ENTROPY, SPEED AND SPECTRAL RADIUS OF RANDOM WALKS

FERNANDO AL ASSAL

ABSTRACT. In this paper, we give an introduction to random walks on infinite discrete groups by focusing on three invariants – entropy, speed and spectral radius. Entropy is a measure of the randomness of the stochastic process, whereas the other two measure drift from the origin. We show a triple equivalence between positive entropy, speed and existence of nontrivial tail events. Moreover, we further motivate the spectral radius ρ by a theorem of Kesten which says $\rho < 1$ if and only if the group is amenable. We go through many examples and computations.

1. INTRODUCTION

By a *right random walk* on a group G it is meant a sequence of random variables

$$X_n = \xi_0 \xi_1 \cdots \xi_n$$

where the $(\xi_i)_{i\geq 1}$ are independent G-valued random variables with distribution μ.

The *transition probabilities* $p^n(x, y)$ are defined by

$$p^n(x, y) = \mathbb{P}_x(X_n = y) = \mu^{\star n}(x^{-1}y)$$

where $\mathbb{P}_x$ is the probability conditioned at the event that $\xi_0 = x$ and $\mu^{\star n}$ is the n-th convolution power of μ. (Recall that the convolution product of two measures is defined by $\mu \star \nu(E) = \int_G \int_G 1_E(yz) d\mu(y) d\nu(z)$ and that if $(\xi_i)_{i\geq 1}$ are i.i.d. with distribution μ, then $\xi_0\xi_1 \cdots \xi_n$ has distribution $\mu^{\star n}(\xi_0^{-1} \cdot)$.) The probability measure $\mathbb{P}_x$ is a measure on the probability space $\Omega = G^{\mathbb{N}}$ of infinite sequences of elements of G.

We shall assume the distribution is *symmetric*, i.e., $\mu(x) = \mu(x^{-1})$ and *irreducible*, i.e., given $x, y \in G$, there is m such that $p^m(x, y) > 0$. We also assume μ has *finite support*, i.e., the random walk has *finite range*.

In this paper we will mostly work with *finitely-generated* groups. Thus, we can think geometrically of our random walk on G as a random walk on a Cayley graph $\mathrm{Cay}(G, S)$, where S is a symmetric generating set such that $\operatorname{supp} \mu \subset S$. In this setting, our assumption of symmetry just means that the random walker is as likely to traverse the edge corresponding to multiplication by s as it is to traverse the s^{-1} edge. Irreducibility of the chain means the Cayley graph must be connected. Finite range means there is an universal m such that at every step the random walker cannot walk more than m units of distance in the graph metric d.

2. Entropy of a random variable

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X : \Omega \to F$ be a random variable, where $F = \{x_1, \ldots, x_m\}$. Let $p(X) = \mathbb{P}(X = x)$. We define the *entropy* of X by

$$\begin{aligned} H(X) &= -\mathbb{E} \log p(X) \\ &= -\sum_{i=1}^{m} p(x_i) \log p(x_i). \end{aligned}$$

Entropy, as defined above, is a concept in information theory introduced in 1948 by Shannon that measures the unpredictability (randomness) of the information content in a random variable. Suppose we see our F-valued X as the value of a roll of a $|F|$-sided biased die and suppose we keep throwing the die. (Equivalently, consider a sequence $(X(\omega_1), X(\omega_2), \ldots)$ where $(\omega_1, \omega_2, \ldots) \in \Omega^{\mathbb{N}}$.)

If X has high entropy, one obtains a lot of information at every die roll. That is, after a few rolls one is still uncertain about the value of next roll. Conversely, if X has low entropy, after observing a few outcomes one obtains little information afterwards.

For example, say X is a biased coin toss:

$$\mathbb{P}(X = 1) = p \quad \text{and} \quad \mathbb{P}(X = -1) = 1 - p.$$

Then, the entropy is

$$H(p) = p \log p + (1-p) \log(1-p).$$

A simple exercise in calculus shows $H(p)$ has a strict maximum at $p = 1/2$. This makes sense – if p is very close to 1 or 0, then a sequence of coin tosses will likely have mostly heads or mostly tails, being more predictable.

More generally, solving the constrained maximization problem

$$\begin{aligned} \max_{p(x_1), \ldots, p(x_m)} &\sum_{i=1}^{m} p(x_i) \log p(x_i) \\ \text{s.t.} &\sum_{i=1}^{m} p(x_i) = 1 \end{aligned}$$

gives us $p(x_i) = \frac{1}{n}$ for $1 \leq i \leq n$. In particular, *entropy is maximized by the uniform distribution*.

If X is uniformly distributed, a calculation shows $H(X) = \log |F|$. In particular,

$$H(X) \leq \log |F| \tag{2.1}$$

for every F-valued random variable X.

3. Three invariants of a random walk

Above we discussed a measure of randomness for a random variable X. A random walk X_n is, however, a *sequence* of random variables. Since the value of X_n depends on the value of X_{n-1}, the idea is that X_n gets more and more unpredictable as n gets large. Thus $H(X_n)$ should grow.

Hence, a reasonable measure of the randomness of the random process X_n is to see how much faster than linear is the growth of $H(X_n)$. In this spirit we define the *entropy of a random walk* (or *asymptotic entropy*) as

$$h = \limsup_{n\to\infty} \frac{H(X_n)}{n}.$$

Another property of a random walk one might want to discuss is how fast (if at all) the random walker drifts away from a certain reference point, say the identity 1 of G. To measure that, we define the *speed of a random walk* as

$$\ell = \limsup_{n\to\infty} \frac{\mathbb{E}|X_n|}{n},$$

where $|X_n| := d(X_n, 1)$, where 1 is the identity of the group.

Finally, another way to measure drift is seeing how fast the probability the random walker returns to where it started decays. Thus we define the *spectral radius* of the random walk as

$$\rho = \limsup_{n\to\infty} p^{2n}(1,1)^{1/2n}.$$

We have $0 < \rho \leq 1$, and $\rho < 1$ means that the probability the random walker returns to where it started decays exponentially.

The sequences $H(X_n)$, $\mathbb{E}|X_n|$ and $-\log p^{2n}(1,1)$ are all *subadditive*, i.e., $a_{n+m} < a_n + a_m$. A famous lemma due to Fekete says that for a subadditive sequence a_n,

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_n \frac{a_n}{n}.$$

Thus, we can substitute all of the lim sup's above by ordinary limits.

Moreover, the limits in the definitions of speed and entropy are almost sure limits – it is not necessary to take expectations. To see that, we recall the famous *Kingman subadditive ergodic theorem*:

Theorem 3.1 (Kingman)**.** *Let $f_n : G^n \to \mathbb{R}$ be bounded, measurable functions such that*

$$f_{m+n}(x_1, \dots, x_{m+n}) \leq f_m(x_1, \dots, x_m) + f_n(x_{m+1}, \dots, x_{m+n}).$$

Then, if $(\xi_n)_{n\geq 1}$ are i.i.d. G-valued random variables,

$$\lim_{n\to\infty} \frac{1}{n} f_n(\xi_1, \dots, \xi_n) = \inf_n \frac{1}{n} \mathbb{E} f_n(\xi_1, \dots, \xi_n).$$

Applying the theorem to sequences

$$f_n(\xi_1, \dots, \xi_n) = d(\xi_1 \cdots \xi_n, 1) \quad \text{and} \quad g_n(\xi_1, \dots, \xi_n) = H(\xi_1 \cdots \xi_n)$$

we conclude that *almost surely*

$$h = -\lim_{n\to\infty} \frac{1}{n} \log p_n(X_n) \quad \text{and} \quad \ell = \lim_{n\to\infty} \frac{|X_n|}{n}.$$

Example 3.2. Consider a simple random walk on X_n on the grid $\mathbb{Z}^d$. Jensen's inequality gives us that

$$\begin{aligned}(\mathbb{E}|X_n|)^2 &\leq \mathbb{E}|X_n|^2 \\ &= \mathbb{E}(Y_1 + \cdots + Y_n)^2 = n,\end{aligned}$$

so $\mathbb{E}|X_n| \leq \sqrt{n}$, whence the speed of the random walk is $\ell_{\mathbb{Z}^d} = 0$.

The group $\mathbb{Z}^d$ has *polynomial growth* – the size of a ball of radius n is $|B_n| = Cn^d$ for some constant C. Therefore (2.1) gives that

$$H(X_n) \leq \log C + d \log n,$$

so the entropy of X_n is $h_{\mathbb{Z}^d} = 0$. In particular, this shows that *finite-range random walks on groups of polynomial growth have zero entropy*.

There is a *local limit theorem* that says that for simple random walk on $\mathbb{Z}^d$,

$$p^{2n}(0,0) \sim Cn^{-d/2}.$$

In particular, $\rho_{\mathbb{Z}^d} = 1$.

Example 3.3. Let $\mathbb{F}_k$ be the free group on k generators, with $k \geq 2$. Consider a simple random walk X_n on $\mathbb{F}_k$, which is the same thing as a random walk on the Cayley graph of $\mathbb{F}_k$–the regular $2k$-valent tree.

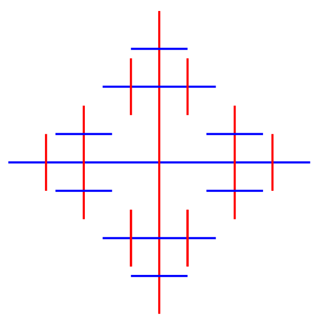

Fig. 1: Cayley graph of the free group on two generators $\mathbb{F}_2 = \langle a, b \rangle$ with respect to the symmetric generating set $S = \{a, a^{-1}, b, b^{-1}\}$.

Notice that as soon as the random walker picks a direction, the probability it backtracks is $p = (2k-1)/2k$. For the purposes of studying asymptotic behavior, we can therefore think of a random walk on $\mathbb{F}_k$ as a biased random walk on $\mathbb{N}$ with $p(n, n+1) = (2k-1)/2k$ and $p(n, n-1) = 1/2k$. Thus, the return probabilities decay exponentially – $\rho_{\mathbb{F}_k} < 1$.

Moreover, the speed of the simple random walk on $\mathbb{F}_k$ is almost surely equal to

$$\lim_{n\to\infty} \frac{Y_n}{n},$$

where $Y_n = \xi_1 + \cdots + \xi_n$ is the said biased random walk on $\mathbb{N}$. By the strong law of large numbers, this limit goes to $\mathbb{E}\xi_i$, whence the speed of the simple random walk on $\mathbb{F}_k$ is a.s.

$$\ell_{\mathbb{F}_k} = \frac{k-1}{k}.$$

Example 3.4. The *lamplighter group* $\mathcal{G}$ is defined as the semidirect product

$$\mathcal{G} = \bigoplus_{x \in G} \mathbb{Z}_2 \rtimes \mathbb{Z}.$$

The elements of $\mathcal{G}$ are pairs (η, x), where $\eta : \mathbb{Z} \to \mathbb{Z}_2$ is a finitely supported function and $x \in \mathbb{Z}$. The product operation is defined as

$$(\eta, x)(\xi, y) := (\eta \oplus (x + \xi), x + y)$$

where $\oplus$ is componentwise sum mod 2 and $(x + \xi)(g) := \xi(g - x)$.

We will consider random walks with distribution supported by the symmetric generating set

$$S = \{(0, 1), (0, -1), (\chi_0, 0)\},$$

where χ_0 is the characteristic function of $\{0\}$.

The way one really should think about this random walk is to imagine that at each site $x \in \mathbb{Z}$ there is a lamplight, which could be on or off. Given a finitely supported $\eta : \mathbb{Z} \to \mathbb{Z}_2$, one see $\eta^{-1}(1)$ as the finite set of lights which are turned on. If the random walker moves along the generators $(0, \pm 1)$ above, it just moves along the integer lattice without switching any lights. Moving along $(\chi_0, 0)$ corresponds to switching a light.

The random walker has a lot of ways to move away from the origin in the Cayley graph $(\mathcal{G}, S)$ as it switches lights on. Thus, one might suspect that the $\mathcal{G}$ random walk has higher speed and entropy than simple random walk on $\mathbb{Z}$. However, this is not the case.

Proposition 3.5 (Kaimanovich-Vershik)**.** *The entropy of a nearest neighbor random walk on the lamplighter group is $h_{\mathcal{G}} = 0$.*

Proof. Let $Z_n = (\eta_n, X_n)$ be the random walk. Observe that the projection on $\mathbb{Z}$,

$$X_n = Y_1 + \cdots + Y_n$$

is a sum of independent random variables. Kolmogorov's inequality gives a lower bound on the probability that X_k stays on the ball $B_{n^{3/4}}$ in $\mathbb{Z}^d$ for all $k \le n$:

$$\mathbb{P}_{(0,0)}(Z_n \colon \forall k \le n, |X_k| \le n^{3/4}) \ge 1 - \frac{\operatorname{Var} X_n}{n^{3/2}}.$$

Moreover observe that since there is $\delta > 0$ such that

$$\min\{\mu(x) \colon x \in \operatorname{supp} \mu\} = \delta,$$

then for all $x \in \operatorname{supp} \mu^{\star n}$ we have

$$\mathbb{P}(X_n = x) \ge \delta^n$$

Now consider the set

$$A_n = \{(\eta, x) \in \mathcal{G}_1 : |x| \le n^{3/4} \text{ and } \operatorname{supp} \eta \subset [-n^{3/4}, n^{3/4}]\}.$$

Observe that its cardinality does not grow exponentially:

$$\log_2 |A_n| = \frac{3}{4}\log_2 n + 2n^{3/4}.$$

Using all these inequalities, we can bound the entropy of X_n

$$\begin{aligned} H(X_n) &= \sum_{x\in A_n} p(X_n)(-\log)p(X_n) + \sum_{x\notin A_n} p(X_n)(-\log)(p(X_n)) \\ &\leq \log|A_n| - \log\delta^n \mathbb{P}_{(0,0)}(X_n \notin A_n) \\ &\leq \log|A_n| - n\log\delta\frac{\operatorname{Var} Y_i}{n^{1/2}} \\ &= o(n). \end{aligned}$$

□

Remark 3.6. One can study more general lamplighter groups, where we have a lamplight at each vertex of an arbitrary Cayley graph instead of the integer line. Given a discrete group G, we define the lamplighter group of G as

$$\operatorname{Lamp} G = \bigoplus_{x\in G} \mathbb{Z}_2 \rtimes G.$$

It is an interesting fact the entropy (and thus the speed) of a simple random walk on $\operatorname{Lamp}\mathbb{Z}^d$ is zero if and only if $d < 3$.

Example 3.7. The speed of a random walk also makes sense for Lie groups. Let $X_n = M_n \cdots M_1$ be a *left* random walk on $\mathrm{SL}(2,\mathbb{R})$, where the distribution μ induced by the M_i is such that

(a) $\int_{\mathrm{SL}(2,\mathbb{R})} \log\|M\| d\mu(M) < \infty$.

(b) The subgroup G_μ generated by $\operatorname{supp}\mu$ is not compact.

(c) There is no finite set $L \subseteq \mathbb{RP}^1$ left invariant by μ-a.e. element of $\mathrm{SL}(2,\mathbb{R})$.

Then, a theorem due to Furstenberg says that the *Lyapunov exponent*

$$\lambda = \lim_{n\to\infty}\frac{1}{n}\log\|M_1\cdots M_n\|$$

is almost surely positive. Under certain conditions, the word metric on the group is analogous to the log of the matrix norm above. Hence we see that this says that the random walk with distribution μ in $\mathrm{SL}(2,\mathbb{R})$ has positive speed.

A basic relation between two of these invariants is

Proposition 3.8. *$h \geq -2\log\rho$.*

In particular, *spectral radius less than one implies positive entropy.* That is, exponential decay of the return probabilities should entail a large amount of randomness.

Proof. By symmetry,

$$p^{2n}(1,1) = \sum_{x \in G} p_n(x)^2.$$

Now we estimate the $p_n(x)$ in the above sum. We know, almost surely, that

$$h = -\lim_{n\to\infty} \frac{1}{n} \log p_n(X_n).$$

Given $\epsilon > 0$, we know by *Egorov's theorem* that there is an event $E \subset \Omega$ with $\mathbb{P}(E) \geq 1-\epsilon$ such that the above limit is uniform on E.

In particular, we may find $N(\epsilon)$ such that $n \geq N(\epsilon)$ implies

$$\left| -\frac{1}{n} \log p_n(X_n(\omega)) - h \right| < \epsilon$$

for all $\omega \in E$. Thus,

$$\mathbb{P}(p_n(X_n) \geq e^{-n(h+\epsilon)}) \geq 1-\epsilon.$$

By Chebyshev's inequality,

$$\mathbb{E}p_n(X_n) \geq e^{-n(h+\epsilon)}(1-\epsilon).$$

But

$$\mathbb{E}p_n(X_n) = \sum_{x\in G} p_n(x)^2.$$

So taking the above inequality to the $1/2n$-th power and letting $n \to \infty$ we get

$$\rho \geq e^{-\frac{h+\epsilon}{2}}.$$

Since $\epsilon > 0$ was arbitrary, this gives the desired relation. □

4. Positive entropy is equivalent to positive speed

Here we prove another, more involved relation between these invariants.

Theorem 4.1. *Let X_n be a symmetric finite range random walk on a finitely generated group G. Then, the random walk has positive speed if and only if it has positive entropy.*

The fact that positive entropy implies positive speed makes some sense a priori. In fact, if your random walk has a large amount of randomness, then one ought to expect it to be able to walk through a large amount of sites. Since the random walk has *finite range*, it cannot jump too far at each step, so it should be walking away from the origin.

The other direction might seem less intuitive. In fact, if your random walker is always walking in a fixed direction with high probability, then the random walk has positive speed but low entropy, since there is barely any randomness. But what makes the theorem work is the assumption of *symmetry* – it prevents the random walker from just walking very biasedly in a single direction.

Proof. **Positive entropy $\implies$ positive speed.**

The inequality (2.1) gives us

$$H(X_n) \le \log |B_n|.$$

Since $h > 0$, the entropy of X_n grows at least linearly, i.e., there is some positive $\alpha > 0$ such that

$$H(X_n) \ge \alpha n$$

for $n \gg 1$. In particular these inequalities imply

$$|B_n| \ge e^{\alpha n}.$$

That is, G has *exponential growth* with rate at least $\alpha > 0$.

Let ℓ be the speed of the random walk. We know, almost surely, that

$$\ell = \lim_{n\to\infty} \frac{|X_n|}{n}.$$

Let $\epsilon > 0$ be given. By Egorov's theorem, we may find $N(\epsilon)$ such that $n \ge N(\epsilon)$ implies

$$|X_n(\omega)| \le n(\ell + \epsilon)$$

uniformly for $\omega \in E$, where E is an event with $\mathbb{P}(E) \ge 1 - \epsilon$. This means

$$\mathbb{P}(X_n \in B_{n(\ell+\epsilon)}) \ge 1 - \epsilon.$$

Now we have a handle on the entropy of X_n. We write

$$H(X_n) = \sum_{x \in B_{n(\ell+\epsilon)}} p_n(x)(-\log)(p_n(x)) + \sum_{x \notin B_{n(\ell+\epsilon)}} p_n(x)(-\log)(p_n(x)).$$

We can bound the first term using the fundamental inequality (2.1) and exponential growth. We have

$$\begin{aligned}\sum_{x \in B_{n(\ell+\epsilon)}} p_n(x)(-\log)(p_n(x)) &\le \log |B_{n(\ell+\epsilon)}| \\ &\le \alpha(\ell + \epsilon)n.\end{aligned}$$

It was observed in the proof of Proposition 3.5 that there is $\delta \in (0, 1]$ such that

$$\mathbb{P}_1(X_n = x) \ge \delta^n$$

for every $x \in \operatorname{supp} \mu^{\star n}$. This gives us a way to control the second term, as

$$(-\log)p_n(x) \le n\delta,$$

where $\delta \in (0, 1]$ is fixed and $x \in \operatorname{supp} \mu^{\star n}$. Therefore,

$$\begin{aligned}\sum_{x \notin B_{n(\ell+\epsilon)}} p_n(x)(-\log)(p_n(x)) &\le n\delta \mathbb{P}(X_n \notin B_{n(\ell+\epsilon)}) \\ &\le n\delta\epsilon.\end{aligned}$$

The two inequalities give

$$H(X_n) \leq \alpha(\ell + \epsilon)n + n\delta\epsilon,$$

whence

$$h \leq \alpha(\ell + \epsilon) + \delta\epsilon.$$

But $\epsilon > 0$ was arbitrary and δ only depends on the distribution μ. Thus we have shown

$$h \leq \alpha \cdot \ell,$$

which concludes the proof.

2. Positive speed $\implies$ positive entropy.

The proof of this hinges on the following estimate:

Theorem 4.2 (Carne-Varopoulos)**.** *Let X_n be a symmetric finite-range random walk on a finitely generated group G with step distribution supported by a symmetric generating set A. Let d be the graph metric on the Cayley graph $C(G, A)$. Then, for every $x \in G$,*

$$\mathbb{P}_1(X_n = x) \leq \rho^n \exp\left(-\frac{|x|^2}{2n}\right)$$

where ρ is the spectral radius of the random walk.

In particular, this implies

$$\log p_n(X_n) \leq -\frac{|x|^2}{2n},$$

for every $x \in \operatorname{supp} \mu^{\star n}$. Thus,

$$-\frac{1}{n} \log p_n(X_n) \geq \frac{1}{2}\left(\frac{|X_n|}{n}\right)^2.$$

Taking the pointwise limit,

$$h \geq \frac{1}{2}\ell^2.$$

Thus positive speed implies positive entropy. Notice that this all works even if $\rho = 1$, even though the estimate is sharper when $\rho < 1$. □

5. Entropy and tail events

There is an alternate characterization of the property of a random walk having positive speed and entropy, which has to do with *tail events*. An event A is said to be a tail event if its occurrence does not depend on the first finitely many values of the random walk.

More precisely, we can define an equivalence relation $\sim$ on the probability space $\Omega = G^{\mathbb{N}}$ by saying $\omega_1 \sim \omega_2$ if the ω_1 and ω_2 differ (as sequences) by finitely many values. Then, we say A is a tail event if $\omega \in A$ implies $\omega' \in A$ for all $\omega' \sim \omega$. The set of tail events forms a σ-algebra, the *tail σ-algebra $\mathscr{T}$*.

As an exercise, the reader can check that

$$\mathscr{T} = \bigcap_{n \geq 0} \sigma(X_n, X_{n+1}, \ldots),$$

where $\sigma(X_n, X_{n+1}, \ldots)$ is the smallest σ-algebra such that $X_n, X_{n+1}, \ldots$ are measurable.

Example 5.1. Let X_n be a simple random walk on $\mathbb{F}_2$. The event that X_n ends in the top branch of the tree (Figure 1) is a tail event. It has probability 1/4.

Example 5.2. Let X_n be a simple random walk on $\mathbb{Z}$. The event that X_n visits the origin infinitely many times is a tail event. It has probability zero.

Example 5.3. Let X_n be a simple random walk on Lamp $\mathbb{Z}^3$. The event that the lamp in the origin remains lit forever is a tail event. This event has probability p such that $0 < p < 1$, which we will not prove here.

This suggests that a random walk has nontrivial tail events (i.e., with probability strictly between 0 and 1) if and only if it has positive speed or entropy. This makes sense because positive entropy means there is a large amount of randomness in the process. Thus, there should be many different ways for it to drift away from the origin, i.e., nontrivial tail events.

Theorem 5.4. *The tail σ-algebra $\mathscr{T}$ has nontrivial events if and only if the entropy is positive.*

The proof of this theorem requires the concept of *conditional entropy*, which measures the amount of randomness of a random variable X if we are given the value of Y. We define $H(X|Y)$, the conditional entropy of X given Y, as

$$H(X|Y) = \sum_{y \in G} \mathbb{P}(Y = y) H(X|Y = y),$$

where

$$H(X|Y = y) = \sum_{x \in G} \mathbb{P}(X = x|Y = y)(-\log)\mathbb{P}(X = x|Y = y).$$

It is a fact, that we shall not prove here, that entropy does not increase as you condition it on more information. That is,

$$H(X|Y) \leq H(X)$$

with equality being achieved if and only if X and Y are independent. Moreover,

$$H(X|Y, Z) \leq H(X|Y),$$

where $H(X|Y, Z) = H(X|(Y, Z))$ is the entropy of X conditioned on the joint distribution (Y, Z).

Proof of Theorem 5.4. It is a straightfoward calculation to show that X and Y are random variables, then

$$H(X) + H(Y|X) = H(Y) + H(X|Y).$$

In particular, if X_n is the random walk, then

$$H(X_1) + H(X_n|X_1) = H(X_n) + H(X_1|X_n).$$

It is easy to see that $H(X_n|X_1) = H(X_{n-1})$. Moreover, by the Markov property,

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

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

Entropy of a random variable does not increase if you condition it on more information. Therefore, $H(X_1|(X_n, X_{n+1}, \ldots))$ does not decrease as n as you increase n. This shows $\delta_n = H(X_n) - H(X_{n-1})$ is nonincreasing. Since entropy is nondecreasing as n increases, i.e., $\delta_n \geq 0$, we have that $\delta_n \to \delta$ for some $\delta \in [0, \infty)$.

As we take $n \to \infty$, we get
$$H(X_1) - \delta = H(X_1|\mathscr{T}).$$
If the asymptotic entropy h is positive, then $H(X_n)$ grows at least linearly. That is, there is some $\alpha > 0$ such that
$$H(X_{n+1}) - H(X_n) \geq \alpha$$
for every $n \gg 1$. In particular, $\delta > 0$, so $H(X_1|\mathscr{T}) < H(X_1)$. This shows the tail σ-algebra $\mathscr{T}$ cannot be trivial.

Conversely, if the asymptotic entropy h is zero, then $\delta = 0$, so $H(X_1) = H(X_1|\mathscr{T})$. Thus X_1 is independent from every tail event. But given $T \in \mathscr{T}$, we know $T \in \sigma(X_1, X_2, \ldots)$, so in particular T is X_1-measurable. In particular $T = \{X_1 \in B\}$ for some Borel set B. Since X_1 is independent from T, T is independent from itself, so
$$\mathbb{P}(T \cap T) = \mathbb{P}(T)^2,$$
which shows T is trivial. □

6. Amenability and spectral radius

The triple equivalence proven in the last two sections gives us a good feel for what speed and entropy actually mean. But what about the spectral radius ρ? On the one hand, exponential decay of return probabilities ($\rho < 1$) tells us that the random walk is somehow walking away from the origin. In fact, we saw that positive speed implies $\rho < 1$. However, the converse is not true – the lamplighter group $\mathcal{G}$ has spectral radius 1 and zero speed.

So what does the condition $\rho < 1$ mean? It is connected to the concept of *amenability* of groups.

We define the *isoperimetric constant* of a group G to be
$$\kappa = \inf_F \frac{|\partial F|}{|F|},$$
where F is a finite set in a Cayley graph $\mathrm{Cay}(G, S)$, where S is a symmetric generating set (the choice of S is unimportant). This is measures the ratio between surface and volume in the group. A group is said to be *nonamenable* if it has positive isoperimetric constant.

Example 6.1. The lattice $\mathbb{Z}^d$ and the lamplighter group Lamp $\mathbb{Z}^d$ are amenable. However, the tree $\mathbb{F}_k$ is not. More generally, *hyperbolic groups* – those in which the distance between two sides of a triangle can always be fit in a δ-ball for some constant δ are nonamenable.

The following theorem, which initiated the subject of random walks on groups in 1959, shows that the spectral radius ρ actually tells us about the geometry of the group.

Theorem 6.2 (Kesten)**.** *Let X_n be a symmetric irreducible finite range random walk on G. Then, $\rho < 1$ if and only if G is nonamenable.*

This in particular tells us that random walks on nonamenable groups are transient. The reader is referred to [6] for a proof.

Appendix A. The Carne-Varopoulos bound

The proof of the Carne-Varopoulos bound is an interesting excursion into functional analysis. We use the following version of the spectral theorem:

Definition A.1. The *spectrum* of a linear operator T, denoted $\operatorname{Spec} T$ is defined by

$$\operatorname{Spec} T = \{\lambda \in \mathbb{R} \colon\ T - \lambda \,\mathrm{Id} \ \text{ is not bijective}\}.$$

Theorem A.2. *Let H be a Hilbert space and $T : H \to H$ be a bounded self-adjoint linear operator. Fix $u \in H$. Then, there is a unique positive Radon measure $\nu = \nu_{T,u}$, the so-called* spectral measure *with respect to T and u, such that*

$$(f(T)u, u) = \int_{\operatorname{Spec} T} f \, d\nu.$$

Proof. See [4]. □

Proof of Carne-Varopoulos. Consider the Markov operator P defined by

$$\begin{aligned} Pu(x) &= \sum_{y \in G} p(x, y) u(y) \\ &= \mathbb{E}_x u(X_1). \end{aligned}$$

Observe P is a bounded linear operator. We claim moreover it maps the Hilbert space $\ell^2(G)$ into itself, where

$$\ell^2(G) = \left\{ u : G \to \mathbb{R} \colon\ \sum_{x \in G} u(x)^2 < \infty \right\}.$$

To see this, if we take $u \in \ell^2(G)$, observe that

$$\begin{aligned} \|Pu\|_{\ell^2(G)} &= \sum_{x \in G} (\mathbb{E}_x u(X_1))^2 \\ &\leq \sum_{x \in G} \mathbb{E}_x \left(u(X_1)^2 \right) \\ &= \sum_{x \in G} \sum_{y \in G} u(y)^2 p(x, y) \\ &\leq \sum_{x \in G} \|u^2\|_{\ell^2(G)} \left(\sum_{y \in G} p(x, y)^2 \right)^{1/2}, \end{aligned}$$

where the second step is Jensen's inequality and the final step is Cauchy-Schwartz for the $\ell^2(G)$ inner product. Notice that since $\|u\|_{\ell^2 G} < \infty$, we have $\|u^2\|_{\ell^2(G)} < \infty$ as well.

Now we exploit the fact that the step distribution of the random walk is supported by a finite generating set $A = \{a_1, \dots, a_n\}$. Our estimate then becomes

$$\begin{aligned}
\|Pu\|_{\ell^2(G)} &\le C \sum_{x \in G} \left(\sum_{i=1}^{n} p(x, xa_i)^2 \right)^{1/2} \\
&\le \sum_{x \in G} \sum_{i=1}^{n} p(x, xa_i) \\
&= \sum_{i=1}^{n} \sum_{x \in G} p(xa_i, x) = n < \infty,
\end{aligned}$$

where we used the symmetry of the random walk on the last step.

Since the random walk is symmetric, it follows that the Markov operator P is self adjoint:

$$\begin{aligned}
(Pu, v) &= \sum_{x \in G} Pu(x)v(x) \\
&= \sum_{x \in G} \sum_{y \in G} p(x, y)u(y)v(x) \\
&= \sum_{y \in G} \sum_{x \in G} p(y, x)v(x)u(y) = (u, Pv).
\end{aligned}$$

(The interchange of summation signs above follows from Tonelli's theorem – since $(\cdot, \cdot)$ is bilinear and we can write $u = u^+ - u^-$ where u^+ and u^- are nonnegative, we may assume without loss of generality that u is nonnegative.)

A classical result due to Chebyshev says that for every $k \in \mathbb{Z}$ there is a unique polynomial $Q_k \in \mathbb{R}[x]$, of degree k, satisfying $Q_k(\cos \xi) = \cos k\xi$ for all $\xi \in \mathbb{R}$. In particular, this implies $|Q_k(s)| \le 1$ for $|s| \le 1$.

Let $S_n = Y_1 + \cdots + Y_n$ be a simple random walk on $\mathbb{Z}$. In the proof of the LLT for $\mathbb{Z}^d$ above, we computed the generating function of S_n,

$$\phi_{S_n}(\theta) = \mathbb{E}e^{iS_n\theta} = \cos^n \theta,$$

whence we get the formula

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

In particular, we can write

$$s^n = \sum_{k \in \mathbb{Z}} \mathbb{P}(S_n = k) Q_k(s).$$

Since our operator P is bounded and self-adjoint, a theorem from functional analysis says there is a Banach algebra isometric homomorphism

$$\psi_P : C(\operatorname{Spec} T) \to \mathscr{L}(\ell^2(G))$$

such that $\psi_P(p) = \sum_j a_j P^j$, where $p \in \mathbb{R}[x]$ is a polynomial $p(x) = \sum_j a_j x^j$. In particular,

$$P^n = \|P^n\| \sum_{k \in \mathbb{Z}} \mathbb{P}(S_n = k) Q_k(P/\|P\|),$$

so by continuity and bilinearity of the inner product, we have

$$(\delta_x, P^n \delta_y) = \|P^n\| \sum_{k \in \mathbb{Z}} \mathbb{P}(S_n = k) \left(\delta_x, Q_k(P/\|P\|)\delta_y\right).$$

But observe that

$$\begin{aligned} (\delta_x, P^n \delta_y) &= P^n \delta_y(x) \\ &= \sum_{y \in G} p^{(n)}(x, y) \delta_x(y) = p^{(n)}(x, y), \end{aligned}$$

whence we get the formula for the transition probability

$$p^{(n)}(x, y) = \|P^n\| \sum_{k \in \mathbb{Z}} \mathbb{P}(S_n = k) \left(\delta_x, Q_k(P/\|P\|)\delta_y\right).$$

On the one hand, since P is self-adjoint, we know $\|P^n\| = \|P\|^n$. Moreover, we can show that the operator norm of the Markov operator P is the spectral radius ρ of the random walk. By the spectral theorem, we know there is spectral measure ν such that

$$(u, P^n u) = \int_{\operatorname{Spec} P} u^n d\nu(u).$$

In particular, having the calculations in the above paragraph in mind,

$$p^{(2n)}(0, 0)^{1/2n} = \| \operatorname{Id} \|_{L^{2n}(\operatorname{Spec} P, \nu)}$$

where Id is the identity function on $\mathbb{R}$. So as $n \to \infty$ we obtain

$$\rho = \| \operatorname{Id} \|_{L^\infty(\operatorname{Spec} P)}.$$

But it is a fact in functional analysis (see [1]) that for P bounded and self-adjoint on a Hilbert space,

$$\|P\| = \| \operatorname{Id} \|_{L^\infty(\operatorname{Spec} P)}.$$

Therefore our estimate for the transition probability becomes

$$\begin{aligned} p^{(n)}(x, y) &= \rho^n \sum_{k \in \mathbb{Z}} \mathbb{P}(S_n = k) \left(\delta_x, Q_k(P/\rho)\delta_y\right) \\ &\leq \rho^n \sum_{k \in \mathbb{Z}} \mathbb{P}(S_n = k) \\ &\leq 2\rho^n \sum_{k \geq d(x,y)} \mathbb{P}(S_n = k), \end{aligned}$$

where the first step is by Cauchy-Schwartz.

We can then conclude using Hoeffding's inequality for simple random walk in $\mathbb{Z}$:

$$P(|S_n| \geq \alpha) \leq 2e^{-\frac{2\alpha^2}{n}}.$$

$\square$

Acknowledgments. It is my pleasure to thank my mentor Clark Butler for guiding me through the subject of random walks on groups with deep insights and helpful suggestions. My sincere thanks also go to people behind the organization of the 2015 REU and the Second Summer School in Analysis.

References

[1] Brezis, H. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.*

[2] Kaimanovich, V. A.; Vershik, A. M. Random Walks on Discrete Groups: Boundary and Entropy. Ann. Probab. 11 (1983), no. 3, 457–490. doi:10.1214/aop/1176993497.
`http://projecteuclid.org/euclid.aop/1176993497`.

[3] Komjáthy, J.; Peres, Y. Lecture Notes for Markov Chains: Mixing Times, Hitting Times and Cover Times. Saint Perersburg Summer School, 2012.
`http://www.win.tue.nl/~jkomjath/SPBlecturenotes.pdf`.

[4] Kowalski, E. Spectral theory in Hilbert spaces (ETH Zürich, FS 09).
`https://people.math.ethz.ch/~kowalski/spectral-theory.pdf`

[5] Lalley, S. Measure-Theoretic Probability I.
`http://galton.uchicago.edu/~lalley/Courses/381/MeasureTheory.pdf`

[6] Lalley, S. Random Walks on Hyperbolic Groups I.
`http://galton.uchicago.edu/~lalley/Talks/paris-talkA.pdf`.