Random Walks Revision Notes

Ted Jones-Healey

April 14, 2019

Abstract

These notes are based off the lectures of Dr Tuomas Sahlsten for the course on Analysis, Random Walks and Groups given at the University of Manchester in 2019 with minor additions and subtractions in content.

Contents

0 Introduction

We model a pack of cards in its ordered state (i.e. King Clubs, Queen clubs,..., Ace Clubs, ...Diamonds, ...Hearts,...Spades) as the set $\{0, \dots, 51\}$, then a **shuffle** is $\sigma \in S_{52}$ i.e. a bijection of the set $\{0, \dots, 51\}$.

Example 0.1. Perfect riffle shuffle $\sigma \in S_{52}$ s.t.

$$\sigma(j) := 2j \mod 51, \text{ i.e. } \sigma(j) := \begin{cases} 2j & \text{if } 0 \leq j \leq 25 \\ 2j - 51 & \text{if } 26 \leq j \leq 51 \end{cases}$$

Note: the perfect riffle is deterministic, i.e. as a distribution $\mu : S_{52} \to [0, 1]$, $\mu = \delta_\sigma$ (i.e. dirac distribution on σ), so performs the same shuffle every time. It has order 8:

Solution: We seek minimal k s.t. $\sigma^k = \text{id}$. Since $\sigma^k(j) = 2^k j \mod 51$ we require k s.t. $2^k = 1 \mod 51$. Checking $k = 1, 2, \dots, 8$ shows 8 is the correct value.

First example of a probability distribution over the group S_{52}:

Example 0.2. Random transposition is given by randomly picking two cards, with replacement, and swapping them. If same card is chosen twice do nothing. This a particular probability distribution $\mu : S_{52} \to [0,1]$, where $\mu(\sigma) \neq 0$ iff $\sigma = \text{id}$, or σ swaps exactly two cards. Prob of swapping distinct $i, j \in \{0, \dots, 51\}$ (i.e. transposition $(i\ j) \in S_{52}$) is $\mu((i\ j)) = \frac{2}{52^2}$, while prob of doing nothing $\mu(\text{id}) = \sum^{52} \frac{1}{52^2} = \frac{1}{52}$.

Example 0.3. Pass the broccoli Given p people arranged in a circle, modelled by the group $\mathbb{Z}_p = \{0, \dots, p-1\}$, if the broccoli starts at 0, and at each turn the broccoli is passed either left or right with probability $\frac{1}{2}$, we get a random walk on $\mathbb{Z}_p$.

Definition 0.4. Note in the following all measure-theoretic matters are neglected as this course focuses on the simpler case of finite discrete sets.

- **Sample space** is just a set Ω (for instance $\Omega = \mathbb{Z}_p$), and **event** is element of the power set of sample space, $P(\Omega)$.
- **Probability distribution** is a map $\mu : \Omega \to [0,1]$ such that $\sum_{\omega \in \Omega} \mu(\omega) = 1$.
- **Probability measure** is the extension of μ to events: $\mu : P(\Omega) \to [0,1]$, such that for $A \subset \Omega$: $\mu(A) := \sum_{\omega \in A} \mu(\omega)$, with $\mu(\emptyset) := 0$. By defn of a probability distribution we have: $\mu(\Omega) = 1$.

  For a probability distribution on $\mathbb{Z}_p$, the probability measure extending it satisfies
  - $A \subset B \subset \mathbb{Z}_p \implies \mu(A) \leq \mu(B)$ "monotonicity"
  - $\{A_i\}$ countable pairwise disjoint subsets, then $\mu(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$ "σ-additive"
  - $\mu(\mathbb{Z}_p) = 1$
- **Probability space** is a triple $(\Omega, P(\Omega), \mu)$ - Sample space, event space, and probability measure.
- For set S, an $S-$**valued random variable** is map $X : \Omega \to S$.
- **Probability distribution wrt random variable** $X : \Omega \to S$ **and probability space** $(\Omega, P(\Omega), \mu)$ is the function

$$\widetilde{\mu} : X(\Omega) \to [0,1] \text{ s.t. } \widetilde{\mu}(b) := \mu(X^{-1}(\{b\}))$$

  This allows us to associate probabilities to "areas" in the image space of the random variable.

Note: Normal terminology for probability distribution is what we call a "probability distribution wrt a random variable and probability space". This way, when the sample space Ω is difficult to express explicitly, we can instead define distributions over a domain which is actually the image set of the random variable. For instance a "normally-distributed random variable" X has a probability distribution defined over $\mathbb{R}$ given by the bell-curve. In the simple case $\Omega = \mathbb{Z}_p$, we can define probability distributions wrt an implicit random variable $\text{id} : \mathbb{Z}_p \to \mathbb{Z}_p$, so it is defined directly over Ω and this case coincides with the definition of distribution given in this course.

---

1 Probability on $\mathbb{Z}_p$

Course notes denote addition in $\mathbb{Z}_p$ as $t \oplus s := t + s \mod p$, and subtraction $t \ominus s := t - s \mod p$. Here we just use $+, -$.

Definition 1.1. **Probability distribution on** $\mathbb{Z}_p$ is a map $\mu : \mathbb{Z}_p \to [0,1]$ s.t. $\sum_{t=0}^{p-1} \mu(t) = 1$.

Example 1.2.

- **Uniform distribution** is $\lambda : \mathbb{Z}_p \to [0,1]$ s.t. $\lambda(t) = 1/p \; \forall t \in \mathbb{Z}_p$.
- **Dirac distribution** at $s \in \mathbb{Z}_p$ is $\delta_s : \mathbb{Z}_p \to [0,1]$ s.t. $\delta_s(t) := 1$ if $t = s$, 0 otherwise.
- **Fair Pass the broccoli distribution**

$$\mu(t) = \begin{cases} \frac{1}{2} & t = 1 \\ \frac{1}{2} & t = -1 \\ 0 & \text{otherwise} \end{cases}$$

Solution: They are indeed distributions since they each satisfy $\sum_{t=0}^{p-1} \lambda(t) = 1$.

Definition 1.3. The **expectation** of random variable $f : \mathbb{Z}_p \to \mathbb{C}$ wrt distribution μ is

$$\mathbb{E}_\mu(f) = \mu(f) := \sum_{t \in \mathbb{Z}_p} f(t)\mu(t) \in \mathbb{C}$$

Note: The indicator function for $A \subset \mathbb{Z}_p$, $\mathbf{1}_A(t) := \begin{cases} 1 & t \in A \\ 0 & t \notin A \end{cases}$, has expectation equal to the probability of the event A.

Theorem 1.4. *For distributions $\mu_1, \ldots, \mu_n$ and $\alpha_1, \ldots, \alpha_n \in [0,1]$ s.t. $\sum_i \alpha_i = 1$, then the **convex combination** $\mu(t) := \sum_i \alpha_i \mu_i(t)$ is also a probability distribution.*

Example 1.5. **Biased pass the broccoli** is the convex combination of delta distributions: for $\alpha \in (0,1)$

$$\mu_\alpha = \alpha\delta_1 + (1-\alpha)\delta_{-1}$$

Note to each probability distribution there is a certain degree of **uncertainty** as to what value it will take. For the uniform distribution this uncertainty is maximal since all options in the sample space are equally likely, whilst for the delta distribution the uncertainty is minimal, i.e. we are certain about what will occur. We cover two approaches to measuring the uncertainty contained in a distribution: **total variation distance** (define metric on space of probability distributions and then measure distance from the uniform distribution), or via **entropy**.

Definition 1.6.

- **Total variation distance** between μ, ν is $d(\mu, \nu) := \max_{A \subset \mathbb{Z}_p} |\mu(A) - \nu(A)|$
- L_1**-norm** on the set of functions $\{f : \mathbb{Z}_p \to \mathbb{R}\}$ is $\| f \|_1 := \sum_{t \in \mathbb{Z}_p} |f(t)|$
- L^∞**-norm** on $f : \mathbb{Z}_p \to \mathbb{R}$ is $\| f \|_\infty = \max_{t \in \mathbb{Z}_p} |f(t)|$

Lemma 1.7. *Total variation distance is a metric, i.e. satisfies triangle inequality: $d(\mu, \nu) \leq d(\mu, \tau) + d(\tau, \nu)$, symmetric: $d(\mu, \nu) = d(\nu, \mu)$, equality: $d(\mu, \nu) = 0$ iff $\mu = \nu$.*

Theorem 1.8 (Total variation distance $\equiv L_1$ norm)**.**

$$d(\mu, \nu) = \frac{1}{2} \| \mu - \nu \|_1 = \frac{1}{2} \sum_{t \in \mathbb{Z}_p} |\mu(t) - \nu(t)|$$

Theorem 1.9 (Variational formula)**.**

$$d(\mu, \nu) = \frac{1}{2} \max\{ \ |\mu(f) - \nu(f)| \ | \ f : \mathbb{Z}_p \to \mathbb{R} \ s.t. \ \ \| f \|_\infty \leq 1\}$$

i.e. it is the maximal difference in expections over all random variables f whose max values are ≤ 1.

Definition 1.10.

- **Information** of distribution μ is $I_\mu : \mathbb{Z}_p \to [0, \infty)$, $I_\mu(t) := -\ln(\mu(t))$
- **Entropy** of distribution μ is $H(\mu) = -\sum_{t \in \mathbb{Z}_p} \mu(t) \ln(\mu(t))$ i.e. the expected information.

Note: For uniform distribution λ, $H(\lambda) = \ln(p)$, whilst for dirac distribution δ_s, $H(\delta_s) = 0$.

Theorem 1.11 (Pinsker's Inequality)**.** *For arbitrary distribution $\mu : \mathbb{Z}_p \to [0, 1]$ and uniform distribution λ (so $H(\lambda) = \ln(p)$):*

$$\frac{1}{2(H(\lambda) + 1)} |H(\mu) - H(\lambda)| \leq d(\mu, \lambda) \leq \sqrt{2|H(\mu) - H(\lambda)|}$$

---

2 Dynamics

Definition 2.1. The **convolution** of $f, g : \mathbb{Z}_p \to [0, 1]$ is $(f * g)(t) := \sum_{s \in \mathbb{Z}_p} f(t \ominus s) g(s)$
Let $\mu^{*n} := \mu^{*(n-1)} * \mu$, with $\mu^{*0} := \delta_0$

Theorem 2.2. *Convolution has following properties*

- *Commutative:* $f * g = g * f$
- *Associative:* $f * (g * h) = (f * g) * h$
- *Bilinear:* $f * (\lambda g + \mu h) = \lambda f * g + \mu f * h$

Theorem 2.3. *The convolution of two distributions on* $\mathbb{Z}_p$ *is*

- *also a distribution*
- *for general* μ *and uniform dist* λ*, we have:* $\mu * \lambda = \lambda$
- *for dirac dist* δ_s*, have* $(\delta_s * \mu)(t) = \mu(t - s)$ *(so* $\delta_0 * \mu = \mu$*).*

Example 2.4. Fair pass the broccoli: $(\mu*\mu)(t) = \frac{\delta_2(t)}{4} + \frac{\delta_0(t)}{2} + \frac{\delta_{-2}(t)}{4}$ i.e. the probability distribution after two moves of the broccoli.

We can represent a finite group G ($|G| = n$) by its Cayley graph. A **transition kernel** P is an $n \times n$ matrix where $P_{ts} = P(t, s)$ is the probability of transitioning from s to t. On the cayley graph, the graph becomes a weighted directed graph. We require this since the distribution μ only tells us where to move if starting at 0. P is essentially n distributions, one for each element of G, arranged in matrix form. For the pass the broccoli example, we take $P(t, s) = \mu(t - s)$. So the distribution at each point is really the same as the one defined at 0, just with its domain translated. In the case $G = \mathbb{Z}_5$:

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

Represent μ as p-dim'l vec with s-th entry: $\mu(s)$. Let $\mu^{*0} = \delta_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}$ denote starting at 0, then the probability distribution after the first move is $P\mu^{*0}$, the first column of P, which is precisely μ. Similarly, starting at $i \in \mathbb{Z}_p$ is given by δ_i, and the distribution after one move is the i-th column of P. The distribution after two moves when starting at 0, is $\mu * \mu = P\mu = P^2\delta_0$ and after three moves: $\mu^{*3} = P^2\mu$.

Note: To generalise the above to several distributions, let $P_\mu(t, s) := \mu(t - s)$. Define distributions μ_1, μ_2,... and transition kernels P_{μ_1}, P_{μ_2},... as above. Then starting at 0, with δ_0, and taking the first move wrt μ_1, second move wrt μ_2 etc, we get the distribution after n-moves given by $\mu_1 * \cdots * \mu_n = P_{\mu_n} \ldots P_{\mu_1}\delta_0$.

Entropy grows under convolution:

Theorem 2.5. *Distributions* $\mu, \nu : \mathbb{Z}_p \to [0,1]$ *then entropy satisfies*

$$\max\{H(\mu), H(\nu)\} \leq H(\mu * \nu) \leq H(\mu) + H(\nu)$$

Proof. Via convexity of $x \log x$...? □

Definition 2.6.
- **Sumset** of $A, B \subset \mathbb{Z}_p$ is $A \oplus B := \{t + s \ : \ t \in A, s \in B\}$.
- Let $A^{\oplus n} := A^{\oplus(n-1)} \oplus A$, with $A^{\oplus 0} := \emptyset$.
- **Support** of distribution μ is $\text{spt}(\mu) = \{t \in \mathbb{Z}_p \ : \ \mu(t) > 0\} \subset \mathbb{Z}_p$

Note the following properties:

- $\max\{|A|, |B|\} \leq |A \oplus B| \leq |A||B|$
- (Cauchy-Davenport ineq) p prime, then: $\min\{|A| + |B| - 1, p\} \leq |A \oplus B|$

Theorem 2.7. *For distributions* μ, ν, *spt*$(\mu * \nu) =$ *spt*$(\mu) \oplus$ *spt*(ν)

Because the probability distribution over each point $t \in \mathbb{Z}_p$ is essentially the same, i.e. μ, we can denote:

Definition 2.8. A **random walk** on $\mathbb{Z}_p$ with n steps is the $\mathbb{Z}_p$-valued random variable $X_n := t_1 + \cdots + t_n$ for n $\mathbb{Z}_p$-valued random variables $t_1, \ldots, t_n$ identically distributed wrt μ (i.e. $\mathbb{P}(t_j = t) = \mu(t)\ \forall j$). Note the distribution of the sum of random variables is the convolution of each distributions, so $\mathbb{P}(X_n = t) = \mu^{*n}(t)$.

Definition 2.9.

$$\mathbb{P}(X_1 = s,\ X_n = t) := \mathbb{P}(X_1 = s)\mathbb{P}(X_n = t) = \mu(s)\mu^{*n}(t)$$

Here we take X_1 and X_n to be the random variables denoting the first step and n-th step of two distinct random walks. Hence they are independent, and the probability of the intersection of events on each is the product of the respective probabilities.

- For fixed $s \in \mathbb{Z}_p$, the probability n-th step is t given first step was s is:

$$\mathbb{P}(s + t_2 + \cdots + t_n = t) = \delta_s * \mu^{*(n-1)}(t)$$

- Because we defined $\mathbb{P}(X_1 = s,\ X_n = t)$ with X_1, X_n referring to distinct random walks, they will always be independent and hence

$$P(X_n = t \mid X_1 = s) := \frac{\mathbb{P}(X_1 = s,\ X_n = t)}{\mathbb{P}(X_1 = s)} = \mathbb{P}(X_n = t)$$

  I disagree with how this was defined in class since we should have $P(X_n = t \mid X_1 = s) = \mathbb{P}(s + t_2 + \cdots + t_n = t)$.

Definition 2.10. $\mu_\infty : \mathbb{Z}_p \to [0,1]$ is the **limit** of sequence of distributions $\mu_1, \mu_2, \cdots : \mathbb{Z}_p \to [0,1]$ if $\lim_{n\to\infty} \mu_n(t) = \mu_\infty(t)$. (Note if μ_∞ is a limit, it also a distribution).

Theorem 2.11 (Characterisation of limits). *μ_∞ is the limit of $\mu_1, \ldots,$ iff* $\lim_{n\to\infty} d(\mu_n, \mu_\infty) = 0$

Definition 2.12. μ is **ergodic** if $\lim_{n\to\infty} \mu^{*n}(t) = \lambda(t)$, for λ the uniform distribution.

Lemma 2.13. *If $A \subset \mathbb{Z}_p$ is not contained within a coset of a proper subgroup, then $\exists n \in \mathbb{N}$ s.t. $A^{\oplus n} = \mathbb{Z}_p$.*

Theorem 2.14. *For distribution μ, spt(μ) is not contained within coset of proper subgroup iff $\exists n \in \mathbb{N}$ s.t. spt(μ^{*n}) $= \mathbb{Z}_p$.*

Theorem 2.15 (Ergodic theorem). *μ is ergodic iff spt(μ) is not contained within a coset of a proper subgroup.*

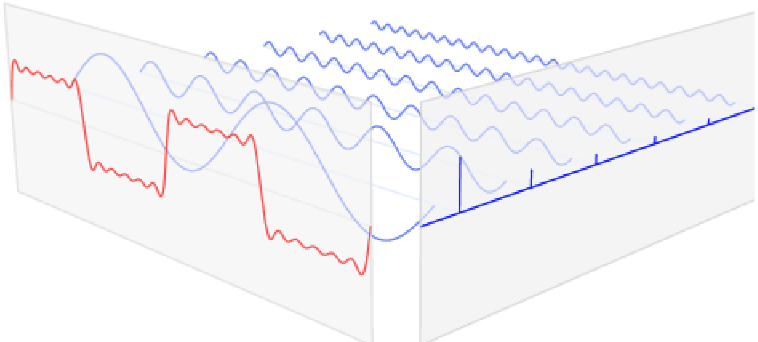
Definition 2.16.

- For **threshold** $\epsilon > 0$, the **mixing time** of a random walk driven by μ, is $n_{\text{mix}}(\epsilon) \in \mathbb{N}$ s.t. $\forall n \geq n_{\text{mix}}(\epsilon)$: $d(\nu * \mu^{*n}, \lambda) < \epsilon$. ($\nu$ denotes the initial starting distribution).
- For map ϕ s.t. $\phi(n) \to 0$ as $n \to \infty$, then μ is **mixing with rate** ϕ if $d(\nu * \mu^{*n}, \lambda) \leq \phi(n)\ \forall n$
- μ is **exponentially mixing** if for some $C > 0,\ 0 \leq \theta < 1$, it is mixing with rate $\phi(n) \leq C\theta^n$.

---

3 Harmonic Analysis

The following image is taken from Tuomas's notes, so all credit goes to him. I think it is the most interesting/instructive visualisation for Fourier analysis that I have come across:



Definition 3.1.
- The **discrete fourier transform** of $f : \mathbb{Z}_p \to \mathbb{C}$ is $\hat{f} : \mathbb{Z}_p \to \mathbb{C}$ such that:

$$\hat{f}(k) := \sum_{t=0}^{p-1} f(t) e^{\frac{-2\pi i k t}{p}}$$

- The maps $\chi_k(t) := e^{\frac{-2\pi i k t}{p}}$ are **characters** of $\mathbb{Z}_p$ (i.e. group homomorphism $\mathbb{Z}_p \to \mathbb{C}$).

Theorem 3.2 (Exponential sum formula)**.** *For $\theta \neq 0$,*

$$\sum_{t=0}^{p-1} e^{it\theta} = \frac{1 - e^{ip\theta}}{1 - e^{i\theta}}$$

Example 3.3.
- $\hat{\lambda} = \delta_0$
- $\hat{\delta_0}(k) = 1 \ \forall k$, while $\hat{\delta_s}(k) = e^{\frac{-2\pi i k t}{p}}$
- $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, then $\hat{\mu}(k) = \frac{1}{2} e^{\frac{-2\pi i k}{p}} + \frac{1}{2} e^{\frac{2\pi i k}{p}} = \cos(\frac{2\pi k}{p})$

The moral is that if $\hat{\mu}$ takes large values then μ is close to δ_0, if it takes small values (for most k) then it close to λ.

Theorem 3.4 (Fourier series theorem)**.** *Every map $f : \mathbb{Z}_p \to \mathbb{C}$ has **fourier expansion***

$$f(t) = \frac{1}{p} \sum_{k=0}^{p-1} \hat{f}(k) e^{\frac{2\pi i k t}{p}}$$

*Also called the **inverse fourier transform**.*

Definition 3.5. • The **inner product** of $f, g : \mathbb{Z}_p \to \mathbb{C}$ is $\langle f, g\rangle := \sum_t f(t)\overline{g(t)}$

- The L_2**-norm** is $\| f \|_2 := \sqrt{\langle f, f\rangle}$
- For $1 < p < \infty$, the L_p**-norm** is $\| f \|_p = (\sum_t |f(t)|^p)^{\frac{1}{p}}$

Lemma 3.6. • *Note the characters χ_k are orthonormal: $\langle \chi_k, \chi_l\rangle = 1$ if $k = l$, 0 otherwise.*

- *(Cauchy-Schwartz) $\forall f, g$: $|\langle f, g\rangle| \leq \| f \|_2 \| g \|_2$*
- *(Holders Inequality) For $1 < p, q < \infty$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$, then $|\langle f, g\rangle| \leq \| f \|_p \| g \|_q$*

Theorem 3.7 (Plancherels theorem/Parsevals identity)**.**

$$\langle f, g\rangle = \frac{1}{p}\langle \hat{f}, \hat{g}\rangle \qquad\qquad \| f \|_2 = \frac{1}{\sqrt{p}} \| \hat{f} \|_2$$

Theorem 3.8 (Convolution theorem)**.**

$$\widehat{f * g} = \hat{f}\hat{g}$$

---

4 Mixing Time

Theorem 4.1. • *(Upper Bound Lemma) For distribution μ, then $\forall n \in \mathbb{N}$*

$$d(\mu^{*n}, \lambda) \leq \frac{1}{2}\sqrt{\sum_{k \in \mathbb{Z}_p \setminus \{0\}} |\hat{\mu}(k)|^{2n}}$$

- *Generalisation:*

$$d(\mu_1 * \cdots * \mu_n, \lambda) \leq \frac{1}{2}\sqrt{\sum_{k \in \mathbb{Z}_p \setminus \{0\}} \prod_{j=1}^{n} |\hat{\mu_j}(k)|^2}$$

- *(Lower Bound Lemma) $\forall n \in \mathbb{N}$*

$$d(\mu^{*n}, \lambda) \geq \frac{1}{2}\sqrt{\frac{1}{p}\sum_{k \in \mathbb{Z}_p \setminus \{0\}} |\hat{\mu}(k)|^{2n}}$$

- *(Entropy)*

$$H(\mu^{*n}) \geq \ln(p) - (\ln(p) + 1)\sqrt{\sum_{k \in \mathbb{Z}_p \setminus \{0\}} |\hat{\mu}(k)|^{2n}}$$

Definition 4.2. Distribution μ has a **spectral gap** if $|\hat{\mu}(k)| < 1 \ \forall k \in \mathbb{Z}_p \backslash \{0\}$.

Note that by definition of μ being a distribution that: $\hat{\mu}(0) = 1$, and (by $\triangle$-ineq): $\hat{\mu}(k) \leq 1 \ \forall k$.

Theorem 4.3.

- *If μ has a spectral gap then it is exponentially mixing.*
- *μ has a spectral gap iff it is ergodic.*
- *(See Theorem 3.9 in Course notes for upper and lower bounds on mixing of the pass the broccolli distribution)*

5 Beyond $\mathbb{Z}_p$

For general finite group G, a **probability distribution** is defined similarly as $\mu : G \to [0, 1]$ s.t. $\sum_g \mu(g) = 1$. The total variation distance and L_1-norm are defined identically as for $\mathbb{Z}_p$.

Definition 5.1. For distributions f, g

- The **left convolution** $(f *_L g)(a) := \sum_{b \in G} f^{-1}(b^{-1}a)g(b)$
- The **right convolution** is $(f *_R g)(a) := \sum_{b \in G} f^{-1}(ab^{-1})g(b)$

These coincide for abelian G.

A random walk is defined similarly as $X_n = a_1 \ldots a_n$ (i.e. product, rather than additive, notation since G not necessarily abelian). Ergodicity of distribution μ is defined as before.

Theorem 5.2. *μ is ergodic iff the support of spt(μ) is not contained within a coset of a proper subgroup.*

Proof. Proof requires care for noncommutativity of group. □

Definition 5.3. $\mathbb{Z}_p^d = \{(t_1, \ldots, t_d) : t_j \in \mathbb{Z}_p\}$ which is a group using the usual vector addition.

Example 5.4. Ehrenfests Urn model is made up of d balls in 2 urns, whose states are modelled via a vector $v \in \mathbb{Z}_2^d$, one coordinate per ball, with $v_j = 0$ if j-th ball is in the left urn, $v_j = 1$ if in the right urn.

Each move selects a ball/coordinate randomly and swaps its urn/value. Based on starting configuration $t \in \mathbb{Z}_2^d$, a move is equivalent to adding a standard basis vec $e_j \in \mathbb{Z}_2^d$. The distribution for moves is then

$$\mu(t') = \begin{cases} \frac{1}{d} & t' = e_j \text{ for some } 1 \leq j \leq n \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The distribution after one move is then $\mu * \delta_t$, and after n moves: $\mu^{*n} * \delta_t$.

Definition 5.5. The **fourier transform** of $f : \mathbb{Z}_2^d \to \mathbb{C}$ is $\hat{f} : \mathbb{Z}_2^d \to \mathbb{C}$ s.t.

$$\hat{f}(k) := \sum_{t \in \mathbb{Z}_2^d} f(t)(-1)^{k \cdot t}$$

With this definition, all the Harmonic analysis such as the Plancheral/Parseval/Convolution theorems for $\mathbb{Z}_p$ hold for $\mathbb{Z}_2^d$ too.

Theorem 5.6 (Upper Bound Lemma for $\mathbb{Z}_2^d$)**.** $\forall n \in N$

$$d(\mu^{*n}, \lambda) \leq \frac{1}{2} \sqrt{\sum_{k \in \mathbb{Z}_2^d \setminus \{0\}} |\hat{\mu}(k)|^{2n}}$$

Lemma 5.7. *The fourier transform of the distribution in* (1) *is:* $\hat{\mu}(k) = 1 - \frac{2}{d}|\{1 \leq j \leq d \ : \ k_j = 1\}|$

---

6 Additional Questions and Notes

- Calculate the expectation of the convolution of two distributions. Show it is the sum of the expectations of each distribution.
- For group G, and G-valued random variable $f : \mathbb{Z}_p \to G$ wrt distribution $\mu : \mathbb{Z}_p \to [0, 1]$, the expectation is an element of the group algebra kG ($k = \mathbb{R}$): $\mu(f) = \sum_g \mu(g)g$.
- No spectral gap implies the support of dist is within a coset of a subgroup of G. Can different subgroups of G be characterised by this property? If so how? Does this generalise to quantum groups? i.e could one define "quantum subgroups" via sets for which distributions defined over them have no spectral gap?
- See comments for discussion of my issues with how $P(X_1 = s, X_n = t)$ is defined.