0 Introduction

Example 0.1. Perfect riffle shuffle $\sigma \in S_{52}$ s.t. $\sigma(j) := 2j \mod 51$. I.e.

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

The perfect riffle is deterministic, i.e. as a distribution $\mu : S_{52} \to [0,1]$, $\mu = \delta_\sigma$. It has order 8:

Proof. 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 $k = 1, 2, \ldots, 8$ shows 8 is the correct value. □

Example 0.2. Random transposition Randomly pick two cards and swap, with replacement after first choice (so doing nothing is possible). Has probability distribution $\mu : S_{52} \to [0,1]$,

$$\mu(\sigma) = \begin{cases} \sum^{52} \frac{1}{52^2} = \frac{1}{52} & \sigma = \text{id} \\ \frac{2}{52^2} & \sigma = (ij) \\ 0 & \text{otherwise} \end{cases}$$

Example 0.3. Pass the broccoli (i.e. $\mathbb{Z}_p$) $\mathbb{Z}_p = \{0, \ldots, p-1\}$, the broccoli starts at 0, and at each step the broccoli is passed either left or right with probability $\frac{1}{2}$, generating random walk on $\mathbb{Z}_p$.

Definition 0.4.

- **Sample space** is a set Ω (i.e. $\Omega = \mathbb{Z}_p$), and an **event** is element of the power set of the 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 the distribution to the event space: $\mu : P(\Omega) \to [0,1]$, such that for $A \subset \Omega$: $\mu(A) := \sum_{\omega \in A} \mu(\omega)$, with $\mu(\emptyset) := 0$. Its properties:
  - $\mu(\Omega) = 1$
  - $A \subset B \subset \mathbb{Z}_p \implies \mu(A) \leq \mu(B)$ "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 triple $(\Omega, P(\Omega), \mu)$ - Sample space, event space, probability measure.
- For a set S, an $S-$**valued random variable** is a map $X : \Omega \to S$.
- **Probability distribution wrt random variable** $X : \Omega \to S$ **and probability space** $(\Omega, P(\Omega), \mu)$ is the function $\widetilde{\mu} : X(\Omega) \to [0,1]$ 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.

1 Probability on $\mathbb{Z}_p$

Course notes use $t \oplus s := t + s \mod p$, $t \ominus s := t - s \mod p$. We use $+, -$ here.

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**

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

Definition 1.3. The **expectation** of random var $f : \mathbb{Z}_p \to \mathbb{C}$ wrt distribution $\mu : \mathbb{Z}_p \to [0,1]$ is:

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

Note the probability of event $A \subset \Omega$ is the expectation of the indicator function (rv) over A:

$$\mathbf{1}_A(t) := \begin{cases} 1 & t \in A \\ 0 & t \notin A \end{cases}$$

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

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

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

Definition 1.6.

- **Total variation distance** of dists 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)**.** *The tvd = the max difference in expectations over all rv's with max value (L_∞) at most 1:*

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

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: the uniform and dirac distributions have entropies:

$$H(\lambda) = \ln(p), \qquad H(\delta_s) = 0$$

Theorem 1.11 (Pinsker's Inequality)**.** *For distribution $\mu : \mathbb{Z}_p \to [0,1]$ and uniform dist λ*

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

2 Dynamics

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

Theorem 2.2.
- *Commutativity:* $f * g = g * f$
- *Associative:* $f * (g * h) = (f * g) * h$
- *Bilinear:* $f * (\lambda g + \mu h) = \lambda(f * g) + \mu(f * h)$
- *for distributions μ, ν, then $\mu * \nu$ is also a distribution.*
- *for distribution μ and uniform dist λ, we have:* $\mu * \lambda = \lambda$
- *for distribution μ and dirac dist δ_s, we have:* $(\delta_s * \mu)(t) = \mu(t - s)$*, hence:* $\delta_0 * \mu = \mu$.

Example 2.3. Fair pass the broccoli: $(\mu * \mu)(t) = \frac{\delta_2(t)}{4} + \frac{\delta_0(t)}{2} + \frac{\delta_{-2}(t)}{4}$ (distribution after two steps).

Theorem 2.4. *For distributions $\mu, \nu : \mathbb{Z}_p \to [0,1]$, "entropy grows under convolution":*

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

Definition 2.5.
- **Sumset** of $A, B \subset \mathbb{Z}_p$ is $A \oplus B := \{t + s \ : \ t \in A, s \in B\}$.
- Iterated sumsets: $A^{\oplus n} := A^{\oplus(n-1)} \oplus A$, with $A^{\oplus 0} := \emptyset$.
- **Support** of distribution μ is: $\mathrm{spt}(\mu) = \{t \in \mathbb{Z}_p \ : \ \mu(t) > 0\} \subset \mathbb{Z}_p$

Properties of sumsets:

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

Theorem 2.6. *For distributions μ, ν, $spt(\mu * \nu) = spt(\mu) \oplus spt(\nu)$*

Definition 2.7. 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 $\mathbb{Z}_p$-valued random variables $t_1, \ldots, t_n$ identically distributed wrt distribution μ, so for each j: $\mathbb{P}(t_j = t) = \mu(t)$. By probability, probability distribution for the sum of random variables is the convolution of the distributions: $\mathbb{P}(X_n = t) = \mu^{*n}(t)$.

Definition 2.8.

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

Note: here X_1 and X_n are two **distinct** rw's, so are independent, and therefore the probability of both events is the product of the probabilities of each event. Hence

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

This is stupid as we should have: $P(X_n = t \mid X_1 = s) = \mathbb{P}(s + t_2 + \cdots + t_n = t)$ (see below).

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

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

Definition 2.9. The **limit** of the sequence of distributions $\mu_1, \mu_2, \cdots : \mathbb{Z}_p \to [0,1]$ is $\mu_\infty : \mathbb{Z}_p \to [0,1]$ such that: $\lim_{n\to\infty} \mu_n(t) = \mu_\infty(t)\ \forall t \in \mathbb{Z}_p$. In this case: μ_∞ is also a distribution.

Theorem 2.10 (Characterisation of limits)**.** *μ_∞ is limit of $\mu_1, \mu_2, \cdots \iff \lim_{n\to\infty} d(\mu_n, \mu_\infty) = 0$*

Definition 2.11. Distribution μ is **ergodic** if $\lim_{n\to\infty} \mu^{*n}(t) = \lambda(t)$, for λ the uniform distribution.

Lemma 2.12. *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 2.13. *For distribution μ, the support spt(μ) is not contained within a coset of a proper subgroup $\iff \exists n \in \mathbb{N}$ such that spt$(\mu^{*n}) = \mathbb{Z}_p$.*

Theorem 2.14 (Ergodic theorem)**.** *μ is ergodic $\iff$ spt(μ) is not contained within a coset of a proper subgroup.*

Definition 2.15.

- The **mixing time** $n_{\text{mix}}(\epsilon) \in \mathbb{N}$ of a random walk driven by distribution μ, with a **threshold** $\epsilon > 0$, is such that: $d(\nu * \mu^{*n}, \lambda) < \epsilon\ \forall n \geq n_{\text{mix}}(\epsilon)$, $\forall$ starting dists ν.
- For $\phi : \mathbb{N} \to [0, \infty)$ s.t. $\lim_{n\to\infty} \phi(n) = 0$, μ **is mixing with rate** ϕ if

$$d(\nu * \mu^{*n}, \lambda) \leq \phi(n)\ \forall n, \nu$$

- μ is **exponentially mixing** if $\exists C \in (0, \infty), \theta \in [0, 1)$ such that μ is mixing with a rate function ϕ that is expontentially decaying: $\phi(n) \leq C\theta^n$.

3 Harmonic Analysis

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 $\mathbb{Z}_p \to \mathbb{C}$).

Lemma 3.2.

1. *$\hat{\mu}(0) = 1$ (by definition of μ being a distribution).*
2. *$\hat{\mu}(k) \leq 1\ \forall k$*
3. *(Exponential sum formula) For $\theta \neq 0$, $\sum_{t=0}^{p-1} e^{it\theta} = \frac{1-e^{ip\theta}}{1-e^{i\theta}}$*

Proof. Pt 2: $|\hat{\mu}(k)| = |\sum_t \mu(t) e^{\frac{-2\pi i k t}{p}}| \leq \sum_t |\mu(t) e^{\frac{-2\pi i k t}{p}}| = \sum_t |\mu(t)| \cdot |e^{\frac{-2\pi i k t}{p}}| = \sum_t |\mu(t)| = 1$ ☐

Example 3.3. Fourier transform of distributions. The moral is the more spread out $\hat{\mu}$ is, the more confined μ is, and vice versa.

- $\hat{\lambda} = \delta_0$

- $\hat{\delta_0} = \lambda$. So $\hat{\delta_0}(k) = 1\ \forall k$. More generally: $\hat{\delta_s}(k) = e^{\frac{-2\pi iks}{p}}$
- $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, then $\hat{\mu}(k) = \frac{1}{2}e^{\frac{-2\pi ik}{p}} + \frac{1}{2}e^{\frac{2\pi ik}{p}} = \cos(\frac{2\pi k}{p})$

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

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

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.
- *Characters χ_k are orthonormal, i.e.* $\langle \chi_k, \chi_l\rangle = \begin{cases} 1 & \textit{if } k = l, \\ 0 & \textit{otherwise.} \end{cases}$
- *(Cauchy-Schwartz Inequality) $\forall f, g$:* $|\langle f, g\rangle| \le \| f \|_2 \| g \|_2$
- *(Holders Inequality) For $1 < p, q < \infty$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$, then* $|\langle f, g\rangle| \le \| f \|_p \| g \|_q$

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 $\forall n \in \mathbb{N}$*

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

- *Generalisation:*

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

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

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

- *(Entropy)*

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

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

Theorem 4.3.
- *Distribution μ has a spectral gap $\implies$ it is exponentially mixing.*
- *Distribution μ has a spectral gap $\iff$ it is ergodic.*

5 Beyond $\mathbb{Z}_p$

For general group G, probability distributions, total variation distance, L_1-norms, ergodicity are defined identically. A random walks are denoted $X_n = a_1 \dots a_n$, as products rather sums since G not necessarily abelian.

Definition 5.1. For distributions f, g over G (the following coincide for abelian G):

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

Theorem 5.2. *μ is ergodic $\iff$ the support of $spt(\mu)$ is not contained within a coset of a proper subgroup.(Result as above, but requires new proof).*

Definition 5.3. $\mathbb{Z}_p^d = \{(t_1, \dots, t_d) : t_j \in \mathbb{Z}_p\}$ is a vector space, so group over 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, so is equivalent to adding a standard basis vec $e_j \in \mathbb{Z}_2^d$. Hence the distribution of possible moves is:

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

The distribution after one move, starting at configuratio $t \in \mathbb{Z}_2^d$ 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 of moves in Ehrenfests Urn model (in Equation (1)) is:* $\hat{\mu}(k) = 1 - \frac{2}{d}|\{1 \leq j \leq d \ : \ k_j = 1\}|$.