For $s \in S$, denote by $F_s \subset \mathbb{F}_2$ the set of words that begin with s. Then $Z_{N+n} \in F_{Z_N}$. By the symmetry of the random walk,

$$\mathbb{P}[Z_n \in F_a \text{ for all } n \text{ large enough}] = \frac{1}{4}.$$

For any subset $F \subset \mathbb{F}_2$, the event Note that the event $E_F := \{Z_n \in F \text{ for all } n \text{ large enough}\}$ is a tail event of the process $(Z_1, Z_2, \ldots)$. Moreover, it is a *shift-invariant* event. A random variable W is measurable with respect to the shift-invariant sigma-algebra if there is some f such that

$$W = f(Z_1, Z_2, \ldots) = f(Z_2, Z_3, \ldots).$$

Note that this implies that W is also a tail event with respect to $(Z_1, Z_2, \ldots)$. We have thus proved the following claim.

Claim 5.3. *The simple random walk on $\mathbb{F}_2$ admits a non-constant shift-invariant random variable.*

5.5 Distance from the origin of the simple random walk

Denote $L_n = |Z_n|$. Note that conditioned on $Z_{n-1} = e$, $L_n = L_{n-1} + 1 = 1$. And for any $g \neq e$

$$\mathbb{P}[L_n = L_{n-1} + 1 | Z_n = g] = \frac{3}{4}$$
$$\mathbb{P}[L_n = L_{n-1} - 1 | Z_n = g] = \frac{1}{4}.$$

Define the process $(\tilde{X}_1, \tilde{X}_2)$ on $\mathbb{Z}$ by $\tilde{X}_0 = 0$ and

$$\tilde{X}_n = \begin{cases} L_n - L_{n-1} & \text{if } Z_n \neq e \\ Y_n & \text{otherwise,} \end{cases}$$

where Y_n are independent with $\mathbb{P}[Y_n = +1] = 3/4$ and $\mathbb{P}[Y_n = -1] = 1/4$. It can be shown that $(\tilde{X}_1, \tilde{X}_2, \ldots)$ are i.i.d. and so $\tilde{Z}_n = \tilde{X}_1 + \cdots + \tilde{X}_n$ is a random walk on $\mathbb{Z}$, with drift 1/2. Thus

$$\lim_n \frac{1}{n} \tilde{Z}_n = \frac{1}{2}$$

by the strong law of large numbers. By transience, the event $\{Z_n = e\}$ happens only finitely often, and so $\tilde{Z}_n$ and L_n never differ by more than a (random) constant: $\max_n |L_n - \tilde{Z}_n|$ is finite almost surely. Hence

$$\lim_n \frac{1}{n} L_n = \lim_n \frac{1}{n} \tilde{Z}_n + \frac{1}{n}(L_n - \tilde{Z}_n) = \lim_n \frac{1}{n} \tilde{Z}_n = \frac{1}{2}.$$

Thus $L_n = |Z_n|$ concentrates around $n/2$. Since $\tilde{X}_n \leq L_n - L_{n-1}$, $\tilde{Z}_n \leq L_n$. Hence, by the Hoeffding bound,

$$\mathbb{P}[Z_n = e] = \mathbb{P}[L_n = 0] = \mathbb{P}[L_n \leq 0] \leq \mathbb{P}\left[\tilde{Z}_n \leq 0\right] \leq \mathrm{e}^{-n/8}, \tag{5.1}$$

so that the probability of return to the origin decays exponentially with n.

6 The lamplighter group

6.1 Lamplighters

The *lamplighter* is a person located at some point $x \in \mathbb{Z}$. At each $z \in \mathbb{Z}$ there is a lamp that is either off or on. We imagine that initially all lamps are off. The lamplighter has three things that she can do:

1. Move one step to the right.

2. Move one step to the left.

3. Flip the state of the lamp at her current location.

Thus, a sequence of actions of the lamplighter is a word in the alphabet $S = \{a, a^{-1}, b\}$, corresponding to the three options above. After executing such a sequence, we can describe the current state by a pair (f, x), where $x \in \mathbb{Z}$ is the location of the lamplighter, and finitely supported $f : \mathbb{Z} \to \mathbb{Z}_2$ is the indicator of the lamps that are on. We denote by $\bigoplus_{\mathbb{Z}} \mathbb{Z}_2$ the set of such finitely supported f, which we call lamp configurations. Denote by $\alpha \colon \bigoplus_{\mathbb{Z}} \mathbb{Z}_2 \to \bigoplus_{\mathbb{Z}} \mathbb{Z}_2$ the shift operation on configurations given by $[\alpha f](x) = f(x-1)$.

Suppose that g_1 culminates in (f_1, x_1) and that g_2 culminates in (f_2, x_2). Then the state of the system when executing g_1 followed by g_2 will be

$$g_1 g_2 = (f_1, x_1)(f_2, x_2) = (f_1 + \alpha^{x_1} f_2, x_1 + x_2).$$

It is easy to see that this operation is associative and invertible, and so we have defined a group, which is denoted by $\bigoplus_{\mathbb{Z}} \mathbb{Z}_2 \rtimes \mathbb{Z}$. This is also sometimes written as $\mathbb{Z}_2 \wr \mathbb{Z}$. Using this notation our generating set is

$$S = \{a, a^{-1}, b\} = \{(1,0), (-1,0), \delta_0\},$$

where $\delta_0 \in \bigoplus_{\mathbb{Z}} \mathbb{Z}_2$ is the indicator of 0.

Another way to think of this group is as follows: $f \in \bigoplus_{\mathbb{Z}} \mathbb{Z}_2$ is an instruction to the lamplighter located at x to flip the lamps at all $x + z$ such that $f(z) = 1$. The group is defined by $f^2 = 0$ for all f, $f_1 f_2 = f_2 f_1$ and $af = (\alpha f)a$.

Given $g \in \mathbb{Z}_2 \wr \mathbb{Z}$ we denote by $|g|$ the minimum number of generators in S whose product is equal to g. We denote by B_r the set $\{g : |g| \le r\}$. It is easy to see that every f with support contained in $\{0, 1, \dots, r/3\}$ is in B_r, and thus B_r is of size at least $2^{r/3}$, and in particular grows exponentially with r, like the free group and unlike $\mathbb{Z}^d$.

6.2 The flip-walk-flip random walk

Let Y_1, Y_2 be independent and uniform on $\{e, b\}$, where e is the identity $(0,0)$ of the lamplighter group, and $b \in S$ is equivalent to δ_0. Let W be uniform on $\{a, a^{-1}\}$, two of the generators. Let $X_1 = Y_1 W Y_2$, and let μ be the distribution of X_1. So X_1 is chosen at random by uniformly and independently (1) telling the lamplighter to flip or not (2) telling the lamplighter to move either left or right, and (3) again telling the lamplighter to flip or not.

As usual, we will take X_n i.i.d. μ and $Z_n = X_1 X_2 \cdots X_n$. The map $\pi\colon \mathbb{Z}_2 \wr \mathbb{Z} \to \mathbb{Z}$ given by $\pi(f,x) = x$ is a group homomorphism (i.e., $\pi(g_1, g_2) = \pi(g_1) + \pi(g_2)$, and so $\pi(Z_n)$ is the simple random walk on $\mathbb{Z}$. Let $c\colon \mathbb{Z}_2 \wr \mathbb{Z} \to \bigoplus_{\mathbb{Z}} \mathbb{Z}_2$ be the configuration $c(f,x) = f$.

The support of this random walk at time n is B_{3n}, and in particular the support has exponential growth, as in the free group. So a natural guess is that the return probabilities $\mathbb{P}[Z_n = e]$ decay exponentially. As we will see, this turns out to be false. Nevertheless, the return probabilities are summable, and hence the random walk is transient.

The reason to look at this particular random walk is that given the locations $V_n = \{\pi(Z_1), \ldots, \pi(Z_n)\}$ visited by the lamplighter up to time n, the configuration $c(Z_n)$ is distributed uniformly on V_n. Thus,

$$\mathbb{P}[Z_n = e | V_n] \leq 2^{-|V_n|},$$

since $Z_n = e$ implies in particular that all lamps are off. Recall that $\pi(Z_n)$ is with high probability order of $\sqrt{n}$, and hence $|V_n|$ is, with high probability, at least $\sqrt{n}$. It can be furthermore shown that the probability that $|V_n|$ is less than (say) $n^{1/4}$ is of order $1/n^{1+\delta}$ for some $\delta > 0$. Hence

$$\mathbb{P}[Z_n = e] \leq \frac{1}{n^{1+\delta}} + 2^{-n^{1/4}},$$

and in particular $\sum_n \mathbb{P}[Z_n = e]$ is finite. So the random walk is transient.

7 Random walks on finitely generated groups

7.1 Finitely generated groups

Let $G = \langle S \rangle$ be a group generated by a finite, symmetric set S. We have seen a few examples. Another one is the group $\mathrm{SL}(2,\mathbb{Z})$ of two-by-two integer matrices with integer entries and determinant 1, with the operation of multiplication. This is a group since the determinant of each such matrix is one, and so its inverse is also in $\mathrm{SL}(2,\mathbb{Z})$. Multiplication is clearly associative and remains in $\mathrm{SL}(2,\mathbb{Z})$. What is less obvious is that $\mathrm{SL}(2,\mathbb{Z})$ is finitely generated. We will not prove this, but it turns out that it is generated by

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

and their inverses.

An even simpler example is $\mathrm{Iso}(\mathbb{Z})$. This is the group of linear bijections $g\colon \mathbb{Z} \to \mathbb{Z}$ such that $|z_1 - z_2| = |g(z_1) - g(z_2)|$ (it is also called the infinite dihedral group). These are the functions of the form $g(z) = rz + d$, where $r \in \{-1,+1\}$ and $d \in \mathbb{Z}$. It is generated by $a(z) = z+1$, $a^{-1}(z) = z-1$ and $b(z) = -z$.

For a given generating set S, we can define a norm on G by letting $|g|$ equal the minimal k such that g can be written at the product of k elements of S. This is called a norm since $|gh| \le |g| + |h|$, $|g^{-1}| = |g|$, $|g| \ge 0$ with equality iff $g = e$, where e denotes the identity element. We can use this norm to define the metric $d\colon G \times G \to \mathbb{N}$ by $d(g,h) = |g^{-1}h|$. This is equal to the minimal k such that $h = g s_1 \cdots s_k$ for $s_i \in S$. The norm $|g|$ is the distance of g from e in the Cayley graph, and $d(g,h)$ is the distance between g and h. Note that d is *left-invariant* in the sense that $d(kg,kh) = d(g,h)$ for all $g,h,k \in G$.

The norm and metric clearly depend on the choice of generating set, and when we want to be explicit about that we will write $|g|_S$ and d_S. Nevertheless, the following claim shows that the choice of generating set does not substantially affect either.

Claim 7.1. *Let $G = \langle S \rangle = \langle T \rangle$. Then there exists a constant $m > 0$ such that, for all $g \in G$,*

$$\frac{1}{m}|g|_S \le |g|_T \le m|g|_S.$$

Denote the by $B_n = \{g \in G : |g| \le n\}$ the ball of radius n in G. The *exponential growth rate* of G is given by

$$\mathrm{GR}(G) = \lim_n \frac{1}{n} \log |B_n|. \tag{7.1}$$

By Claim 7.1, the growth rate is independent of the choice of generating set. However, it is not a priori obvious that the limit exists. To show this, we will first show that the sequence

$$b_n = \log |B_n|$$

is *subadditive*.

Claim 7.2. $b_{n+m} \le b_n + b_m$.

Proof. Write $g \in B_{n+m}$ as $g = s_1 \cdots s_{n+m}$. Then $g = g_1 g_2$ where $g_1 = s_1 \cdots s_n$ and $g_2 = s_{n+1} \cdots s_{n+m}$. Thus $g_1 \in B_n$ and $g_2 \in B_m$. Hence the map $B_n \times B_m \to B_{n+m}$ given by $(g_1, g_2) \mapsto g_1 \cdot g_2$ is onto, and so $|B_{n+m}| \le |B_n| \cdot |B_m|$. □

We can now apply the *Fekete Lemma*.

Lemma 7.3 (Fekete Lemma)**.** *Let* $(a_n)_n$ *be a subadditive sequence. Then* $\lim_n a_n/n$ *exists and is equal to* $\inf_n a_n/n$.

This lemma, together with the previous claim, show that the limit in (7.1) exists. It furthermore shows that it is equal to $\inf_n \frac{1}{n} \log |B_n|$.

7.2 Random walks

Let μ be a finitely supported probability measure on G. We define the μ-random walk on G as before, by letting $(X_1, X_2, \ldots)$ be i.i.d. μ, setting $Z_0 = e$ and $Z_n = X_1 X_2 \cdots X_n$. We assume that μ is *non-degenerate* in the sense that for every $g \in G$ there is some n such that $\mathbb{P}[Z_n = g] > 0$. We say that μ is symmetric if $\mu(g) = \mu(g^{-1})$ for all $g \in G$. We denote by $\mu^{(n)}$ the distribution of Z_n. This is the n-fold convolution of μ with itself. Convolution of measures on G is given by

$$[\eta * \nu](g) = \sum_{h \in G} \eta(gh^{-1}) \nu(h) = \sum_{k \in G} \eta(k) \nu(k^{-1} g),$$

where the second equality follows by the change of variables $k = gh^{-1}$. Note that when G is not commutative then convolution is not commutative either. It is, however, associative.

7.3 The max-entropy

We define the *max entropy* $h_\infty(\mu)$ by

$$h_\infty(\mu) = \lim_n -\frac{1}{n} \log \left(\max_g \mathbb{P}[Z_n = g] \right) = \lim_n -\frac{1}{n} \max \log \mu^{(n)}.$$

Thus, if $h_\infty(\mu) = r \ge 0$ then the highest probability at time n is $e^{-rn + o(n)}$. Of course, we need to prove that this limit exists for this to be well defined.

Claim 7.4. *Let* $\zeta = \eta_1 * \eta_2$ *for* η_1, η_2 *probability measures on* G. *Then* $\max \zeta \ge (\max \eta_1) \cdot (\max \eta_2)$.

Proof. Suppose that the maxima of η_1 and η_2 are attained at g_1 and g_2 respectively. Then $\zeta(g_1 g_2) \ge \eta_1(g_1) \cdot \eta_2(g_2) = (\max \eta_1) \cdot (\max \eta_2)$. □

For a probability measure ν on G let

$$H_\infty(\nu) = -\max \log \mu.$$

Then we have shown that

$$H_\infty(\eta_1 * \eta_2) \le H_\infty(\eta_1) + H_\infty(\eta_2).$$

It follows that the sequence $a_n = H_\infty(\mu^{(n)})$ is subadditive. We can now apply the Fekete Lemma, which implies that $\lim_n \frac{1}{n} H_\infty(\mu^{(n)})$ exists. But this is exactly equal to $h_\infty(\mu)$.

Proposition 7.5. *Suppose that μ is symmetric. Then*

$$h_\infty(\mu) = \lim_n -\frac{1}{2n} \log \mathbb{P}[Z_{2n} = e].$$

Proof. Pick $g_n \in \operatorname{argmax}_g \mathbb{P}[Z_n = g]$ that maximizes the probability that Z_n visits g. I.e., $\mathbb{P}[Z_n = g_n] = \max \mu^{(n)}$. Then

$$\mathbb{P}[Z_{2n} = e] \ge \mathbb{P}[X_1 \cdots X_n = g_n] \cdot \mathbb{P}\left[X_{n+1} \cdots X_{2n} = g_n^{-1}\right] = \mu^{(n)}(g_n) \cdot \mu^{(n)}(g_n^{-1}) = (\max \mu^{(n)})^2.$$

Therefore, and since $\max \mu^{(2n)} \ge \mu^{(2n)}(e) = \mathbb{P}[Z_{2n} = e]$,

$$\max \mu^{(2n)} \ge \mathbb{P}[Z_{2n} = e] \ge (\max \mu^{(n)})^2$$

and

$$-\frac{1}{2n} \log \max \mu^{(2n)} \le -\frac{1}{2n} \log \mathbb{P}[Z_{2n} = e] \le -\frac{1}{n} \log \max \mu^{(n)}.$$

Taking the limit $n \to \infty$ yields the result. $\square$

8 The Markov operator and the spectral norm

8.1 The Markov operator of a random walk

For a finitely generated group G, denote by $\mathbb{R}^G$ the vector space of real functions $G \to \mathbb{R}$. Denote by $\ell^2(G)$ the Hilbert space of functions $\varphi \colon G \to \mathbb{R}$ such that $\sum_g \varphi(g)^2 < \infty$. This space is equipped with inner product $\langle \varphi, \psi \rangle = \sum_g \varphi(g)\psi(g)$ and, as usual, the norm

$$\|\varphi\|_2^2 = \langle \varphi, \varphi \rangle$$

We will refer to $(\delta_g)_{g \in G}$ as the standard basis of $\ell^2(G)$. In this basis we can write

$$\varphi = \sum_{g \in G} \varphi(g)\delta_g.$$

More generally, for $p \geq 1$, denote by $\ell^p(G)$ the Banach space of functions $\varphi \colon G \to \mathbb{R}$ such that

$$\|\varphi\|_p^p := \sum_g \varphi(g)^p < \infty. \tag{8.1}$$

As usual $\ell^\infty(G)$ will be the Banach space of bounded functions with norm $\|\varphi\|_\infty = \sup_g |\varphi(g)|$.

For each $h \in G$ define the *right translation* linear operator $R_h : \mathbb{R}^G \to \mathbb{R}^G$

$$[R_h\varphi](g) = \varphi(gh).$$

Applying a change of variable to (8.1) shows that $\|R_h\varphi\|_p = \|\varphi\|$, so that R_h is an isometry for all $\ell^p(G)$. Note that $R_h R_k = R_{hk}$ and that $R_{h^{-1}} = R_h^{-1}$. This makes the map $h \mapsto R_h$ a *representation of* G.

Let μ be a non-degenerate, finitely supported symmetric measure on a finitely generated group G. The *Markov operator* $M : \mathbb{R}^G \to \mathbb{R}^G$ associated with μ is the linear operator given by $M = \sum_h \mu(h) R_h$, so that

$$[M\varphi](g) = \sum_h \mu(h)\varphi(gh).$$

One way to think of this operator is as follows: If $\psi = M\varphi$ then $\psi(g) = \mathbb{E}\left[\varphi(gX_1)\right]$ is the expectation of φ at the location visited by the random walk after visiting g. There is another interpretation: the *matrix entries* of M with respect to the standard basis are the transition probabilities of the Markov chain:

$$\langle \delta_h, M\delta_g \rangle = [M\delta_g](h) = \mathbb{P}[Z_{n+1} = g | Z_n = h],$$

provided that $\mathbb{P}[Z_n = g] > 0$. Likewise, the powers of M capture the n-step transition probabilities:

$$\left\langle \delta_h, M^k \delta_g \right\rangle = \mathbb{P}[Z_{n+k} = g | Z_n = h]. \tag{8.2}$$

Claim 8.1. *For $p \geq 1$ and $\varphi \in \ell^p(G)$, $\|M\varphi\|_p \leq \|\varphi\|_p$, with a strict inequality for $p > 1$ and $\varphi \neq 0$.*

Since μ has finite support $\{h_1, \ldots, h_k\}$, this claim can be proved by looking at the finite dimensional space $\text{span}\{\varphi, R_{h_1}\varphi, \ldots, R_{h_k}\varphi\}$. The proof then follows from the fact that ℓ^p-balls in $\mathbb{R}^d$ are convex: every convex combination of unit vectors has norm at most one. For $p > 1$, balls are strictly convex. This implies that we in fact have a strict inequality, unless $\varphi = 0$. The important fact for us is that M is a bounded operator on $\ell^p(G)$.

8.2 Self-adjointness and return probabilities

Since μ is symmetric, an important property of the Markov operator is that it is *self-adjoint*: $M^\dagger = M$. That is, for all $\varphi, \psi \in \ell^2(G)$,

$$\langle \psi, M\varphi \rangle = \langle M\psi, \varphi \rangle.$$

The property of being self-adjoint is a generalization to Hilbert spaces of the symmetry property of finite dimensional (real) matrices. To see that M is self-adjoint, note that the adjoint of R_h is $R_h^\dagger = R_{h^{-1}}$:

$$\begin{aligned}
\langle \varphi, R_h\psi \rangle \sum_g \varphi(g)[R_h\psi](g) \\
&= \sum_g \varphi(g)\psi(gh) \\
&= \sum_k \varphi(kh^{-1})\psi(k) \\
&= \langle R_{h^{-1}}\varphi, \psi \rangle.
\end{aligned}$$

Hence the symmetry of μ implies that the adjoint of $M = \sum_h \mu(h) R_h$ is

$$M^\dagger = \sum_h \mu(h) R_{h^{-1}} = \sum_h \mu(h^{-1}) R_h = \sum_h \mu(h) R_h = M.$$

As a corollary, we provide a simple proof of the following claim.

Claim 8.2. *When μ is symmetric, $\mathbb{P}[Z_{2n} = e] \geq \mathbb{P}[Z_{2n} = g]$ for all $g \in G$.*

Proof.

$$\begin{aligned}
0 &\leq \|M^n(\delta_g - \delta_e)\|^2 \\
&= \langle M^n(\delta_g - \delta_e), M^n(\delta_g - \delta_e) \rangle \\
&= \langle M^n\delta_g, M^n\delta_g \rangle - 2\langle M^n\delta_g, M^n\delta_e \rangle + \langle M^n\delta_e, M^n\delta_e \rangle \\
&= \langle \delta_g, M^{2n}\delta_g \rangle - 2\langle \delta_g, M^{2n}\delta_e \rangle + \langle \delta_e, M^{2n}\delta_e \rangle,
\end{aligned}$$

where the last equality follows from the fact that M is self-adjoint. Now, by (8.2)

$$\langle \delta_g, M^{2n}\delta_g \rangle = \mathbb{P}[gZ_{2n} = g] = \mathbb{P}[Z_{2n} = e]$$

and

$$\left\langle \delta_g, M^{2n}\delta_e \right\rangle = \mathbb{P}[Z_n = g].$$

Hence

$$\mathbb{P}[Z_{2n} = g] \le \mathbb{P}[Z_{2n} = e].$$

□

8.3 The spectral norm

In this section we will denote the ℓ^2 norm by $\|\cdot\|$. The norm of the Markov operator M, as a linear operator on the Hilbert space $\ell^2(G)$, is given by

$$\|M\| = \sup\{\|M\varphi\| : \|\varphi\| = 1\} = \sup\left\{\frac{\|M\varphi\|}{\|\varphi\|} : \varphi \neq 0\right\}.$$

By Claim 8.1, $\|M\| \le 1$. The following theorem relates the norm of M to the max-entropy of the random walk. The norm of M is also known as the *spectral radius* of the random walk.

Theorem 8.3. *For all symmetric, finitely supported μ, $\|M\| = \mathrm{e}^{-h_\infty(\mu)}$.*

By Proposition 7.5, this implies that $\|M\| = \lim_n \mathbb{P}[Z_{2n} = e]^{1/(2n)}$.

To prove this theorem we will need some facts about self-adjoint operators on Hilbert spaces. Before stating our claims, we will discuss the simpler, finite dimensional case.

In $\mathbb{R}^n$, a self-adjoint operator can be represented by a real symmetric matrix A. Such a matrix will have distinct real eigenvalues $\lambda_1, \cdots, \lambda_k$ for some $k \le n$. Furthermore, for every vector $v \in \mathbb{R}^n$ we can find orthonormal eigenvectors $w_1, \ldots, w_k$ (corresponding to the above eigenvalues) such that $v = \sum_{i=1}^k \alpha_i w_i$. It follows that the operator norm of A in this case is $\max_i |\lambda_i|$.

Using the eigenvector basis, we can calculate

$$A^n v = \sum_{i=1}^k \alpha_i \lambda_i^n w_i.$$

Hence

$$\left\|A^n v\right\|^2 = \sum_{i=1}^k |\alpha_i|^2 |\lambda_i|^{2n}.$$

and in particular, denoting $|\lambda_m| = \max\{|\lambda_i| : \alpha_i > 0\}$,

$$\lim_n \left\|A^n v\right\|^{1/n} = |\lambda_m|.$$

and if $\|v\| = 1$ then

$$\|Av\| \le \lim_n \left\|A^n v\right\|^{1/n} \le \|A\|.$$

The following claim shows that the same holds in Hilbert spaces. We say that an operator on a Hilbert space is bounded if it has finite norm.

Lemma 8.4. *Let A be a self-adjoint bounded operator on a Hilbert space $\mathscr{H}$. Then for any unit vector $v \in \mathscr{H}$*

$$\|Av\| \leq \lim_n \left\|A^n v\right\|^{1/n} \leq \|A\|.$$

Proof. Fix a unit vector $v \in \mathscr{H}$. Since A is self-adjoint,

$$\left\|A^{n+1}v\right\|^4 = \left\langle A^{n+1}v, A^{n+1}v\right\rangle^2 = \left\langle A^n v, A^{n+2}v\right\rangle^2.$$

Applying Cauchy-Schwarz we get

$$\left\|A^{n+1}v\right\|^4 \leq \left\|A^n v\right\|^2 \cdot \left\|A^{n+2}v\right\|^2.$$

Dividing both sides by $\left\|A^{n+1}v\right\|^2 \cdot \|A^n v\|^2$ and taking the square root yields

$$\frac{\left\|A^{n+1}v\right\|}{\|A^n v\|} \leq \frac{\left\|A^{n+2}v\right\|}{\left\|A^{n+1}v\right\|}.$$

Thus the sequence $\frac{\|A^{n+1}v\|}{\|A^n v\|}$ is non-decreasing and converges to some ρ:

$$\rho = \lim_n \frac{\left\|A^{n+1}v\right\|}{\|A^n v\|}.$$

Since

$$\left\|A^n v\right\| = \frac{\|Av\|}{\|v\|} \cdots \frac{\|A^n v\|}{\left\|A^{n-1}v\right\|}$$

we can conclude that

$$\lim_n \left\|A^n v\right\|^{1/n} = \rho$$

with

$$\|Av\| \leq \rho \leq \|A\|.$$

$\square$

Denote by $\ell^2_f(G)$ the finitely supported $\varphi \in \ell^2(G)$. Recall that

$$\|M\| = \sup\{\left\|M\varphi\right\| : \left\|\varphi\right\| = 1\}.$$

Since we can approximate any $\varphi \in \ell^2(G)$ by a finitely supported $\varphi' \in \ell^2_f(G)$, in the sense that $\left\|\varphi - \varphi'\right\| < \varepsilon$, the continuity of M implies that

$$\|M\| = \sup\{\left\|M\varphi\right\| : \left\|\varphi\right\| = 1, \varphi \in \ell^2_f(G)\}. \tag{8.3}$$

Choose any $\varphi \in \ell^2_f(G)$ with $\|\varphi\| = 1$. Since M is self-adjoint,

$$\|M^n\varphi\|^2 = \langle M^n\varphi, M^n\varphi\rangle = \langle \varphi, M^{2n}\varphi\rangle.$$

Denote $\operatorname{supp}\varphi = F \subset G$. Then, since $\varphi = \sum_{g\in F}\varphi(g)\delta_g$, we can write the above as

$$\|M^n\varphi\|^2 = \sum_{g,h\in F}\varphi(g)\varphi(h)\langle \delta_g, M^{2n}\delta_h\rangle.$$

Recalling that the matrix entries are the Markov transition properties we have

$$\begin{aligned}\|M^n\varphi\|^2 &= \sum_{g,h\in F}\varphi(g)\varphi(h)\mathbb{P}[hZ_{2n}=g]\\ &\le \sum_{g,h\in F}|\varphi(g)\varphi(h)|\mathbb{P}[hZ_{2n}=g].\end{aligned}$$

By Claim 8.2, $\mathbb{P}[hZ_{2n}=g] \le \mathbb{P}[Z_{2n}=e]$. Hence

$$\begin{aligned}\|M^n\varphi\|^2 &\le \sum_{g,h\in F}|\varphi(g)\varphi(h)|\mathbb{P}[Z_{2n}=e]\\ &= \sum_{g,h\in F}|\varphi(g)|\cdot|\varphi(h)|\mathbb{P}[Z_{2n}=e]\\ &= \mathbb{P}[Z_{2n}=e]\sum_{g\in F}|\varphi(g)|\sum_{h\in F}|\varphi(h)|.\end{aligned}$$

Now, $|\varphi(g)| \le 1$, since $\sum_g \varphi(g)^2 = 1$. Hence

$$\|M^n\varphi\|^2 \le \mathbb{P}[Z_{2n}=e]|F|^2.$$

It follows that

$$\lim_n \|M^n\varphi\|^{1/n} \le \lim_n \mathbb{P}[Z_{2n}=e]^{1/(2n)} = \mathrm{e}^{-h_\infty(\mu)}.$$

By the first inequality of Lemma 8.4

$$\|M\varphi\| \le \lim_n \|M^n\varphi\|^{1/n},$$

and so, by (8.3),

$$\|M\| \le \mathrm{e}^{-h_\infty(\mu)}.$$

Finally,

$$\mathrm{e}^{-h_\infty(\mu)} = \lim_n \mathbb{P}[Z_{2n}=e]^{1/(2n)} = \lim_n \langle \delta_e, M^{2n}\delta_e\rangle^{1/(2n)} = \lim_n \|M^n\delta_e\|^{1/n}.$$

and so applying the second inequality of Lemma 8.4 to $v = \delta_e$ yields that

$$\mathrm{e}^{-h_\infty(\mu)} = \lim_n \|M^n\delta_e\|^{1/n} \le \|M\|.$$

This concludes the proof of Theorem 8.3.

9 Amenability and Kesten's Theorem

9.1 Følner sequences and the isoperimetric constant

Let $G = \langle S \rangle$ be a finitely generated group. Given a set $F \subset G$, we denote the *boundary* of F by

$$\partial F = \{g \not\in F : \exists s \in S \text{ s.t. } gs \in F\}.$$

This is the set of vertices in the Cayley graph that are not in F but are connected to a vertex in F. Note that this definition depends on S, and we write $\partial_S F$ when we want to make this dependence explicit.

The *surface-to-volume ratio* of a finite $F \subset G$ is $|\partial F|/|F|$. The *isoperimetric* constant of G (with respect to S) is

$$\Phi(G,S) = \inf_{F \subset G} \frac{|\partial_S F|}{|F|},$$

where the infimum is taken over finite F.

A group G is said to be *amenable* if $\Phi(G,S) = 0$. This notion is well-defined (i.e., independent of the choice of S) since, by Claim 7.1, if S and T are generating sets then there exists a constant $m > 0$ such that

$$\frac{1}{m}|\partial_S F| \leq |\partial_T F| \leq m|\partial_S F|. \tag{9.1}$$

Equivalently, G is amenable if there is a sequence of finite subsets F_n with surface-to-volume ratio tending to zero. Such sequences are called *Følner sequences*. By (9.1), a sequence is Følner with respect to one generating set if it is Følner with respect to another.

It is useful to also define the *inner boundary* $\partial^i F$

$$\partial F = \{f \in F : \exists s \in S \text{ s.t. } fs \not\in F\}.$$

This is the set of vertices in F that are connected to a vertex outside of F. Since each vertex has $|S|$ edges,

$$\frac{1}{|S|} \cdot |\partial F| \leq |\partial^i F| \leq |S| \cdot |\partial F|. \tag{9.2}$$

We can thus equivalently define Følner sequences and amenability using the inner boundary.

9.2 Examples

To see that $\mathbb{Z}^d$ is amenable, we can verify that $F_n = \{1,\dots,n\}^d$ is a Følner sequence.

Claim 9.1. *$G = \langle S \rangle$ is amenable if* $\mathrm{GR}(G) = 0$.

Proof. Since $B_{n+1} = B_n \cup \partial B_n$, $|B_{n+1}| \geq |B_n| \cdot (1+\Phi(G,S))$. Hence $|B_{n+1}| \geq (1+\Phi(G,S))^n$ and

$$\mathrm{GR}(G) = \lim_n \frac{1}{n} \log |B_n| \geq \log(1+\Phi(G,S)).$$

Thus, if G is non-amenable then $\mathrm{GR}(G) > 0$. $\square$

It may be tempting to imagine that the converse of Claim 9.1 is true. However, the lamplighter group has exponential growth even though it is amenable. Fix the generating set $S = \{(0,+1),(0,-1),(\delta_0,0)\}$. Denote $I_n = \{-n,\dots,n-1\}$. Consider the set

$$F_n = \big\{(f,z) : \operatorname{supp} f \subseteq I_n, z \in I_n\big\}.$$

it is of size exactly $2n \cdot 2^{2n}$ and is contained in B_{6n}, and so $|B_{6n}| \geq 2^n$. Thus the lamplighter has exponential growth. To see that it is amenable, note that

$$\partial F_n = \big\{(f,z) : \operatorname{supp} f \subseteq I_n, z \in \{-n-1, n\}\big\}$$

and so $|\partial F_n| = 2 \cdot 2^{2n}$. Thus F_n is a Følner sequence.

9.3 Kesten's Theorem

In the next claim we denote symmetric differences by $\triangle$.

Claim 9.2. *Let $G = \langle S \rangle$ be a finitely generated group. Let $(F_1, F_2, \dots)$ be a sequence of finite subsets of G. The following are equivalent.*

1. *F_n is a Følner sequence.*
2. *For every $s \in S$*

$$\lim_n \frac{|F_n \triangle F_n s|}{|F_n|} = 0.$$

3. *For every $h \in G$*

$$\lim_n \frac{|F_n \triangle F_n h|}{|F_n|} = 0.$$

In this claim, $F_n h$ is the set $\{fh : f \in F\}$. The proof of this claim relies on (9.1), as well as the observation that $F \triangle F s \subseteq \partial F \cup \partial^i F$.

Theorem 9.3 (Kesten)**.** *Let G be a finitely generated group, and let μ be a finitely supported, symmetric, non-degenerate probability measure on G. If G is amenable then $\|M\| = 1$.*

Proof. Let $S = \operatorname{supp}\mu$ be a symmetric generating set. Let F be a finite subset of G, and let $\varphi\colon G \to \{0,1\}$ be the indicator of F: $\varphi(g) = \mathbb{1}_{\{g \in F\}}$. Let $\psi = M\varphi$, and note that

1. $\psi(g) \in [0,1]$.

2. $\psi(g) = 1$ for all $g \in F \setminus \partial^i F$.

3. $\psi(g) = 0$ for all $g \notin F \cup \partial F$.

In particular, $\psi(g) \neq \varphi(g)$ only for $g \in \partial F \cup \partial^i F$. Hence

$$\left\|\varphi - M\varphi\right\|_2^2 = \left\|\varphi - \psi\right\|_2^2 \leq |\partial F \cup \partial^o F| \leq (1+|S|)|\partial F|,$$

by (9.2). In particular, by the triangle inequality

$$\left\|M\varphi\right\| \geq \left\|\varphi\right\| - \sqrt{(1+|S|)|\partial F|},$$

and so

$$\frac{\left\|M\varphi\right\|}{\left\|\varphi\right\|} \geq 1 - \frac{\sqrt{(1+|S|)|\partial F|}}{\left\|\varphi\right\|} = 1 - \sqrt{\frac{(1+|S|)|\partial F|}{|F|}}.$$

Letting φ_n be the indicators of a Følner sequence F_n yields that

$$\lim_n \frac{\left\|M\varphi_n\right\|}{\left\|\varphi_n\right\|} \geq 1,$$

and so $\|M\| = 1$, since we already showed in Claim 8.1 that $\|M\| \leq 1$. □

Let η, ν be probability measures on G. We view them as elements of $\ell^1(G)$. As such, the distance between them is

$$\left\|\eta - \nu\right\| = \sum_{g \in G} |\eta(g) - \nu(g)|.$$

We can also apply the right translation operators R_h to them:

$$[R_h \nu](g) = \nu(gh).$$

Theorem 9.4. *Let $G = \langle S \rangle$ be a finitely generated group. The following are equivalent.*

1. *G is amenable.*

2. *There is a sequence ν_n of probability measures on G such that*

$$\lim_n \|\nu_n - R_s \nu_n\|_1 = 0$$

for all $s \in S$.

3. *There is a sequence φ_n of vectors in $\ell^2(G)$ such that*

$$\lim_n \left\|\varphi_n - R_s \varphi_n\right\|_2 = 0$$

for all $s \in S$.

In the latter two conditions, one can replace $s \in S$ with $s \in G$, and get two more equivalent conditions.

Theorem 9.5 (Kesten)**.** *Let G be a finitely generated group, and let μ be a finitely supported, symmetric, non-degenerate probability measure on G. If G is not amenable then $\|M\| < 1$.*

This Theorem, together with (5.1), implies that the free group $\mathbb{F}_2$ is not amenable.

To prove this theorem we will need a simple lemma on Markov operators. A Hilbert space is separable if it has a countable basis. For example, our space $\ell^2(G)$ is separable because it admits the countable basis δ_g.

Lemma 9.6. *Let A be a self-adjoint operator on a separable Hilbert space $\mathscr{H}$ with $\|A\| = 1$. Suppose that the matrix entries $\langle e_i, Ae_j \rangle$ are non-negative for some countable orthonormal basis e_i, $i \in I$. Then there is a sequence of unit vectors $w_n \in \mathscr{H}$ such that*

$$\lim_n \langle w_n, Aw_n \rangle = 1.$$

To see that the assumption that A has positive entries is necessary, consider the operator $A : \mathbb{R} \to \mathbb{R}$ given by $a(x) = -x$. For finite dimensional $\mathscr{H}$ this is part of the statement of the Perron-Frobenius Theorem.

Proof of Lemma 9.6. We identify each vector $v = \sum_i \langle v, e_i \rangle e_i$ with the function $I \to \mathbb{R}$ given by $v(i) = \langle v, e_i \rangle$.

Since $\|A\| = 1$ there is a sequence of unit vectors $v_n \in \mathscr{H}$ such that $\lim_n \|Av_n\| = 1$, and hence $\lim_n \langle v_n, A^2 v_n \rangle = 1$, since A is self-adjoint. We would like to have vectors for which this holds for A rather than A^2.

Since the matrix entries $\langle e_i, Ae_j \rangle$ are non-negative, the matrix entries $\langle e_i, A^2 e_j \rangle$ are non-negative, and for every v

$$\|Av\| = \langle v, A^2 v \rangle = \sum_{i,j} v(i)v(j) \langle e_i, Ae_j \rangle \leq \sum_{j,i} |v(i)| \cdot |v(j)| \langle e_i, A^2 e_j \rangle.$$

Thus we can assume that $v_n(i)$ is non-negative. Hence $[Av_n](i)$ is also non-negative, and $\langle v_n, Av_n \rangle > 0$. We further can assume that $\langle v_n, Av_n \rangle \in [0,1]$ converges to some $\alpha \in [0,1]$.

Define $u_n = v_n + Av_n$ then

$$\begin{aligned}
\lim_n \langle u_n, Au_n \rangle &= \lim_n \langle v_n + Av_n, Av_n + A^2 v_n \rangle \\
&= \lim_n \langle v_n, Av_n \rangle + \langle v_n, A^2 v_n \rangle + \langle Av_n, Av_n \rangle + \langle Av_n, A^2 v_n \rangle \\
&= \lim_n 2\alpha + 2.
\end{aligned}$$

Now,

$$\lim_n \|u_n\|^2 = \lim_n \|v_n\|^2 + \|Av_n\|^2 + 2 \langle v_n, Av_n \rangle = 2 + 2\alpha > 0,$$

and so we have that for $w_n = u_n / \|u_n\|$

$$\lim_n \langle w_n, Aw_n \rangle = 1.$$

□

Given this, we can proceed with the proof of the theorem.

Proof of Theorem 9.5. By Lemma 9.6, there is a sequence of unit vectors $\varphi_n \in \ell^2(G)$ such that

$$1 = \lim_n \langle \varphi_n, M\varphi_n \rangle = \lim_n \sum_h \mu(h) \langle \varphi_n, R_h \varphi_n \rangle .$$

Observe that each term $\langle \varphi_n, R_h \varphi_n \rangle$ on the right hand side is at most 1, since $\|R_h \varphi_n\| = 1$. And since the right hand side is a finite (weighted) average of these terms,

$$\lim_n \langle \varphi_n, R_h \varphi_n \rangle = 1$$

and

$$\lim_n \|\varphi_n - R_h \varphi_n\| = 0.$$

So by Theorem 9.4, G is amenable, since $\operatorname{supp} \mu$ is a generating set. $\square$

10 The Carne-Varopoulos bound

10.1 Theorem statement

The Hoeffding bound for $\mathbb{Z}^d$ can be stated as follow:

$$\mathbb{P}[Z_n = z] \leq 2\mathrm{e}^{-\frac{|z|^2}{2n}},$$

where $|z|$ is the norm of z, calculated using the generating set $\operatorname{supp}\mu$. The next theorem generalizes this to all finitely generated groups.

Theorem 10.1 (Carne-Varopoulos)**.** *Let $G = \langle S \rangle$ be a finitely generated group, and let μ be a symmetric measure with support S. Let M be the corresponding Markov operator. Then for any $g \in G$,*

$$\mathbb{P}[Z_n = g] \leq 2\,\|M\|^n\, \mathrm{e}^{-\frac{|g|^2}{2n}}.$$

It follows that if G has sub-exponential growth, then the random walk Z_n is concentrated with distance roughly $\sqrt{n}$, just like on $\mathbb{Z}^d$.

10.2 Harmonic oscillator

To prove this theorem we will need to adapt some techniques from physics. Consider a mass that can move up or down. We denote its position at (continuous) time t by φ_t, and its speed by ψ_t, so that

$$\frac{\mathrm{d}\varphi_t}{\mathrm{d}t} = \psi_t.$$

It connected to a spring that pulls it back in, with a force equal to $-L \cdot \varphi_t$, so that the further it is the stronger the pull. Thus

$$\frac{\mathrm{d}\psi_t}{\mathrm{d}t} = -L\varphi_t.$$

We can write these equation as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\varphi_t\\ \psi_t\end{pmatrix} = V\begin{pmatrix}\varphi_t\\ \psi_t\end{pmatrix}$$

where

$$V = \begin{pmatrix}0 & 1\\ -L & 0\end{pmatrix}.$$

The solution is

$$\begin{pmatrix}\varphi_t\\ \psi_t\end{pmatrix} = \mathrm{e}^{tV}\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix},$$

or

$$\begin{pmatrix}\varphi_t\\ \psi_t\end{pmatrix} = \begin{pmatrix}\cos(\sqrt{L}t) & \frac{1}{\sqrt{L}}\sin(\sqrt{L}t)\\ -\sqrt{L}\sin(\sqrt{L}t) & \cos(\sqrt{L}t)\end{pmatrix}\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix}.$$

Note that the energy $E_t = L\varphi_t^2 + \psi_t^2$ is conserved, so that e^{tV} is an orthogonal operator on $\mathbb{R}^2$ for the norm given by the energy.

We would like to do the same thing in discrete time. It is tempting, in analogy to the continuous time differential equations, to consider the discrete time system

$$\begin{aligned}\varphi_{n+1} &= \varphi_n + \psi_n\\ \psi_{n+1} &= \psi_n - L\varphi_t,\end{aligned}$$

or

$$\begin{pmatrix}\varphi_{n+1}\\ \psi_{n+1}\end{pmatrix} = (I+V)\begin{pmatrix}\varphi_n\\ \psi_n\end{pmatrix}.$$

The problem is that energy is no longer preserved: this is not an orthogonal operator. The mistake is that we have taken the operator to be $I+V$ rather than e^V. Indeed, we need a matrix with unit determinant. We will take

$$U = \begin{pmatrix}M & 1\\ -(1-M^2) & M\end{pmatrix}$$

for $M<1$ which corresponds to $1-\frac{1}{2}L \approx \cos(\sqrt{L})$. Our discrete time system is thus

$$\begin{pmatrix}\varphi_{n+1}\\ \psi_{n+1}\end{pmatrix} = U\begin{pmatrix}\varphi_n\\ \psi_n\end{pmatrix},$$

so that

$$\begin{pmatrix}\varphi_n\\ \psi_n\end{pmatrix} = U^n\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix}.$$

The energy that is conserved is

$$E_n = (1-M^2)\varphi_n^2 + \psi^2.$$

10.3 Coupled harmonic oscillators and the continuous time wave equation

Consider now a unit mass located at each $g \in G$. The masses can again move up and down, and we denote the height of the mass at g at time t by $\varphi_t(g)$ and its velocity by $\psi_t(g)$, so that

$$\frac{\mathrm{d}\varphi_t(g)}{\mathrm{d}t} = \psi_t(g).$$

The masses are connected by springs to their neighbors in the Cayley graph, where the strength of the spring between g and gs is $\mu(s)$ for some symmetric probability measure μ on g. The strength of the attraction is proportional to the distance between them, and attraction translates to force on the mass at g (and thus acceleration) equal to $\mu(s)(\varphi(gs)-\varphi(g)$. We thus have that

$$\frac{\mathrm{d}\psi_t(g)}{\mathrm{d}t} = \sum_s \mu(s)\left(\varphi_t(gs)-\varphi_t(g)\right).$$

This system has an energy

$$E_t = \sum_g \psi_t(g)^2 + \frac{1}{2}\sum_g\sum_s \mu(s)(\varphi_t(gs)-\varphi_t(g))^2, \tag{10.1}$$

which is conserved over time:

$$\begin{aligned}
\frac{\mathrm{d}E_t}{\mathrm{d}t} &= 2\sum_g \psi_t(g)\frac{\mathrm{d}\psi_t(g)}{\mathrm{d}t} + \sum_g\sum_s \mu(s)\left(\varphi_t(gs)-\varphi_t(g)\right)\left(\frac{\mathrm{d}\varphi_t(gs)}{\mathrm{d}t} - \frac{\mathrm{d}\varphi_t(g)}{\mathrm{d}t}\right) \\
&= 2\sum_g \psi_t(g)\sum_s \mu(s)\left(\varphi_t(gs)-\varphi_t(g)\right) + \sum_g\sum_s \mu(s)\left(\varphi_t(gs)-\varphi_t(g)\right)\left(\psi_t(gs)-\psi_t(g)\right) \\
&= \sum_g \psi_t(g)\sum_s \mu(s)\left(\varphi_t(gs)-\varphi_t(g)\right) + \sum_g\sum_s \mu(s)\left(\varphi_t(gs)-\varphi_t(g)\right)\psi_t(gs).
\end{aligned}$$

This is equal to zero by applying the change of variable $g \mapsto gs$ the first summand and using the fact that μ is symmetric.

10.4 The Laplacian

We introduce some notation to help us write this more elegantly. Given $\varphi \in \mathbb{R}^G$, denote by $\nabla\varphi\colon G \to \mathbb{R}^S$ the map

$$[\nabla\varphi]_s(g) = \varphi(gs) - \varphi(g).$$

It is useful to think of $\nabla\varphi$ as the derivative of φ, with the component $[\nabla\varphi]_s$ being the derivative in the "direction" s. Clearly, it is a linear operator. Note that for $\theta = \nabla\varphi$ and $h = gs$ it holds that

$$\theta_s(g) = -\theta_{s^{-1}}(h).$$

We call such functions anti-symmetric.

In the context of a symmetric measure μ supported on a generating set S we define an inner product on the space of functions $G \to \mathbb{R}^S$ by

$$\langle\theta,\theta'\rangle = \frac{1}{2}\sum_g \langle\theta(g),\theta'(g)\rangle = \frac{1}{2}\sum_g\sum_s \mu(s)\theta_s(g)\theta'_s(g).$$

Of course, this is not defined for all θ, θ' and we restrict ourselves to $\theta: G \to \mathbb{R}^S$ such that $\|\theta\|^2 := \langle\theta,\theta\rangle < \infty$. We also restrict ourselves to anti-symmetric ψ. We denote the Hilbert space of such θ by $\ell^2(G,\mathbb{R}^S,AS)$.

For $\varphi \in \ell^2(G)$,

$$\begin{aligned}
\left\|\nabla\varphi\right\|^2 = \left\langle \nabla\varphi, \nabla\varphi \right\rangle &= \frac{1}{2}\sum_g\sum_s \mu(s)[\nabla\varphi]_s(g)[\nabla\varphi]_s(g) \\
&= \frac{1}{2}\sum_g\sum_s \mu(s)(\varphi(gs)-\varphi(g))^2 \\
&= \frac{1}{2}\sum_g\sum_s \mu(s)\left(\varphi(gs)^2+2\varphi(gs)\varphi(g)+\varphi(g)^2\right) \\
&= \left\langle \varphi,\varphi \right\rangle - \left\langle \varphi, M\varphi \right\rangle \\
&= \left\langle \varphi, (I-M)\varphi \right\rangle,
\end{aligned}$$

where I is the identity operator on $\ell^2(G)$. Thus ∇ is a bounded operator from $\ell^2(G)$ to $\ell^2(G,\mathbb{R}^S,AS)$. A similar calculation yields

$$\left\langle \nabla\psi, \nabla\varphi \right\rangle = \left\langle \psi, (I-M)\varphi \right\rangle. \tag{10.2}$$

The "opposite" of the "differentiation" operator ∇ is the "divergence" operator $\nabla^\dagger: (\mathbb{R}^S)^G \to R^G$ given by

$$[\nabla^\dagger\psi](g) = \sum_s \mu(s)\psi_{s^{-1}}(gs).$$

Indeed, the adjoint of ∇ is $\nabla^\dagger$:

$$\begin{aligned}
\left\langle \nabla^\dagger\psi, \varphi \right\rangle &= \sum_g [\nabla^\dagger\psi](g)\varphi(g) \\
&= \sum_g\sum_s \mu(s)\psi_{s^{-1}}(gs)\varphi(g) \\
&= \frac{1}{2}\sum_g\sum_s \mu(s)(\psi_{s^{-1}}(gs)-\psi_s(g))\varphi(g) \\
&= \frac{1}{2}\sum_g\sum_s \mu(s)(\psi_s(g)\varphi(gs)-\psi_s(g)\varphi(g)) \\
&= \frac{1}{2}\sum_g\sum_s \mu(s)\psi_s(g)(\varphi(gs)-\varphi(g)) \\
&= \frac{1}{2}\sum_g\sum_s \mu(s)\psi_s(g)[\nabla\varphi]_s(g) \\
&= \frac{1}{2}\sum_g \left\langle \psi(g), [\nabla\varphi](g) \right\rangle \\
&= \left\langle \psi, \nabla\varphi \right\rangle
\end{aligned}$$

Hence, by (10.2), $\nabla^\dagger\nabla = I - M$, which we denote by L and call the *Laplacian* of the random walk and.

Going back to our masses, recall that the equations governing the system are

$$\frac{\mathrm{d}\varphi_t(g)}{\mathrm{d}t} = \psi_t(g)$$
$$\frac{\mathrm{d}\psi_t(g)}{\mathrm{d}t} = \sum_s \mu(s)\big(\varphi(gs) - \varphi(g)\big).$$

Note that

$$[L\varphi](g) = [\nabla^\dagger\nabla\varphi](g) = \sum_s \mu(s)[\nabla\varphi]_{s^{-1}}(gs) = \sum_s \mu(s)\left(\varphi(g) - \varphi(gs)\right),$$

and so we write our equations as

$$\frac{\mathrm{d}\varphi_t}{\mathrm{d}t} = \psi_t$$
$$\frac{\mathrm{d}\psi_t}{\mathrm{d}t} = -L\varphi_t.$$

We can write our energy as

$$E_t = \|\psi\|^2 + \|\nabla\varphi\|^2 = \langle\psi,\psi\rangle + \langle\varphi, L\varphi\rangle.$$

Note that this is a norm on the Hilbert space $\mathscr{H} := \ell^2(G)\otimes\ell^2(G)$, and thus the dynamics is (differential) orthogonal operator that preserves this norm.

If we think of $\begin{pmatrix}\varphi_t\\ \psi_t\end{pmatrix}$ as an element of $\mathscr{H}$, we can write our equation as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\varphi_t\\ \psi_t\end{pmatrix} = V\begin{pmatrix}\varphi_t\\ \psi_t\end{pmatrix} \tag{10.3}$$

where $V : \mathscr{H} \to \mathscr{H}$ is given by

$$V = \begin{pmatrix}0 & I\\ -L & 0\end{pmatrix}.$$

The solution to (10.3) is

$$\begin{pmatrix}\varphi_t\\ \psi_t\end{pmatrix} = \mathrm{e}^{tV}\begin{pmatrix}\varphi_0\\ \psi_0\end{pmatrix}.$$

10.5 Proof using the discrete time wave equation

To connect this to our random walks, we would now like to do the same exercise but in discrete time. As in the one-dimensional case, we take

$$U = \begin{pmatrix}M & I\\ -(I - M^2) & M\end{pmatrix},$$

but where now M is the Markov operator and U is an operator on $\mathscr{H}$. This operator is now orthogonal, i.e., it preserves the norm

$$\left\|\begin{pmatrix}\varphi\\ \psi\end{pmatrix}\right\|^2 = \langle \psi, \psi\rangle + \frac{1}{2}\langle \varphi, (I - M^2)\varphi\rangle.$$

We can recover M from U by

$$\begin{pmatrix}M & 0\\ 0 & M\end{pmatrix} = \frac{1}{2}\left(U + U^{-1}\right).$$

Likewise,

$$\begin{pmatrix}M & 0\\ 0 & M\end{pmatrix}^n = \frac{1}{2^n}\left(U + U^{-1}\right)^n = \sum_{k=0}^{n} \frac{1}{2^n}\binom{n}{k} U^k.$$

Hence if we let $\tilde{Z}_n$ be the simple random walk on $\mathbb{Z}$ then

$$\begin{pmatrix}0\\ M^n\varphi\end{pmatrix} = \begin{pmatrix}M & 0\\ 0 & M\end{pmatrix}^n \begin{pmatrix}0\\ \varphi\end{pmatrix} = \mathbb{E}\left[U^{\tilde{Z}_n}\begin{pmatrix}0\\ \varphi\end{pmatrix}\right].$$

Write

$$U^n = \begin{pmatrix}A_n & B_n\\ C_n & D_n\end{pmatrix}$$

then

$$U^{n+1} = \begin{pmatrix}A_n & B_n\\ C_n & D_n\end{pmatrix}\cdot\begin{pmatrix}M & I\\ M^2 - I & M\end{pmatrix} = \begin{pmatrix}A_n M + B_n(M^2 - I) & A_n + B_n M\\ C_n M + D_n(M^2 - I) & C_n + D_n M\end{pmatrix}.$$

It thus follows by induction that A_n, B_n, C_n and D_n are respectively polynomials of degrees n, $n-1$, $n+1$, and n in M (in fact, $A_n = B_n$ is the Chebyshev polynomial of order n). Now, $\langle \delta_g, M^k\delta_0\rangle = 0$ when $|k| < |g|$. Thus also

$$\left\langle \begin{pmatrix}0\\ \delta_g\end{pmatrix}, U^k\begin{pmatrix}0\\ \delta_e\end{pmatrix}\right\rangle = 0$$

for all such k (physically, this means that waves propagate at constant speed). Since U is orthogonal, the above inner product is at most 1 for any k, and so we have that

$$\begin{aligned}\langle \delta_g, M^n\delta_e\rangle &= \left\langle \begin{pmatrix}0\\ \delta_g\end{pmatrix}, \begin{pmatrix}M & 0\\ 0 & M\end{pmatrix}^n\begin{pmatrix}0\\ \delta_e\end{pmatrix}\right\rangle\\ &= \mathbb{E}\left[\left\langle \begin{pmatrix}0\\ \delta_g\end{pmatrix}, U^{\tilde{Z}_n}\begin{pmatrix}0\\ \delta_e\end{pmatrix}\right\rangle\right]\\ &\le \mathbb{P}\left[|\tilde{Z}_n| \ge n\right]\\ &\le 2\mathrm{e}^{-\frac{|g|^2}{2n}},\end{aligned}$$

where the last inequality is simply the Hoeffding bound.

Repeating this proof with $\hat{M} := M/\|M\|$ yields an additional $\|M\|^n$ factor. This completes the proof of Theorem 10.1.

11 The Martin boundary and the Furstenberg-Poisson boundary

11.1 The boundary of the free group

Let $\mathbb{F}_2 = \langle S \rangle$, $S = \{a, a^{-1}, b, b^{-1}\}$ be the free group on two generators. Let $\partial\mathbb{F}_2$ denote the set of *infinite reduced words*:

$$\partial\mathbb{F}_2 = \{s_1 s_2 s_3 \cdots : s_n \in S, s_{n+1} \neq s_n^{-1}\}.$$

We can identify each $b \in \partial\mathbb{F}_2$ with an infinite ray, starting from in the origin of the Cayley graph of $\mathbb{F}_2$.

Given $b \in \partial\mathbb{F}_2$, we say that the k-prefix of b is equal to $g \in \mathbb{F}_2$ if $b = s_1 s_2 \cdots s_k \cdots$ and $g = s_1 s_2 \cdots s_k$. We define the k-prefix of $g \in \mathbb{F}_2$ similarly, provided $|g| \geq k$.

We say that a sequence of words in the free group converges to $b \in \partial\mathbb{F}_2$ if for every k it holds for all n large enough that the k-prefix of g_n is equal to the k-prefix of b. When $\mathscr{F}_2$ is endowed with the discrete topology and $\partial\mathbb{F}_2$ is endowed with the product topology, $\partial\mathbb{F})_2$ is a *compactification* of $\mathbb{F}_2$: every sequence $g_n \in \mathbb{F}_2$ has a subsequence that either converges to some $b \in \partial\mathbb{F}_2$ or to some $g \in \mathbb{F}_2$ (and hence eventually equals this g). Indeed, if we define the distance $d(g, b)$ between two (finite or infinite) reduced words as $3^{-r(g,b)}$ where $r(g, b)$ is the maximum k such that the k-prefixes of the words agree, then $\mathbb{F}_2 \cup \partial\mathbb{F}_2$ is a compact metric space and $\partial\mathbb{F}_2$ is the boundary of the discrete set $\mathbb{F}_2$.

Let μ be the simple random walk, given by the uniform distribution over S. Since the random walk is transient, the first generator in Z_n eventually stabilizes, as does the second, etc. Hence there is a random variable B taking value in $\partial\mathbb{F}_2$ such that Z_n converges to B almost surely. Denote by ν the distribution of B. Then ν is a probability measure on $\partial\mathbb{F}_2$ that is called the *exit measure* of the random walk. The symmetry of the simple random walk makes it is easy to calculate ν: the probability that the k-prefix of B is equal to any particular $s_1 s_2 \cdots s_k$ is $\frac{1}{4}3^{-(k-1)}$.

We can associate with each $b \in \partial\mathbb{F}_2$ the harmonic function given by

$$\psi_b(g) = 3^{-|g| + 2r(g,b)}.$$

Equivalently, viewed as a function on the Cayley graph, ψ_b is the function that is equal to 1 at e, increases by a factor of 3 along edges that tend toward the ray b, and decreases by a factor of 3 in the other direction.

Note that B is a *shift-invariant* random variable: there is a measurable function f such that

$$B = f(Z_n, Z_{n+1}, \ldots)$$

for all n; we can take any f such that $f(g_1, g_2, \ldots) = \lim_n g_n$ whenever the limit exists. It turns out that this is the "universal" shift-invariant random variable: $\sigma(B)$ is the shift-invariant sigma-algebra. In other words, every shift-invariant random variable is a function of B.

What does the random walk look like conditioned on B? The answer turns out to be simple: it is not longer a random walk on G, but it is still a Markov chain, with transition probabilities

$$\mathbb{P}[Z_{n+1} = h | Z_n = g, B = b] = \frac{\psi_b(h)}{\psi_b(g)} \mu(g^{-1}h) = \frac{\psi_b(h)}{\psi_b(g)} \mathbb{P}[Z_{n+1} = h | Z_n = g].$$

That is, relative to the unconditioned random walk, there is a threefold increase in the probability of moving in the direction of B, and a threefold decrease in the probability of moving in each of the opposite three directions. It follows from this that

$$\mathbb{P}[Z_1 = g_1, \ldots, Z_n = g_n | B = b] = \psi_b(g_n) \mathbb{P}[Z_1 = g_1, \ldots, Z_n = g_n].$$

To see why this holds, we first note that this conditioned Markov chain indeed converges to $\lim_n Z_n = b$, since the drift towards b will always eventually bring the random walk back to the ray corresponding to b, and will also push it to infinity, away from the origin. Second, observe that

$$\begin{aligned}
\mathbb{E}[\mathbb{P}[Z_1 = g_1, \ldots, Z_n = g_n | B]] &= \mathbb{E}\left[\psi_B(g_n) \mathbb{P}[Z_1 = g_1, \ldots, Z_n = g_n]\right] \\
&= \mathbb{E}\left[\psi_B(g_n)\right] \mathbb{P}[Z_1 = g_1, \ldots, Z_n = g_n] \\
&= \mathbb{P}[Z_1 = g_1, \ldots, Z_n = g_n],
\end{aligned}$$

since $\mathbb{E}\left[\psi_B(g)\right] = 1$ for all g. This proves that these conditional measures form a collection of conditional measures (also called a *disintegration*) of the unconditional measure with respect to B. Such a collection is almost everywhere uniquely determined, by the disintegration theorem.

The rest of this section will be devoted to building a similar theory for every finitely generated group.

11.2 The stopped random walk

Let G be a finitely generated group and let μ be a finitely supported non-degenerate probability measure on G. We assume that μ has symmetric support: $\mu(g) > 0$ implies $\mu(g^{-1}) > 0$.

Let $(Z_0, Z_1, \ldots)$ be the μ-random walk on G. Given a subset$F \subset G$ that includes e, we define the F-*stopped random walk* $(\mathring{Z}_0, \mathring{Z}_1, \ldots)$ by $\mathring{Z}_0 = e$ and

$$\mathring{Z}_{n+1} = \begin{cases} \mathring{Z}_n X_{n+1} & \text{if } \mathring{Z}_n \in F \\ \mathring{Z}_n & \text{otherwise.} \end{cases}$$

Equivalently, let

$$T = \min\{n \geq 0 : Z_n \notin F\},$$

be the first time that the random walk visits an element that is not in F (and hence in ∂F), and let

$$\mathring{Z}_n = \begin{cases} Z_n & \text{if } n \le T \\ Z_T & \text{otherwise.} \end{cases}$$

We say that F is *connected* if for all $g \in F \cup \partial F$ there is an n such that $\mathbb{P}\left[\mathring{Z}_n = g\right] > 0$. Equivalently, the restriction of the Cayley graph to F has a single connected component (since the support of μ is symmetric). We will henceforth assume that F is connected.

Claim 11.1. *If F is finite then T is almost surely finite.*

In cases in which T is finite (such as finite F), $\mathring{Z}_\infty := Z_T = \lim_n \mathring{Z}_n$ is the element of the complement of F that is first visited by the random walk. Since the random walk starts in F (i.e., $e \in F$) then $\mathring{Z}_\infty \in \partial F$.

11.3 Harmonic functions

Suppose that F is connected. We say that a function $\varphi \colon F \cup \partial F$ is μ-harmonic if for every $g \in F$ it holds that $\varphi(g) = \sum_s \mu(s)\varphi(gs)$. Denote by $\ell_\mu(F)$ the collection of μ-harmonic functions on $F \cup \partial F$:

$$\ell_\mu(F) = \left\{\varphi \colon F \cup \partial F \to \mathbb{R} : \varphi(g) = \sum_s \mu(s)\varphi(gs) \text{ for all } g \in F\right\}.$$

Clearly, $\ell_\mu(F)$ is a linear subspace of $\mathbb{R}^{F \cup \partial F}$.

Claim 11.2. *φ is μ-harmonic if and only if*

$$\varphi(\mathring{Z}_n) = \mathbb{E}\left[\varphi(\mathring{Z}_{n+1}) \middle| \mathring{Z}_n\right]. \tag{11.1}$$

(I.e., $\varphi(\mathring{Z}_n)$ is a martingale).

Proof. For $g \in F$, $\mathring{Z}_{n+1} = \mathring{Z}_n X_{n+1}$, and so

$$\begin{aligned} \mathbb{E}\left[\varphi(\mathring{Z}_{n+1}) \middle| \mathring{Z}_n = g\right] &= \mathbb{E}\left[\varphi(\mathring{Z}_n X_{n+1}) \middle| \mathring{Z}_n = g\right] \\ &= \mathbb{E}\left[\varphi(g X_{n+1}) \middle| \mathring{Z}_n = g\right] \\ &= \sum_s \mathbb{P}\left[X_{n+1} = s\right] \mathbb{E}\left[\varphi(g X_{n+1}) \middle| X_{n+1} = s\right] \\ &= \sum_s \mathbb{P}\left[X_{n+1} = s\right] \mathbb{E}\left[\varphi(gs)\right] \\ &= \sum_s \mu(s)\varphi(gs). \end{aligned}$$

Thus (11.1) holds conditioned on $\mathring{Z}_n = g$ iff φ satisfies the harmonicity condition at g. It remains to be shown that no additional constraints are imposed by (11.1) conditioned on $\mathring{Z}_n \in \partial F$. Indeed, there $\mathring{Z}_n = g$ implies $\mathring{Z}_{n+1} = g$, and so (11.1) holds conditioned on $\mathring{Z}_n = g$ for any φ. $\square$

Claim 11.3. *Fix some $h \in \partial F$. The function*

$$\psi(g) := \mathbb{P}\left[\mathring{Z}_\infty = h \middle| \mathring{Z}_n = g\right]$$

is μ-harmonic.

In the definition of ψ we choose for each g some n such that $\mathbb{P}[Z_n = g] > 0$, and the choice of such n is immaterial (by the Markov property).

Proof of Claim 11.3. Note first that if $g \in \partial F$ then the event $\mathring{Z}_n = g$ is the event $\mathring{Z}_\infty = g$, and thus $\psi(g) = 1$ if $g = h$ and $\psi_h(g) = 0$ if $g \neq h$.

For $g \in F$, we condition on the next step of the random walk to arrive at

$$\begin{aligned}\psi(g) &= \mathbb{P}\left[\mathring{Z}_\infty = h \middle| \mathring{Z}_n = g\right] \\ &= \sum_s \mathbb{P}[X_{n+1} = s]\mathbb{P}\left[\mathring{Z}_\infty = h \middle| \mathring{Z}_n = g, X_{n+1} = s\right] \\ &= \sum_s \mathbb{P}[X_{n+1} = s]\mathbb{P}\left[\mathring{Z}_\infty = h \middle| \mathring{Z}_n = gs\right] \\ &= \sum_s \mu(s)\psi(gs).\end{aligned}$$

In the penultimate equality we used the fact that $g \in F$ to identify the event $\{\mathring{Z}_n = g, X_{n+1} = s\}$ with $\{\mathring{Z}_n = gs\}$. $\square$

Lemma 11.4 (The maximum principle)**.** *Let F be connected, let $\varphi \in \ell_\mu(F)$, and let $\varphi(h) = \max\{\varphi(g) : g \in F \cup \partial F\}$. Then either $h \in \partial F$ or φ is constant.*

Proof. Suppose $h \notin \partial F$, i.e. $h \in F$. We show that φ is constant and equal to $C = \varphi(h) = \max\varphi$.

Fix some n so that $\mathbb{P}\left[\mathring{Z}_n = h\right] > 0$. By harmonicity and (11.1),

$$\mathbb{E}\left[\varphi(\mathring{Z}_{n+k}) \middle| \mathring{Z}_n = h\right] = C$$

for all $k \geq 0$. Since F is connected, for all $g \in F \cup \partial F$ there is a k such that $\mathbb{P}\left[\mathring{Z}_{n+k} = g \middle| \mathring{Z}_n = h\right] > 0$. Therefore, since $\varphi(\mathring{Z}_{n+k}) \leq C$, it follows that $\varphi(g) = C$. $\square$

An implication of the maximum principle is the uniqueness principle:

Lemma 11.5 (The uniqueness principle)**.** *Let F be connected and finite. If $\varphi, \psi \in \ell_\mu(F)$ agree on ∂F then they agree everywhere on $F \cup \partial F$.*

Proof. Suppose that $\varphi, \psi \in \ell_\mu(F)$ agree on ∂F. By the maximum principle, $\varphi - \psi$ is either constant, in which case $\varphi = \psi$, or else it attains its maximum on ∂F. Since it vanishes on ∂F we get that $\varphi \leq \psi$. The same argument applied to $\psi - \varphi$ yields $\psi \leq \varphi$. $\square$

11.4 The Poisson formula

Theorem 11.6 (The Poisson formula)**.** *Suppose that F is finite. Fix some $\hat{\varphi}\colon \partial F \to \mathbb{R}$. Then φ is in $\ell_\mu(F)$ and agrees with $\hat{\varphi}$ on ∂F if and only if*

$$\varphi(g) = \mathbb{E}\left[\hat{\varphi}(\mathring{Z}_\infty)\middle|\mathring{Z}_n = g\right] \tag{11.2}$$

for any n such that $\mathbb{P}\left[\mathring{Z}_n = g\right] > 0$.

Proof. Suppose that φ has the form (11.2). Then clearly φ agrees with $\hat{\varphi}$ on ∂F. Furthermore, for $g \in F$

$$\begin{aligned}
\varphi(g) &= \mathbb{E}\left[\hat{\varphi}(\mathring{Z}_\infty)\middle|\mathring{Z}_n = g\right]\\
&= \sum_s \mathbb{E}\left[\hat{\varphi}(\mathring{Z}_\infty)\middle|\mathring{Z}_n = g, X_{n+1} = s\right]\\
&= \sum_s \mathbb{E}\left[\hat{\varphi}(\mathring{Z}_\infty)\middle|\mathring{Z}_{n+1} = gs\right]\mathbb{P}\left[X_{n+1} = s\right]\\
&= \sum_s \mu(s)\varphi(gs).
\end{aligned}$$

Hence $\varphi \in \ell_\mu(F)$. It then follows from the uniqueness principle that conversely, if $\varphi \in \ell_\mu(F)$ agrees with $\hat{\varphi}$ on ∂F, then it must be of the form (11.2). $\square$

An implication of the Poisson formula is that the map

$$\begin{aligned}
\Phi\colon \mathbb{R}^{\partial F} &\to \ell_\mu(F)\\
\hat{\varphi} &\mapsto \mathbb{E}\left[\hat{\varphi}(\mathring{Z}_\infty)\middle|\mathring{Z}_n = \cdot\right],
\end{aligned} \tag{11.3}$$

is linear bijection. Indeed, its inverse is the restriction map $\varphi \mapsto \hat{\varphi}$.

The map Φ has another important property: it is order preserving. I.e., if $\hat{\varphi} \geq \hat{\psi}$, then $\Phi(\hat{\varphi}) \geq \Phi(\hat{\psi})$. It follows that $\hat{\varphi} \geq 0$ iff $\Phi(\hat{\varphi}) \geq 0$.

Since $\ell_\mu(F)$ is a finite dimensional linear space that contains the constant functions, we can always take a $\varphi \in \ell_\mu(F)$, add a constant to it and multiply it by another constant to arrive at a very similar function that is still in $\ell_\mu(F)$ but is also in

$$\ell_\mu(F,1) := \{\varphi \in \ell_\mu(F) : \varphi \geq 0, \varphi(e) = 1\}.$$

Claim 11.7. *$\ell_\mu(F,1)$ is compact.*

Proof. Clearly $\ell_\mu(F,1)$ is closed. It remains to show that it is bounded. By the Poisson formula, if $\varphi \in \ell_\mu(F,1)$ then $\mathbb{E}\left[\varphi(\mathring{Z}_\infty)\right] = 1$. Hence

$$\sum_{h \in \partial F} \varphi(h)\mathbb{P}\left[\mathring{Z}_\infty = h\right] = 1. \tag{11.4}$$

Hence $\varphi(h) \leq \mathbb{P}\left[\mathring{Z}_\infty = h\right]^{-1}$, and $\varphi \leq \min_h \mathbb{P}\left[\mathring{Z}_\infty = h\right]^{-1}$. $\square$

The set $\ell_\mu(F,1)$ is compact, and furthermore convex. Furthermore, it can be identified with convex combinations of the functions

$$\psi_h = \frac{1}{\mathbb{P}\left[\mathring{Z}_\infty = h\right]}\Phi(\mathbb{1}_{\{h\}}),$$

where $\mathbb{1}_{\{h\}} : \partial F \to \{0,1\}$ is the indicator of $h \in \partial F$. That is, every $\varphi \in \ell_\mu(F,1)$ can be written as

$$\begin{aligned}
\varphi &= \Phi(\hat{\varphi}) \\
&= \Phi(\sum_h \varphi(h)\mathbb{1}_{\{h\}}) \\
&= \sum_h \varphi(h)\mathbb{P}\left[\mathring{Z}_\infty = h\right]\frac{1}{\mathbb{P}\left[\mathring{Z}_\infty = h\right]}\Phi(\mathbb{1}_{\{h\}}) \\
&= \sum_h \varphi(h)\mathbb{P}\left[\mathring{Z}_\infty = h\right]\psi_h \\
&=: \sum_h \lambda_h \psi_h
\end{aligned}$$

where, by (11.4), $\sum_h \lambda_h = 1$. That is, φ is the *barycenter* of the probability measure λ defined on the set $\{\psi_h : h \in B_{n+1}\}$.

The functions $\Phi(\mathbb{1}_{\{h\}})$ are the harmonic functions of the form described in Claim 11.3. The functions $\psi_h = \frac{1}{\mathbb{P}\left[\mathring{Z}_\infty = h\right]}\Phi(\mathbb{1}_{\{h\}})$ are the *extreme points* of $\ell_\mu(F,1)$: these functions cannot be written as non-trivial convex combinations of functions in $\ell_\mu(F,1)$.

The constant function on $F \cup \partial F$ is

$$1 = \sum_h \mathbb{P}\left[\mathring{Z}_\infty = h\right]\psi_h.$$

Let ν be a probability measure on the collection $(\psi_h)_h$ given by $\nu(\psi_h) = \mathbb{P}\left[\mathring{Z}_\infty = h\right]$. This is called the *exit measure* of the stopped random walk. By definition,

$$\sum_h \nu(h)\psi_h(g) = 1$$

for all $g \in F$. Note that $\ell_\mu(F,1)$ is a *simplex*: there is a unique way of representing each of each elements as a convex combination of the extreme points. Thus ν is the unique probability measure on $(\psi_h)_h$ for which the above holds.

11.5 The Martin boundary

Fix a finitely supported, non-degenerate μ with symmetric support S so that $G = \langle S \rangle$. Using our notation $\ell_\mu(G)$ is the set of μ-harmonic functions on G, and $\ell_\mu(G,1)$ are the non-negative ones that assign 1 to the identity. We endow $\mathbb{R}^G$ with the topology of pointwise convergence, which is also the product topology. I.e., a sequence of functions $\varphi_n : G \to \mathbb{R}$ converges to φ if $\lim_n \varphi_n(g) = \varphi(g)$ for all $g \in G$, in which case we write $\lim_n \varphi_n = \varphi$.

Clearly, both $\ell_\mu(G)$ and $\ell_\mu(G,1)$ are closed subsets of $\mathbb{R}^G$. The next proposition implies that the latter is compact.

Proposition 11.8. *For every $g \in G$ and $\varphi \in \ell_\mu(G,1)$ it holds that*

$$\sup_n \mathbb{P}\left[Z_n = g^{-1}\right] \leq \varphi(g) \leq \inf_n \frac{1}{\mathbb{P}[Z_n = g]}.$$

Proof. Since φ is harmonic, $(\varphi(hZ_0), \varphi(hZ_1), \dots)$ is a martingale for any $h \in G$. Hence

$$\varphi(h) = \mathbb{E}\left[\varphi(hZ_n)\right] = \sum_{k \in G} \varphi(k)\mathbb{P}[hZ_n = k] \geq \varphi(k)\mathbb{P}[hZ_n = k],$$

and so we have the right inequality by setting $h = e$ and $k = g$. For the left inequality, set $h = g$ and $k = e$. □

An immediate corollary of this proposition is that $\ell_\mu(G,1)$ is compact, since it is closed and contained in the product of compact sets, which is compact.

Let B_n be the ball of radius n in G. Identify each $\varphi \in \ell_\mu(B_n,1)$ with the function in $\mathbb{R}^G$ that agrees with φ on $F \cup \partial F$ and vanishes elsewhere. That is, we now redefine

$$\ell_\mu(B_n,1) = \left\{\varphi \colon G \to \mathbb{R} : \varphi(g) = \sum_s \mu(s)\varphi(gs) \text{ for all } g \in B_n \text{ and } \operatorname{supp}\varphi(g) \subseteq B_{n+1}\right\}.$$

Thus $\ell_\mu(B_n,1)$ is a subset of $\mathbb{R}^G$.

Proposition 11.9. *For every g there is a constant C_g such that for every n and every $\varphi \in \ell_\mu(B_n,1)$ it holds that $\varphi(g) \leq C_g$.*

The proof is similar to that of Proposition 11.8. This implies that the set $\{\psi_h : h \in G\}$, which we identify with G, is precompact: its closure is compact, or, alternatively, every sequence in it has a converging subsequence (even if the limit may not be in G).

Suppose that a sequence $\varphi_n \in \ell_\mu(B_n,1)$ converges pointwise to $\varphi \in \mathbb{R}^G$. Then $\varphi \in \ell_\mu(G,1)$, since clearly $\varphi(e) = 1$ and since at each g the harmonicity condition is satisfied for all n large enough. Conversely, let

$$\begin{aligned} \pi_n \mathbb{R}^G &\to \mathbb{R}^G \\ \varphi &\mapsto \varphi \cdot \mathbb{1}_{\{B_n\}} \end{aligned}$$

be the natural projection to functions supported on the ball of radius n, and note that $\lim_n \pi_n(\varphi) = \varphi$ for any $\varphi \in \mathbb{R}^G$. If $\varphi \in \ell_\mu(G,1)$, then the projection $\varphi_n = \pi_n(\varphi)$ is in $\ell_\mu(B_{n-1},1)$. Since $\lim_n \varphi_n = \varphi$, $\ell_\mu(G,1)$ is the limit of the *sets* $\ell_\mu(B_n,1)$.

An element of $\ell_\mu(G,1)$ is an *extreme point* if it cannot be written as a non-trivial convex combination of two other functions in $\ell_\mu(G,1)$. The topological closure of the set of extreme points of $\ell_\mu(1)$ is called the *Martin boundary* of G with respect to μ, and we will denote it by $\partial_\mu G$.

The reason that $\partial_\mu G$ is called a *boundary* of G is that, if we identify g with $\psi_g \in \ell_\mu(B_{|g|},1)$ then $\partial_\mu G$ is a compactification of G:

Proposition 11.10. *The Martin boundary $\partial_\mu G$ is the set of limit points of G in $\mathbb{R}^G$, and $G \cup \partial_\mu G$ is compact.*

Proof. By Proposition 11.9, every sequence in G has a converging subsequence. Thus the union of G with its limit points is compact, and it remains to be shown that the set of limit points of G is equal to $\partial_\mu G$.

To see that the set of limits points in G contains $\partial_\mu G$, fix an extreme point $\psi \in \ell_\mu(G,1)$, and denote $\psi_n = \pi_n \psi$. By the Poisson formula we can write each ψ_n as the barycenter of a probability measure λ_n on $G \cup B_n$: $\psi_n(g) = \sum_{h \in B_n} \lambda_n(h)\psi_h(g)$.

This sequence of probability measures will have a converging subsequence, which will converge to some probability measure λ on $\ell_\mu(G,1)$ with barycenter ψ. But since ψ is extreme, this measure must be a point mass at ψ, which is thus a limit point of G.

In the other direction, suppose φ is not in $\partial_\mu G$. Then there exists a finite set $F \subset G$ and $\varepsilon > 0$ such that every φ' with $|\varphi'(g) - \varphi(g)| < \varepsilon$ for all $g \in F$ is not extreme. In particular, φ is in the interior of $\ell_\mu(G,1)$, and furthermore φ is in the interior of $\pi_n \ell_\mu(G,1)$ for all n large enough. Thus the interior of $\ell_\mu(G,1)$ is equal to the union of these interiors. Now, G is disjoint from this set, since each ψ_h, is not in any $\pi_n \ell_\mu(G,1)$: for $n < |h|$ the support of ψ_h is too big, and for $n \geq |h|$ the maximum principle is violated. Thus there are no limits points of G in the interior of $\ell_\mu(G,1)$, and they are all contained in $\partial_\mu G$. □

11.6 Bounded harmonic functions

Denote by $\ell^\infty_\mu(G)$ the set of bounded harmonic functions. Let $\mathscr{I}$ be the shift-invariant sigma-algebra of $(Z_0, Z_1, \ldots)$. Recall that a random variable W is measurable with respect to $\mathscr{I}$ if there is some f such that

$$W = f(Z_1, Z_2, \ldots) = f(Z_2, Z_3, \ldots) = f(Z_n, Z_{n+1}, \ldots).$$

An example of a shift-invariant event is the event that $Z_n \in P$ eventually, for some $P \subseteq G$:

$$\{\exists N \text{ s.t. } Z_n \in P \text{ for all } n \geq N\}.$$

We denote by $L^\infty(\mathscr{I})$ the collection of bounded, $\mathscr{I}$-measurable random variables. To each shift-invariant bounded random variable W we can associate the bounded harmonic function $\varphi = \Phi(W)$ given by

$$\varphi(g) = \mathbb{E}[W | Z_n = g],$$

for some (any) n such that $\mathbb{P}[Z_n = g] > 0$. It is simple to check that φ is indeed bounded harmonic. Conversely, to each $\varphi \in \ell^\infty_\mu(G)$ we can assign the $W \in L^\infty(\mathscr{I})$ given by

$$W = \lim_n \varphi(Z_n).$$

The limit exists because $\varphi(Z_n)$ is a bounded martingale, and hence converges.

Indeed, in analogy to (11.3), define

$$\begin{aligned}\Phi\colon L^\infty(\mathscr{I}) &\to \ell^\infty_\mu(G)\\ W &\mapsto \mathbb{E}[W|Z_n=\cdot].\end{aligned}$$

This map is sometimes called the *Furstenberg transform*.

Note that both $\ell^\infty_\mu(G)$ and $L^\infty(\mathscr{I})$ are normed vector spaces when equipped with the supremum norm:

$$\begin{aligned}\|W\|_\infty &= \sup\{x\in\mathbb{R}_+ : \mathbb{P}[|W|\geq x]>0\}\\ \|\varphi\|_\infty &= \sup_g |\varphi(g)|.\end{aligned}$$

It turns out that Φ is not just a bijection between these vector spaces, but moreover preserves these norms.

Proposition 11.11. *The map Φ is an isometry between $L^\infty(\mathscr{I})$ and $\ell^\infty_\mu(G)$.*

Proof. Since $\mathbb{E}[W|Z_n=g]\leq\|W\|_\infty$, $\|\Phi(W)\|_\infty\leq\|W\|_\infty$. In the other direction, given $\varphi\in\ell^\infty_\mu(G)$, the process $W_n=\varphi(Z_n)$ is a bounded martingale and hence converges to $W=\lim_n W_n=\lim_n\varphi(Z_n)$, and W is easily seen to be a shift-invariant random variable. Now,

$$\mathbb{E}\left[\lim_n \varphi(Z_n)\middle| Z_n=g\right]=\varphi(g)$$

by the martingale property of $\varphi(Z_n)$ and the Markov property of Z_n. Thus the map $\varphi\mapsto W$ is the inverse of Φ. Furthermore, $W=\lim_n\varphi(Z_n)\leq\|\varphi\|_\infty$, and so $\|W\|_\infty\leq\|\Phi(W)\|_\infty$. Thus $\|\Phi(W)\|_\infty=\|\varphi\|_\infty$. □

It follows from Proposition 11.11 that if there are no non-constant bounded μ-harmonic functions then the shift-invariant sigma-algebra is trivial: every shift-invariant random variable is constant.

Another consequence of Proposition 11.11 is the following claim. In this statement we identify two events if their symmetric difference has zero measure; equivalently, if their indicators coincide as random variables.

Claim 11.12. *Every shift-invariant event is of the form $Z_n\in P$ eventually, for some $P\subseteq G$.*

Proof. Let $E\in\mathscr{I}$ be a shift-invariant event, and let W be its indicator. Let $\varphi=\Phi(W)$. Since $W=\Phi^{-1}(\varphi)=\lim_n\varphi(Z_n)$, W is the indicator of the event that $\lim_n\varphi(Z_n)=1$.

Let $P=\{g\in G : \varphi(g)>1/2\}$. Then $\lim_n\varphi(Z_n)=1$ iff Z_n is in P for all n large enough. Hence W is also the indicator of the event that Z_n is eventually in P. □

Recall that for each $h\in G$ we defined the right translation linear operator $R_h\colon\mathbb{R}^G\to\mathbb{R}^G$

$$[R_h\varphi](g)=\varphi(gh).$$

We now define the *left translation* operator $L_h : \mathbb{R}^G \to \mathbb{R}^G$ by

$$[L_h\varphi](g) = \varphi(h^{-1}g).$$

As with right translations, this is a representation of G: $L_h L_g = L_{hg}$ and $L_h^{-1} = L_{h^{-1}}$. Will will now be interested in L because is preserves harmonicity. To see this, note that L commutes with R:

$$[L_g R_h\varphi](k) = [R_h\varphi](g^{-1}k) = \varphi(g^{-1}kh) = [L_g\varphi](kh) = [R_h L_g\varphi](k).$$

Since $M = \sum_h \mu(h)R_h$, it follows that L and M commute, and so if $M\varphi = \varphi$ then $M(L\varphi) = LM\varphi = L\varphi$.

The following theorem is known as the Choquet-Deny Theorem, even though it was first proved by David Blackwell. The proof below is due to Margulis.

Theorem 11.13. *Suppose that G is abelian. Then for any μ, every bounded μ-harmonic function is constant.*

To prove this theorem we will need an important result about compact convex sets.

Theorem 11.14 (Krein-Milman Theorem)**.** *Let C be a compact convex subset of a nice topological vector space.*[1] *Then every $c \in C$ is the limit of convex combinations of the extreme points of C.*

Proof of Theorem 11.13. Let $C \subset \ell^\infty_\mu(G)$ be the bounded harmonic functions that take values in $[0,1]$. This is a compact convex set (in the topology of pointwise convergence) and thus by the Krein-Milman theorem has extreme points. Suppose $\psi \in C$ is extreme. Since it is harmonic,

$$\psi = M\psi = \sum_h \mu(h)R_h\psi.$$

Since G is abelian, $R_h\psi = L_{h^{-1}}\psi$, and so

$$\psi = \sum_h \mu(h)L_{h^{-1}}\psi.$$

Now, each $L_{h^{-1}}\psi$ is also in C. Hence we have written ψ as a convex combination of elements of C. But ψ is harmonic, and so $L_{h^{-1}}\psi = \psi$ for all $h^{-1} \in \operatorname{supp}\mu$. Since $\operatorname{supp}\mu$ generates G, we write any $g \in G$ as a product $g = h_1 h_2 \cdot h_n$ of elements of $\operatorname{supp}\mu$. We then have that $L_{g^{-1}}\psi = \psi$. In particular $\psi(g) = \psi(e)$ and ψ is constant. Thus all extreme points in C are constant. And since, again by Krein-Milman, every $\varphi \in C$ is the limit of convex combinations of extreme points, every $\varphi \in C$ is constant. Hence every $\varphi \in \ell^\infty_\mu(G)$ is constant. □

[1]By nice we mean Hausdorff and locally convex. We will only need that $\mathbb{R}^G$ (equipped with pointwise convergence) is nice.

12 Random walk entropy and the Kaimanovich-Vershik Theorem

In this section, as usual, we consider a finitely supported, non-degenerate μ on a finitely generated $G = \langle S \rangle$.

12.1 Random walk entropy

Claim 12.1. $H(Z_{n+m}) \le H(Z_n) + H(Z_m)$.

Proof.

$$Z_{n+m} = (X_1 \cdots X_n) \cdot (X_{n+1} \cdots X_{n+m}),$$

and so

$$H(Z_{n+m}) \le H(X_1 \cdots X_n, X_{n+1} \cdots X_{n+m}).$$

These two random variables are independent, and so

$$H(Z_{n+m}) \le H(X_1 \cdots X_n) + H(X_{n+1} \cdots X_{n+m}).$$

The distribution of $Z_m = X_1 \cdots X_m$ is identical to that of $X_{n+1} \cdots X_{n+m}$, and so

$$H(Z_{n+m}) \le H(Z_n) + H(Z_m).$$

□

This claim shows that the sequence $H(Z_n)$ is subadditive. It thus follows from Fekete's Lemma (Lemma 7.3) that $\frac{H(Z_n)}{n}$ converges. We accordingly define the *random walk entropy* $h(\mu)$ by

$$h(\mu) = \lim_{n \to \infty} \frac{1}{n} H(Z_n).$$

Note that $\frac{1}{n} H(Z_n) \le \frac{1}{n} H(X_1, \ldots, X_n) = H(X_1)$, and thus $h(\mu)$ is finite.

12.2 The Kaimanovich-Vershik Theorem

Theorem 12.2. *The random walk $(Z_0, Z_1, Z_2, \ldots)$ has a trivial tail sigma-algebra if and only if $h(\mu) = 0$.*

Proof. We calculate the mutual information $I(Z_1; \mathscr{T})$, where $\mathscr{T}$ is the tail sigma-algebra. Recall that $\mathscr{T} = \cap_n \mathscr{T}_n$, where $\mathscr{T}_n = \sigma(Z_n, Z_{n+1}, \ldots)$. Hence, by Claim A.4,

$$H(Z_1 | \mathscr{T}) = \lim_n H(Z_1 | Z_n, Z_{n+1}, \ldots).$$

By the Markov property it follows that

$$H(Z_1|\mathscr{T}) = \lim_n H(Z_1|Z_n).$$

By (A.1)

$$H(Z_1|\mathscr{T}) = \lim_n H(Z_n|Z_1) - H(Z_n) + H(Z_1).$$

Now, $Z_1 = X_1$, and $Z_n = X_1 \cdots X_n$, and so

$$H(Z_1|\mathscr{T}) = \lim_n H(X_1 \cdots X_n|X_1) - H(Z_n) + H(Z_1).$$

Note that conditioned on $X_1 = g$, the distribution of $X_1 \cdots X_n$ is identical to the distribution of $gX_1 \cdots X_{n-1}$, which has the same entropy as $X_1, \ldots, X_{n-1} = Z_{n-1}$. Hence $H(X_1 \cdots X_n|X_1) = H(Z_{n-1})$, and we get that

$$H(Z_1|\mathscr{T}) = \lim_n H(Z_{n-1}) - H(Z_n) + H(Z_1).$$

Thus

$$I(Z_1;\mathscr{T}) = \lim_n H(Z_n) - H(Z_{n-1}) = h(\mu).$$

It follows that if $h(\mu) > 0$ then $\mathscr{T}$ is not independent of Z_1, and in particular $\mathscr{T}$ is non-trivial.

For the other direction, a calculation similar to the one above shows that $I(Z_1, \ldots, Z_n;\mathscr{T}) = nh(\mu)$. Thus, if $h(\mu) = 0$, then $\mathscr{T}$ is independent of $(Z_1, \ldots, Z_n)$ for all n, and, as in the proof of Kolmogorov's zero-one law, is trivial. □

We say that G has *subexponential growth* if $\mathrm{GR}(G) = 0$. That is, if $\lim_r \frac{1}{r} \log|B_r| = 0$; see (7.1).

Corollary 12.3. *If G has subexponential growth then $\mathscr{T}$ is trivial.*

Proof. Since Z_n is supported on B_r, $H(Z_n) \leq \log|B_n|$. Hence

$$h(\mu) = \lim_n \frac{1}{n} H(Z_n) \leq \lim_n \frac{1}{n} \log|B_n|.$$

Hence if G is subexponential then $h(\mu) = 0$ and $\mathscr{T}$ is trivial. □

A Basics of information theory

A.1 Shannon entropy

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X be a (simple) random variable taking values in some finite set Θ. We define the *Shannon entropy* of X by

$$H(X) = -\sum_{\theta \in \Theta} \mathbb{P}[X = \theta] \log \mathbb{P}[X = \theta],$$

where we use the convention $0 \log 0 = 0$.

Denote by $\mathbb{P}[X]$ the random variable given by $\mathbb{P}[X](\omega) = \mathbb{P}[X = X(\omega)]$. Then we can write the entropy as

$$H(X) = \mathbb{E}[-\log \mathbb{P}[X]].$$

The first important property of Shannon entropy is the following form of monotonicity:

Claim A.1. *Let X, Y be simple random variables. Suppose Y is $\sigma(X)$-measurable (i.e., $Y = f(X)$ for some function f). Then $H(Y) \leq H(X)$.*

Proof. Note that $\mathbb{P}[Y] \leq \mathbb{P}[X]$ almost surely. Hence

$$H(Y) = \mathbb{E}[-\log \mathbb{P}[Y]] \leq \mathbb{E}[-\log \mathbb{P}[X]] = H(X).$$

$\square$

Given two random variables X and X' taking values in Θ, Θ', we can consider the pair (X, X') as a single random variable taking values in $\Theta \times \Theta'$. We denote the entropy of this random variable as $H(X, X')$. The second important property of Shannon entropy is additivity with respect to independent random variables.

Claim A.2. *Let X, Y be independent simple random variables. Then $H(X, Y) = H(X) + H(Y)$.*

Proof. By independence, $\mathbb{P}[X, Y] = \mathbb{P}[X] \cdot \mathbb{P}[Y]$. Hence

$$H(X, Y) = \mathbb{E}[-\log \mathbb{P}[X, Y]] = \mathbb{E}[-\log \mathbb{P}[X] - \log \mathbb{P}[Y]] = H(X) + H(Y).$$

$\square$

A.2 Conditional Shannon entropy

Let $\mathscr{G}$ be a sub-sigma-algebra of $\mathscr{F}$. For a simple random variable X, define the random variable $\mathbb{P}[X|\mathscr{G}](\omega) = \mathbb{P}[X = X(\omega)|\mathscr{G}](\omega)$, and denote the conditional Shannon entropy by

$$H(X|\mathscr{G}) = \mathbb{E}[-\log \mathbb{P}[X|\mathscr{G}]].$$

For a simple random variable X and any random variable Y, we denote $H(X|Y) = H(X|\sigma(Y))$.

Claim A.3. *$H(X|\mathscr{G}) \le H(X)$, with equality if and only if X is independent of $\mathscr{G}$.*

Proof. By the law of total expectation, $\mathbb{P}[X|\mathscr{G}] = \mathbb{E}[\mathbb{P}[X]|\mathscr{G}]$. Since $x \mapsto -\log(x)$ is a convex function, it follows from Jensen's inequality that

$$\begin{aligned} H(X|\mathscr{G}) &= \mathbb{E}[-\log \mathbb{P}[X|\mathscr{G}]] \\ &= \mathbb{E}[-\log \mathbb{E}[\mathbb{P}[X]|\mathscr{G}]] \\ &\le \mathbb{E}[\mathbb{E}[-\log \mathbb{P}[X]|\mathscr{G}]] \\ &= \mathbb{E}[-\log \mathbb{P}[X]] \\ &= H(X). \end{aligned}$$

When X is independent of $\mathscr{G}$, $\mathbb{P}[X] = \mathbb{P}[X|\mathscr{G}]$, and we therefore have equality. It thus remains to be shown if X is not independent of $\mathscr{G}$ then the inequality is strict. Indeed, in that case $\mathbb{P}[X] \neq \mathbb{P}[X|\mathscr{G}]$ with positive probability, and thus Jensen's inequality is strict with positive probability, from which it follows that our inequality is also strict. □

The same proof shows more generally that if $\mathscr{G}_1 \subseteq \mathscr{G}_2$ then $H(X|\mathscr{G}_1) \ge H(X|\mathscr{G}_2)$.

Claim A.4. *Suppose $\mathscr{G} = \cap_{i=n}^{\infty} \mathscr{G}_n$, and $\mathscr{G}_{n+1} \subseteq \mathscr{G}_n$. Then*

$$H(X|\mathscr{G}) = \lim_n H(X|\mathscr{G}_n) = \sup_n H(X|\mathscr{G}_n).$$

A.3 Mutual information

We denote the *mutual information* of X and $\mathscr{G}$ by $I(X;\mathscr{G}) = H(X) - H(X|\mathscr{G})$. By the above, I is non-negative, and is equal to 0 if and only if X is independent of $\mathscr{G}$. For two random variables X, Y, we denote $I(X;Y) = I(X;\sigma(Y))$.

Claim A.5. *Let X, Y be simple random variables. Then*

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = I(Y;X).$$

Proof. By definition,

$$I(X;Y) = \mathbb{E}[-\log \mathbb{P}[X]] - \mathbb{E}[-\log \mathbb{P}[X|Y]]$$

By Bayes' Law, $\mathbb{P}[X|Y]\mathbb{P}[Y] = \mathbb{P}[X,Y]$. Hence $\log \mathbb{P}[X|Y] = \log \mathbb{P}[X,Y] - \log \mathbb{P}[Y]$, and

$$\begin{aligned} I(X;Y) &= \mathbb{E}[-\log \mathbb{P}[X]] - \mathbb{E}[-\log \mathbb{P}[X,Y] + \log \mathbb{P}[Y]] \\ &= \mathbb{E}[-\log \mathbb{P}[X]] - \mathbb{E}[-\log \mathbb{P}[X,Y]] + \mathbb{E}[-\log \mathbb{P}[Y]] \\ &= H(X) - H(X,Y) + H(Y). \end{aligned}$$

□

It follows that

$$H(X|Y) = H(X) - I(X;Y) = H(X) - I(Y;X) = H(X) + H(Y|X) - H(Y),$$

and so

$$H(X|Y) = H(Y|X) - H(Y) + H(X). \tag{A.1}$$

A.4 The information processing inequality

Let $X_1, X_2, X_3, \ldots$ be a Markov chain, with each X_n simple.

Claim A.6. *$I(X_3; X_1, X_2) = I(X_3; X_2)$. Likewise, for $m > n$, $I(X_n; \sigma(X_m, X_{m+1}, \ldots)) = I(X_n; X_m)$.*

The claim is a consequence of the fact that by the Markov property, $\mathbb{P}[X_3 | X_1, X_2] = \mathbb{P}[X_3 | X_2]$.

B Exercises

1. Let $(X_1, X_2, \ldots)$ be a sequence of independent (but not necessarily identically distributed) integer random variables with $\mathbb{E}[X_n] = 0$ and $|X_n| \leq M$ almost surely for all n and some M. Let $Z_n = X_1 + \cdots + X_n$. Prove a strong law of large numbers, i.e., $\frac{1}{n} \lim_n Z_n = 0$ almost surely.

   Hint. Use the Hoeffding lemma (Lemma 1.4).

2. Let μ be a finitely supported distribution on $\mathbb{Z}^d$ for some $d \geq 1$, and let $(Z_1, Z_2, \ldots)$ be the μ-random walk on $\mathbb{Z}^d$. I.e., $(X_1, X_2, \ldots)$ are i.i.d. μ and $Z_n = X_1 + \cdots + X_n$.

   Using the SLLN for $\mathbb{Z}$ (Theorem 1.6), prove a strong law of large numbers, i.e., $\lim_n \frac{1}{n} Z_n = \mathbb{E}[Z_1]$ almost surely.

   Hint. for $i \in \{1, \ldots, d\}$ consider the projection $\pi_i(x_1, \ldots, x_d) = x_i$ and the process $(Z_1^i, Z_2^i, \ldots)$ given by $Z_n^i = \pi_i(Z_n)$. Prove that $(Z_1^i, Z_2^i, \ldots)$ is a random walk on $\mathbb{Z}$ and use the SLLN for $\mathbb{Z}$.

3. Let Z_n be a μ-random walk on $\mathbb{Z}$ with drift $\alpha = \mathbb{E}[Z_1]$. Prove that for every $\beta > \alpha$ and every $\gamma > \beta$ with $\beta, \gamma < \max \operatorname{supp} \mu$ there is an $r > 0$ such that

$$\lim_n \mathbb{P}\left[Z_n \leq \gamma n \middle| Z_n \geq \beta n\right] \geq 1 - \mathrm{e}^{-rn + o(n)}.$$

4. Let μ be a non-degenerate, finitely supported probability measure on $\mathbb{Z}$ (i.e., for all $x \in \mathbb{Z}$ there exists an n such that $\mu^{(n)}(x) > 0$). Let F be a finite subset of $\mathbb{Z}$. Suppose that $\varphi(x) = \varphi(y)$ for all $x, y \notin F$, and that $\varphi \colon \mathbb{Z} \to \mathbb{R}$ is μ-harmonic at all $x \in F$ (i.e., $\varphi(x) = \sum_y \varphi(x+y)\mu(y)$). Prove that φ is constant.

5. Prove Claim 3.9 from the lecture notes.

   Hint. Define $\varphi(x) = \mathbb{P}[\{x + Z_0, x + Z_1, x + Z_2, \ldots\} \subset F]$ and use (4).

6. Let Z_n be a μ-random walk on $\mathbb{Z}$ with drift $\mathbb{E}[Z_1] = 0$. For $M > 0$, let A_n^M be the event that $Z_n \geq \sqrt{n} M$. Prove that for every M, the probability of $(A_n^M)_n$ i.o. is 1.

   Hint. Use the Central Limit Theorem and the fact that $\limsup_n Z_n / \sqrt{n}$ is a tail random variable with respect to $(X_1, X_2, \ldots)$.

7. Let μ_0 be the simple random walk on $\mathbb{Z}$, let $\mu = \mu_0 \times \cdots \times \mu_0$ be the product measure on $\mathbb{Z}^d$, and let Z_n be the μ-random walk on $\mathbb{Z}^d$. Let

$$P = \{(z_1, \ldots, z_d) \in \mathbb{Z}^d : z_1 > 0, \ldots, z_d > 0\} \subset \mathbb{Z}^d$$

   be the positive octant in $\mathbb{Z}^d$. Show that

   (a) $\lim_n \mathbb{P}[Z_{n+1} \in P | Z_n \in P] = 1$.

   (b) $\mathbb{P}[Z_n \in P \text{ for all } n \text{ large enough}] = 0$.

Hint. Use the Central Limit Theorem for $\mathbb{Z}$ for the first part. Use the recurrence of the simple random walk on $\mathbb{Z}$ for the second.

8. Recall that the lamplighter group $\bigoplus_{\mathbb{Z}} \mathbb{Z}_2 \rtimes \mathbb{Z}$ is generated by $\{(0,1),(0,-1),(\delta_0,0)\}$. Consider the random walk on this group given by $\mu(0,1) = 1/3$, $\mu(0,-1) = 1/6$ and $\mu(\delta_0,0) = 1/2$: the lamplighter moves right with probability 1/3, left with probability 1/6, and flips the lamp at the current location with probability 1/2. Find a non-trivial event in the tail of the μ-random walk $(Z_1, Z_2, \ldots)$.

   Hint. Write each Z_n as a pair $Z_n = (F_n, \tilde{Z}_n)$ where F_n takes values in $\bigoplus_{\mathbb{Z}} \mathbb{Z}_2$ and $\tilde{Z}_n$ takes values in $\mathbb{Z}$. Show that $F_n(0)$ converges almost surely and is a non-trivial tail random variable.

9. Prove that the simple random walk on the infinite dihedral group is recurrent. This is the group generated by $\{a, a^{-1}, b\}$ where $a, b \colon \mathbb{Z} \to \mathbb{Z}$ are given by $a(z) = z+1$ and $b(z) = -z$. The simple random walk is given by $\mu(a) = \mu(a^{-1}) = \mu(b) = 1/3$.

   Hint. Draw the Cayley graph of this group and relate this random walk to a symmetric random walk on $\mathbb{Z}$.

10. Let $S = \{a, a^{-1}, b, b^{-1}\}$ be the standard generating set of the free group on two generators. Let μ be a measure whose support is equal to S (so that, in particular, μ is non-degenerate), and let Z_n be the μ-random walk.

    (a) Suppose that $\mu(s) < 1/2$ for all $s \in S$. Show that Z_n is transient.
    Hint. Let $p = \max_{s \in S} \mu(s)$ and let $\beta = (1-p)/p$. Show that $\varphi(g) = \beta^{-|g|}$ is a positive non-constant μ-superharmonic function on $\mathbb{F}_2$ and deduce that the random walk is transient from Theorem 5.1.

    (b) Suppose that $\mu(s) \geq 1/2$ for some $s \in S$. Show that Z_n is transient.
    Hint. Suppose that $\mu(a) \geq 1/2$. Consider the quotient $\pi \colon \mathbb{F} \to \mathbb{Z}$ given by $\pi(a) = 1$, $\pi(b) = 0$ and $\pi(gh) = \pi(g) + \pi(h)$. This is the map that sums the number of occurrences of a minus the number of occurrences of a^{-1} in a word of the free group. Show that the $\pi_*\mu$-random walk on $\mathbb{Z}$ is transient, and conclude that so is the μ-random walk on $\mathbb{F}$.

    (c) Let $\tau(g) = \mathbb{P}[\exists n \geq 0 : Z_n = g]$. Show that for $g \in S$

    $$\tau(g) = \mu(g) + \tau(g) \sum_{s \in S \setminus \{g\}} \tau(s^{-1})\mu(s).$$

    Hint. Follow the calculation for the simple random walk in §5.3.

11. Prove Claim 7.1 from the lecture notes. Use it to prove that the exponential growth rate of a finitely generated group vanishes for one generating set if and only if it does for another.

12. Prove (8.2).

13. Let M be the Markov operator of a symmetric non-degenerate probability measure μ on a finitely generated group G. Suppose that $\mu(e) > 0$. Show that for every $g \in G$

$$\|M\| = \lim_n \mathbb{P}[Z_n = g]^{1/n}.$$

Hint. Approximate $\mathbb{P}[Z_n = g]$ by $\mathbb{P}[Z_{2m} = e]$ for m close to $n/2$ and apply Theorem 8.3.

14. Let $G = \langle S \rangle = \langle T \rangle$. Let F_n be a sequence of finite subsets of G. Show that

$$\lim_n \frac{|\partial_S F_n|}{|F_n|} = 0 \qquad \text{iff } \lim_n \frac{|\partial_T F_n|}{|F_n|} = 0.$$

15. Let $G = \langle S \rangle$ be a finitely generated group, and let $S = \{s_1, \ldots, s_k\}$. We call $\mathbb{Z}_2 \wr G = \bigoplus_G \mathbb{Z}_2 \rtimes G$ the lamplighter group on G. An element of this group is a pair (f, x) where $f : G \to \mathbb{Z}_2$ is finitely supported and $x \in G$. As in the case of $G = \mathbb{Z}$, the operation is given by

$$(f_1, x_1)(f_2, x_2) = (f_1 + \alpha_{x_1}(f_2), x_1 \cdot x_2),$$

where $\alpha_x : \bigoplus_G \mathbb{Z}_2 \to \bigoplus_G \mathbb{Z}_2$ is the shift

$$[\alpha_x(f)](y) = f(x^{-1}y).$$

(a) Show that $\mathbb{Z}_2 \wr G$ is generated by

$$S_d = \Big\{(\delta_0, 0), (0, s_1), \ldots, (0, s_k)\Big\}.$$

(b) Show that if G is amenable then $\mathbb{Z}_2 \wr G$ is amenable.
Hint. Use a Følner sequence on G to construct a Følner sequence on $\mathbb{Z}_2 \wr G$.

(c) Show that if G is non-amenable then $\mathbb{Z}_2 \wr G$ is non-amenable.
Hint. Project a random walk on $\mathbb{Z}_2 \wr G$ to a random walk on G via $(f, x) \mapsto x$ and argue that the return probabilities of the latter are higher than those of the former. Then use Kesten's theorem (Theorem 9.3).

16. Let μ be a symmetric, finitely supported, non-degenerate probability measure on a finitely generated group $G = \langle S \rangle$ with $\operatorname{supp} \mu = S$. Let M be the associated Markov operator.

As in (10.1), the *energy* of a map $\varphi \ell^2(G)$ is

$$\langle \varphi, (I - M)\varphi \rangle = \frac{1}{2} \sum_{g \in G} \sum_s \mu(s)(\varphi(gs) - \varphi(g))^2.$$

Suppose that F is a connected finite subset of G. Fix a function $\hat{\varphi} : \partial F \to \mathbb{R}$. Denote by Ω the set of functions in $\ell^2(G)$ that agree with $\hat{\varphi}$ on ∂F and vanish outside $F \cup \partial F$. Show that $\varphi \in \Omega$ has minimal energy among all elements of Ω iff $\varphi \in \ell_\mu(F)$.

Hint. Show that if $\varphi \in \Omega$ does not satisfy μ-harmonicity at some $g \in F$ then there is a $\varphi' \in \Omega$ that has lower energy. For the other direction, argue that the energy is continuous and strictly convex, then explain why this implies that there is a unique minimizer of the energy.

17. Let μ be a finitely supported, non-degenerate probability measure on $\mathbb{Z}^d$. We say that $\psi \colon \mathbb{Z}^d \to \mathbb{R}_+$ is *multiplicative* if $\psi(x+y) = \psi(x)\psi(y)$.

    (a) Prove that every multiplicative $\psi \colon \mathbb{Z}^d \to \mathbb{R}_+$ with $\psi(0) = 1$ is of the form $\psi(z) = \mathrm{e}^{t \cdot z}$ for some $t \in \mathbb{R}^d$. Show that such a ψ is furthermore μ-harmonic iff $\mathbb{E}\left[\mathrm{e}^{t \cdot X}\right] = 1$, where X has distribution μ.

    (b) Prove that every $\psi \in \partial_\mu(\mathbb{Z}^d)$ is multiplicative.

    Hint. First suppose that ψ is extremal. Then use the facts that if $\psi \in \ell_\mu(\mathbb{Z}^d, 1)$ then $\psi = \sum_s R_s \psi \mu(s)$ and $\sum_s \psi(s)\mu(s) = 1$. Then prove that $\frac{1}{\psi(s)}[R_s\psi] \in \ell_\mu(\mathbb{Z}^d, 1)$, and use the extremality of ψ. Finally, use this to extend the proof to all of $\partial_\mu G$.