Analysis, Random Walks and Groups

Tuomas Sahlsten
Aalto University
Email: tuomas.sahlsten@aalto.fi

If you found any errors, typos, etc. let me know!

Version 1.5
February 15, 2023

Department of Mathematics and Systems Analysis, Aalto University

Contents

Chapter 1

Introduction

1.1 Shuffling cards

Consider the problem of card shuffling of a deck of 52 cards. Suppose we are shuffling the deck with a usual **riffle shuffle**: we cut the deck roughly from the middle into two packs and then "riffle" (interleave) the two packs together, see Figure 1.1.

Figure 1.1: A riffle shuffle, Johnny Blood, CC-by-sa 2.0

In a normal situation we will observe that the order of the cards in the deck begin to look random (see Figure 1.2 below) so that it is very hard to predict the order of the cards.

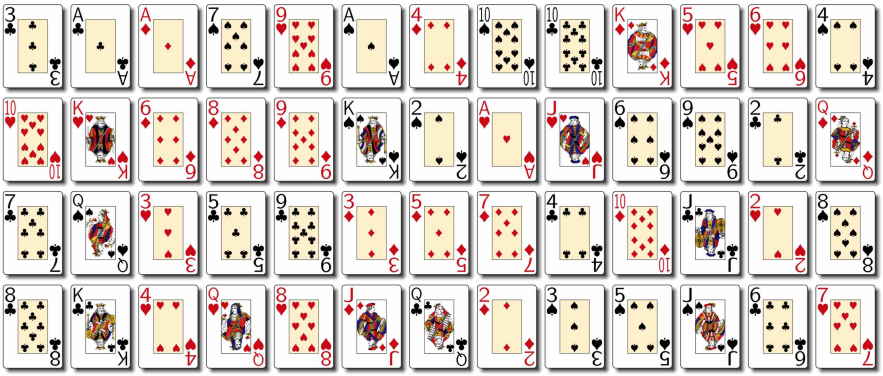

Figure 1.2: A relatively random looking order of cards in a deck of 52 cards.

It should be noted that if we are doing a "perfect shuffle" in the sense that we split the deck into exactly two piles of 26 cards and then manage to riffle the cards perfectly on top of each other (with no errors), then one can check that the shuffle returns to the initial state after 8 shuffles (this is a so called **perfect riffle / faro / dovetail shuffle**). However, in practise this can be hard to do, but some people can do it, see the YouTube video:

`https://www.youtube.com/watch?v=7lNk7bfkFq8`

In a normal situation there is always some errors in the shuffling and it is very hard to do a perfect riffle shuffle. These human errors introduces **randomness** into the situation and are exactly what explain why the deck eventually becomes very random.

In fact, Gilbert, Shannon, and Reeds (1955) did experiments on skilled human card shuffling and were able to make a probabilistic model that predicts very well how the order of the cards evolve as we do a riffle shuffle. In the late 80s / early 90s Persi Diaconis realised the card shuffling problem as question about **mixing of a random walk on the symmetric group** S_{52}, that is, the set of all permutations σ of $\{0, 1, \ldots, 51\}$. Then Diaconis applied methods from **harmonic analysis** and the statistical predictions done by Gilbert, Shannon, and Reeds to obtain estimates on how fast does a deck mix in a human riffle shuffle.


Science Times

The New York Times

Giant Telescope Prepares to Take Its First Look Skyward

In Shuffling Cards, 7 Is Winning Number

Computers confirm gamblers' intuition on how decks mix.

Wide Beliefs On Depression In Women Contradicted

European studies find few problems in hysterectomy, childbirth or menopause.

Figure 1.3: The New York Times article "*In Shuffling Cards, 7 Is Winning Number*", January 9, 1990 by Gina Kolata. © **The New York Times Archives**, available at `https://www.nytimes.com/1990/01/09/science/in-shuffling-cards-7-is-winning-number.html`

Diaconis obtained a very surprising answer in the case of riffle shuffle model done by Gilbert, Shannon and Reeds: it turned out that after roughly 6 shuffles the deck will still be quite ordered, but at the 7th shuffle the deck suddenly becomes very random, see for example *The New York Times* article from 1990 in Figure 1.3). The "very random" here means that almost every possible order of cards is possible in the deck (which can be formalised using something so called total variation distance to uniform or entropy, see later of the course).

Let us give some notation and model card shuffling as a random walk on the symmetric group. We say that $\sigma : \{0, 1, \ldots, 51\} \to \{0, 1, \ldots, 51\}$ is a **permutation** if σ is a bijection. Write

$$S_{52} := \{\sigma \text{ is a permutation of } \{0, 1, \ldots, 51\}\}$$

and equip S_{52} with the binary operation, which assigns to two permutations $\sigma, \sigma' \in S_{52}$ a new permutation, the **product**, defined for $j \in \{0, 1, \ldots, 51\}$ by

$$\sigma\sigma'(j) := \sigma(\sigma'(j)).$$

Hence, $\sigma\sigma'$ is just formally the **composition** $\sigma \circ \sigma'$ of the functions σ and σ', we just do not want to repeat the notation $\circ$ everywhere. In S_{52} let us write $e \in S_{52}$ just the **identity permutation** defined by

$$e(j) := j$$

that keeps each $j \in \{0, 1, \ldots, 51\}$ fixed.

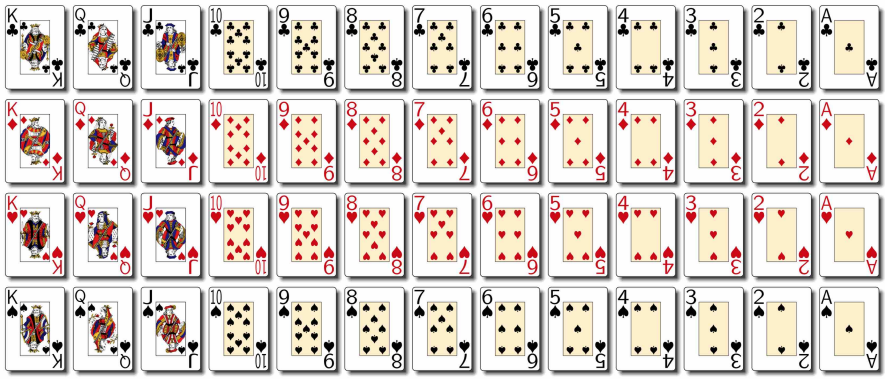

Figure 1.4: The initial state of the deck of cards.

To model card shuffling, let us now think the deck has 52 slots ordered from $0, 1, \ldots, 51$, where 0 denotes the top card and 51 denotes the bottom card. Suppose initially we have the following order for the cards (as in Figure 1.4):
we have for clubs:

$$K\clubsuit, Q\clubsuit, \ldots, A\clubsuit \text{ are in slots } 0, 1, \ldots, 12;$$

for diamonds:

$$K\diamondsuit, Q\diamondsuit, \ldots, A\diamondsuit \text{ are in slots } 13, 14, \ldots, 25;$$

for hearts:

$$K\heartsuit, Q\heartsuit, \ldots, A\heartsuit \text{ are in slots } 26, 27, \ldots, 38;$$

and for spades:

$$K\spadesuit, Q\spadesuit, \ldots, A\spadesuit \text{ are in slots } 39, 40, \ldots, 51.$$

Now if we perform permutations $\sigma \in S_{52}$ of $\{0, 1, \ldots, 51\}$, we move the card around from the initial order $0, 1, \ldots, 51$. For example, the top card of the deck is initially $K\clubsuit$ (corresponds to the label 1), so if the permutation $\sigma(0) = 51$, this means that we move $K\clubsuit$ to the bottom of the deck. Also, if we apply the identity permutation, we keep the deck order the same and the deck is not shuffled at all. Using this identification, every permutation $\sigma \in S_{52}$ is a shuffle of the deck.

If we have n permutations $\sigma_1, \sigma_2, \ldots, \sigma_n \in S_{52}$, then their product

$$\sigma_1\sigma_2 \ldots \sigma_n \in S_{52}$$

corresponds to a permutation where we have shuffled the deck n times with some choices of shuffles $\sigma_1, \sigma_2, \ldots, \sigma_n$. If the permutation is always the same, that is, $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$, we just write σ^n as the product n times.

Now the big question in the course is that how many times should we shuffle a deck of 52 cards to make it "sufficiently random"? What types of shuffling work best? These questions can be relevant when trying to maximise unpredictability of outcomes. As we mentioned earlier, in riffle shuffles we seem to be able to get random orders for cards. If we do not have randomness though, then it is possible that we end up to the original state. A classical example of such behaviour is the faro shuffle (also known as dovetail shuffle or perfect riffle shuffle):

Example 1.1 (Perfect riffle shuffle / dovetail shuffle / faro shuffle)

The perfect riffle shuffle splits the deck into exactly two piles of 26 cards and then riffles the cards perfectly on top of each other. How do we model this as a permutation $\sigma \in S_{52}$? Define

$$\sigma(j) := \begin{cases} 2j, & 0 \leqslant j \leqslant 25; \\ 2j - 51, & 26 \leqslant j \leqslant 51. \end{cases}$$

This is the same as saying that for $j = 0, 1, \ldots, 50$ we have:

$$\sigma(j) := 2j \mod 51$$

and $\sigma(51) = 51$ (the last card remains in the same position). Now for $0 \leqslant j \leqslant 25$, i.e. the first 26 cards on a pile, are put to even slots on $\{0, 1, \ldots, 51\}$, and then the rest ($26 \leqslant j \leqslant 51$) are put to odd slots, which is precisely a perfect riffle shuffle.

For perfect riffle shuffles, there is no randomness present, and it turns out that after 8 shuffles we return to the initial state of the deck:

Theorem 1.2

Let σ be the perfect riffle shuffle. Then after 8 times we return to the initial state of the deck, that is,

$$\sigma^8 = e.$$

Proof
By definition the perfect riffle shuffle σ assigns to $j \in \{0, 1, \ldots, 50\}$ the value

$$\sigma(j) = 2j \mod 51.$$

Thus for $k \in \mathbb{N}$ we have

$$\sigma^k(j) = 2^k j \mod 51.$$

Thus in order for us to have $\sigma^k(j) = j$, we need to find the minimal $k \in \mathbb{N}$ such that

$$1 \equiv 2^k \mod 51,$$

or equivalently, we want to find the minimal $k \in \mathbb{N}$ such that

$$2^k \equiv 1 \mod 51.$$

This is the so called "multiplicative order" in number theory, which we will not go into here, but in this case it is straightforward to compute by looking at the powers up to 8:

$$2^1 = 2 \equiv 2 \mod 51$$

$$2^2 = 4 \equiv 4 \mod 51$$

$$2^3 = 8 \equiv 8 \mod 51$$

$$2^4 = 16 \equiv 16 \mod 51$$

$$2^5 = 32 \equiv 32 \mod 51$$

$$2^6 = 64 \equiv 13 \mod 51$$

$$2^7 = 128 \equiv 26 \mod 51$$

$$2^8 = 256 \equiv 1 \mod 51$$

so $k = 8$ is the minimal k we looked for. Hence 8 perfect riffle shuffles returns the the deck to its initial state. □

However, in a real world situation, typically humans make errors in the riffle shuffles and thus at every step we choose 'random' permutations $\sigma_1, \sigma_2, \ldots, \sigma_n$ and the product

$$\sigma_1 \sigma_2 \ldots \sigma_n$$

after n steps tells us the distribution of the cards after n steps. However, as we mentioned earlier, in human trials we make errors in the riffle shuffles and these errors will eventually accumulate into the order of the cards in the deck to become very hard to predict. How would we formalise this? We can ask the following questions:

Questions 1.3

Q1. How many shuffles does it take for the deck to reach a given order of cards in the deck?

Q2. How many shuffles does it take for the card order to have gone through every possible combination of cards?

Q3. How many shuffles do we need to do such that the deck is close to "close to random"?

First we need to talk about what is a very "uncertain" order of cards. There are in total 52! different permutations $\sigma \in S_{52}$ and each permutation $\sigma \in S_{52}$, when applied to the initial order of cards we agreed in the beginning, gives out some new order of cards. Hence if we don't know which shuffle $\sigma \in S_{52}$ we use, then there are in total 52! different possible orders of the cards. Then we have no knowledge at all on the ordering of the cards: can define this as "very random" state of the cards. Later in the course we will see that this corresponds to the **uniform distribution** or **Lebesgue distribution** of possible permutations $\sigma \in S_{52}$.

Let us now give the first random shuffle example to demonstrate the situation, which is called the random transposition shuffle:

Random transpositions

Place the ordered deck of 52 cards on a table into a single row:

$$K\clubsuit, Q\clubsuit, \ldots, A\clubsuit, K\diamondsuit, Q\diamondsuit, \ldots, A\diamondsuit, K\heartsuit, Q\heartsuit, \ldots, A\heartsuit, K\spadesuit, Q\spadesuit, \ldots, A\spadesuit.$$

We call a permutation $\sigma \in S_{52}$ a **transposition** if it changes the places of two cards, that is, for some $i \neq j$ we have $\sigma(i) = j$ and $\sigma(j) = i$. The random transposition shuffle goes as follows: left hand chooses a random card with probability 1/52, and the right hand chooses a random card with probability 1/52. Then these cards are interchanged. If both hands chose the same card, nothing happens.

Formally this means that if the card i was chosen with probability 1/52 and the card j was chosen with probability 1/52, then the transposition σ that swaps these is chosen with probability

$$\frac{1}{52^2} + \frac{1}{52^2} = \frac{2}{52^2}.$$

If the same card is chosen, then $\sigma = e$, the identity permutation, so the probability of choosing that is 1/52.

Answers to all the Questions 1.3 is then that 270 random transposition shuffles is enough to make the deck sufficiently random as we will see in the final chapter of the course.

Let us give the example of Gilbert, Shannon and Reeds from 1955 on experiments on human riffle shuffles:

Riffle shuffle (Gilbert-Shannon-Reeds, 1955)

The riffle shuffle model by Gilbert-Shannon-Reeds consists of two steps like how humans typically do the riffle shuffle: we first do a random cut roughly from the middle and split the deck into two piles, and then do a riffle, which may not be completely perfect, to produce the shuffle.

(1) **Random cut:** Firstly, we choose $1 \leqslant k \leqslant 52$ randomly with probability:

$$\frac{\binom{52}{k}}{2^{52}}.$$

In probability this would mean that k has binomial distribution $\mathrm{Bin}(n, p)$ with $n = 52$ and $p = 1/2$. Then the player will have k cards on the left hand and $52 - k$ cards on the right hand.

(2) **Random riffle:** Now, given the randomly chosen k cards on the left hand of the player, and $52 - k$ cards on the right hand of the player, we choose a random card either from the bottom of the left pile with probability $k/52$ or and bottom of the right hand pile with probability $(52 - k)/52$.

Then we are remaining with two piles with x cards on the left pile and y cards on the right pile (in total $x + y$ cards, which is in this case 51 cards). Now we continue and choose a random card either from the bottom of the remaining left pile (of x cards) with probability $x/(x + y)$ or from the bottom of the right pile (of y cards) with probability $y/(x + y)$ and place that card on the top of the card chosen in the previous step.

Iterate this until we have gone through all the 52 cards. Then the resulting pile of 52 cards we produced gives us a random permutation $\sigma \in S_{52}$ that gives us this random order of cards.

Answers to all the Questions 1.3 is then that 7 riffle shuffles is enough to make the deck sufficiently random.

Borel and Cheron (1955) suggested also the following type of shuffle in the book of mathematics of Bridge:

Borel's shuffle

Remove the top card of the deck and then insert it into the deck into a random position (it could be any of the 52 positions with probability 1/52). Then also the bottom card of the deck is removed and inserted at a random position of the deck.

This is a relatively slowly mixing shuffle process and answers to all the Questions 1.3 is then that 465 Borel shuffles is enough to make the deck sufficiently random.

Finally we give another popular way to shuffle cards, which is the overhand shuffle:

Overhand shuffles

In the overhand shuffle we transfer a small number of cards at a time from the shuffles right hand to the left. The person shuffling slides a couple of cards from the top of the deck from their right hand to the left. Then we repeat this process until all the cards on the right hand are transferred. Thus cards near the top of the deck end up near to the bottom of the new deck. Since the packet sizes transferred are typically random, this will eventually mix up the deck.

More mathematically we can define overhand shuffle by first choosing k random cut points that split the deck into $k+1$ piles of cards. The sizes of these packets have random size and also the number $k+1$ of packets. Then the overhand shuffle just reverses the order of the packets on the deck producing a new order. Depending on the random $k+1$ packets, this gives a random permutation $\sigma \in S_{52}$.

Answers to all the Questions 1.3 is then that 2500 overhand shuffles is enough to make the deck sufficiently random (proved by R. Pemantle, 1988). This makes sense as overhand shuffle mixes up much more slowly than the riffle shuffle unless the packet sizes chosen are very small.

In all these cases, answering to Questions 1.3 requires us to understand what is the statistics of the **random** product permutation

$$\sigma_1 \sigma_2 \dots \sigma_n$$

for randomly chosen $\sigma_1, \sigma_2, \dots, \sigma_n \in S_{52}$ permutations, say, riffle shuffles, overhand shuffles, random transpositions or Borel shuffles. This is also known as understanding the **mixing of the random walk** $X_n = \sigma_1 \sigma_2 \dots \sigma_n$ on the group S_{52}..

1.2 Mutations in the gene order of chromosomes

One influential motivation for random walks on groups come from the evolution of deoxyribonucleic acid (DNA) sequences, which form a chromosome. Chromosomes can be found from the nucleus of every cell and form the central unit of heredity. Mathematically a *chromosome*, $\mathbf{c}$, is an element $\mathbf{c} \in \{1, 2, \ldots, n\}^m$, where n is the number of possible genes $g_1, \ldots, g_n$, and m is the length of the chromosome. Each gene g_j itself is a sequence (or a block) of DNA, see Figure 1.5 for an illustration.

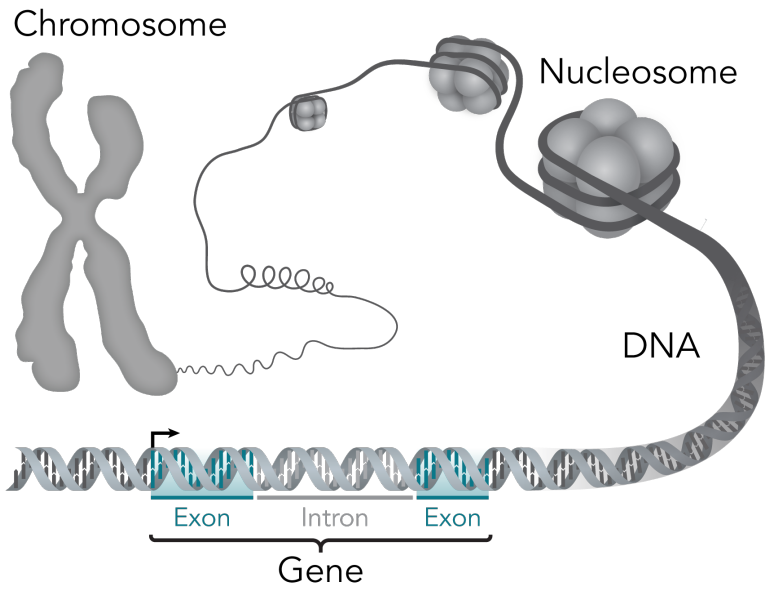

Figure 1.5: A chromosome is formed of blocks (genes) of DNA strings. Picture (c) Wikimedia Commons.

Each gene g_j, $j = 1, 2, \ldots, n$, itself is a sequence of nucleotides in DNA. In cell biology it has been observed that chromosomes $\mathbf{c}$ evolve in various transmutations:

(1) *substitutions*: a gene g_i in a chromosome $\mathbf{c}$ is substituted with another gene (length m remains the same)

(2) *insertions*: a gene is inserted into the chromosome $\mathbf{c}$ to some location (making the length $m + 1$)

(3) *deletions*: a gene is deleted from a chromosome $\mathbf{c}$ from some locartion (making the length $m - 1$)

(4) *inversions*: between two markers of the chromosome $\mathbf{c}$ invert the genes, that is, if $\mathbf{c} = g_{i_1}, \ldots, g_{i_m}$ and we choose two markers $i_j < i_k$, then reverse the order of the genes

$$g_{i_{j+1}}, g_{i_{j+2}} \cdots, g_{i_{k-1}}$$

ino

$$g_{i_{k-1}}, \ldots, g_{i_{j+2}}, g_{i_{j+1}},$$

and various other transformations. Thus if the number of genes is n, then an transmutation evolution of a chromosome **c** to a chromosome **c'** (say along random substitutions of genes) can be modelled as a random walk of random substitutions in the symmetric group S_n, when we identify each label $\{1, 2, \ldots, n\}$ by the corresponding gene $\{g_1, \ldots, g_n\}$.

Example 1.4 (Shuffling genes of flies)

An example for the study of this can be traced to the work of Durrett [6], where he considered the comparisons found by Ranz *et al.* [11] of the chromosomes found in two fly species: *Drosophila repleta* and *Drosophila melanogaster*. Durrett considered the chromosome 2 of *Drosophila repleta*, denoted by $\mathbf{c}_2$, and compared it to the chromosome arm $3R$ of *Drosophila melanogaster*, denoted by $\mathbf{c}_{3R}$. In these examples the number of genes $n = 26$. If we order the genes as $g_1, g_2, \ldots, g_{26}$, then what was observed was that the $\mathbf{c}_{3R}$ chromosome of *Drosophila melanogaster* is equal to the string of genes

$$\begin{aligned}\mathbf{c}_{3R} = (&g_{12}, g_7, g_4, \underline{g_2, g_3, g_{21}, g_{20}}, g_{18}, g_1, g_{13}, g_9, g_{16}, g_6, g_{14},\\ &\underline{g_{26}, g_{25}, g_{24}}, g_{15}, \underline{g_{10}, g_{11}}, g_8, g_5, \underline{g_{23}, g_{22}}, g_{19}, g_{17})\end{aligned}$$

where all the underlined parts match those of the chromosome 2 $\mathbf{c}_2$ of *Drosophila repleta*.

Durrett asked and made a model to answer the question: how many (random) inversions (see (4) above for a definition) from the chromosome $\mathbf{c}_2$ has happened to form $\mathbf{c}_{3R}$? Data analysis on the species suggests that all possible random inversions could have occured with equal probability, why is this true theoretically? Here the model for random inversion is that we choose the end-genes g_{i_j} and g_{i_k}, where $i_j < i_k$, at uniformly randomly, that is, each edge is chosen with equal probability. (Due to biological reasons, one has to do this as a 'continuous time process' in the sense that in practise one expects these random edges are chosen at times of a rate of a Poisson process, but now we just consider this as a discrete time process.)

Durrett then proved that, with high probability, when starting from a gene $\mathbf{c}_2$, then it will take around 85 random inversions of genes to form $\mathbf{c}_{3R}$, which was consistent with the data Ranz *et al.* [11] observed. More details can be found from Durrett's work [6], but this can all be formalised using the tools of this course using the total variation distance for a random process on the symmetric group S_{26} formed of random inversions, similarly to card shuffling above.

More examples of genetic applications can be read for example when comparing the genomes of a mouse and human:

Example 1.5 (Comparing human and mouse genomes)

The differences between the X chromosome of a human $\mathbf{c}_X(human)$ and of a mouse $\mathbf{c}_X(mouse)$ may be encoded as as follows:,

$$\mathbf{c}_X(human) = B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11$$
$$\mathbf{c}_X(mouse) = B1, -B7, B6, -B10, B9, -B8, B2, -B11, -B3, B5, B4$$

The notation above for $\mathbf{c}_X(human)$ means that the X chromosome of a human $\mathbf{c}_X(human)$ consists of 11 blocks of genes $Bk = g_{i_1} \dots g_{i_k}$ of various lengths, $k = 1, 2, 3, \dots, 11$, and so does the mouse, and the numbers denote how $\mathbf{c}_X(human)$ is shuffled to get $\mathbf{c}_X(mouse)$ as follows:

(1) The first block $\mathbf{c}_X(mouse)$ of the X chromosome of the mouse is equal to the first block of the human's X chromosome's block $B1$.

(2) The second block of $\mathbf{c}_X(mouse)$ is $-B7$ indicates that the second segment of a mouse is 7th block $B7$ human segment $\mathbf{c}_X(human)$ with the orientation reversed in $B7$ (hence -1).

(3) The third block in $\mathbf{c}_X(mouse)$ is $B6$ so it indicates that the 6th block $B6$ of a human $\mathbf{c}_X(human)$ is 3th mouse block in $\mathbf{c}_X(mouse)$

... etc.

The *parsimony approach* in evolutionary changes of the X chromosome asks about estimating the minimum number of reversals of the blocks in the mouse

$$\mathbf{c}_X(mouse) = B1, -B7, B6, -B10, B9, -B8, B2, -B11, -B3, B5, B4$$

back to that of a human

$$\mathbf{c}_X(human) = B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11?$$

Hannehalli and Pevzner [9] developed an algorithm to find that the minimal distance can be computed as 7, that is, 7 is the number of reversals needed to transform the X chromosome of a mouse into that of a human. However, there are thoudands of different solutions how the minimum can be achieved, also if one wants to develop the change in practise through mutations, one can ask if this can be modelled as a random walk using random inversions or other transmutations. This was done for example by Berestycki and Durrett [3] using more complex random models of shuffling genes based on graph theory.

1.3 Scrambling the Rubik's cube

Rubik's cube is a puzzle invented by the Hungarian architect and sculptor Ernő Rubik in 1974. The standard Rubik's cube has six squares at the center of each face, which are fixed to a core that enables the other 20 squares to rotate around.

Figure 1.6: Rubik's cube in three states: (1) solved state, (2) an application of a face rotation and (3) a relatively randomly looking scrambled version.

The standard Rubik's cube thus has 6 **faces** and each face has 9 smaller squares we will call **facets**. Thus the whole cube as in total $6 \times 9 = 54$ facets. We will say that the cube is in a **solved state** if all the facets of each corresponding face share the same colour.

A **move** of the cube is one of the following rotations of the 6 faces: 90 degrees, 180 degrees or -90 degrees. In a move of the cube the center facet attached to the mechanism does rotate around its center but will not change its face. The **Singmaster notation** to the Rubik's cube moves are the following:

Face rotations of the Rubik's cube		
90 degrees	180 degrees	-90 degrees
F - front clockwise	F^2 - front clockwise twice	F^{-1} - front counter-clockwise
B - back clockwise	B^2 - back clockwise twice	B^{-1} - back counter-clockwise
U - top clockwise	U^2 - top clockwise twice	U^{-1} - top counter-clockwise
D - bottom clockwise	D^2 - bottom clockwise twice	D^{-1} - bottom counter-clockwise
L - left face clockwise	L^2 - left face clockwise twice	L^{-1} - left face counter-clockwise
R - right face clockwise	R^2 - right face clockwise twice	R^{-1} - right face counter-clockwise

We can identify each of the Rubik's cube moves as a permutation on the set of non-center facets. Recall that there were in total $54 - 6 = 48$ non-center facets. Suppose at the initial state we are have the non-center facets of the solved Rubik's cube's assigned to these slots. More precisely this means that each facets corresponding to $0, 1, \ldots, 7$ share the same color, $8, 9, \ldots, 15$ share the same color, and so on. Then each of the Rubik's cube moves listed above is a bijection

$$\sigma : \{0, 1, \ldots, 47\} \to \{0, 1, \ldots, 47\},$$

i.e a permutation. That is, $\sigma \in S_{48}$. Note that some of the permutations of $\{0, 1, \ldots, 47\}$ is not possible as all the possible moves are listed above.

The key moves are the six face rotation $\{F, B, U, D, L, R\} \subset S_{48}$ since all the other can be obtained as their combinations (for example, $FF = F^2$ or $FFF = F^{-1}$). Hence all the possible states of the Rubik's cube can be identified with an element $\sigma \in S_{48}$ that is obtained as a finite composition of the maps F, B, Y, D, L, R. This gives rise to the Rubik's cube group:

Rubik's cube group

The **Rubik's cube group** $\mathcal{R}$ is the subgroup of S_{48} generated by the 6 face rotations $\{F, B, U, D, L, R\}$, that is,

$$\mathcal{R} = \langle F, B, U, D, L, R \rangle.$$

As in the case of card shuffling, we can model "scrambling" of the Rubik's cube by random choices of the permutations $\sigma \in \mathcal{R}$. One such scramble choice could be, for example, choosing the face rotation F with probability $1/2$ and the face rotation L with probability $1/2$. It is interesting to ask which random choices produce a very random scramble.

The most **uniformly random** state would then correspond to the random permutation with equal probability $1/|\mathcal{R}|$ chooses a permutation from $\mathcal{R}$. Since each $\sigma \in \mathcal{R}$ corresponds to a some order of the non-center facets of the Rubik's cube, then the probability of knowing what is the order of the colors of the non-center facets of the Rubik's cube will be very the smallest possible, that is, $1/|\mathcal{R}|$. Then with very high probability such a permutation will produce a very randomly looking scramble of the Rubik's cube, see Figure 1.6(3).

We can then ask questions such as:

Questions 1.6

Starting from the solved state of the Rubik's cube, apply the face rotation F with probability $1/3$, the face rotation L with probability $1/3$ and the face rotation U with probability $1/3$.

Q1. When we keep on applying these two rotations randomly, is the distribution of the Rubik's cube "close to random"?

Q2. If yes, then how many rotations we need to be close to random?

Q3. After how many rotations of the cube we can reach a given state of the cube with high probability?

Q4. After how many rotations of the cube we have reached every state of the cube with high probability?

These questions can be answered with the same probabilistic language as in the case of card shuffling: by considering random walks on the group $\mathcal{R}$. Thus, we need to find out what is the probability of achieving a given permutation with a given random process and how close is this probability distribution to the uniform distribution.

To attack Questions 1.6, we first need to us to understand the situation in a far simpler settings that do not involve such large groups like S_{48} and its large subgroups. The two key examples we will consider in this course will be based on the symmetric group S_4 of 4 elements, which we can use to model *dice rolling* in the next section (and it is also a very simple *non-abelian group*), and then the even simpler setting of cyclic additive group of p elements $\mathbb{Z}_p$

(also noted C_p in some literature), which is an *abelian group*. However, we emphasise that proving similar results in the simpler setting carry on also to the more abstract setting of Rubik's cube, card shuffling and even Lie groups beyond this course, and we will come back to these in the end of the course.

1.4 Dice rolling

This section will describe another model based on dice rolling, which, like the card shuffling, can be described as a random walk on a symmetric group. However, this time we only have to work with the symmetric group S_4, which makes the computations also easier towards the end of the course.

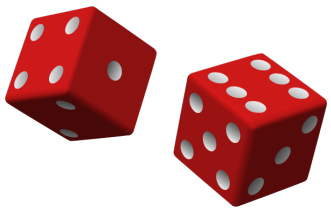

Figure 1.7: Two 6 sided dice (D6), picture from OpenClipart

When we throw a dice, we will perform a symmetry of the cube, which permutes the 8 cornes of the dice but in a way that an edge is sent on another edge. It turns out that there are exactly 24 *physically possible* symmetries that can be represented by an actual dice throw. Formally, you could also have symmetries that map the inside of the cube to the outside, but these are impossible physically and would require deformation of the dice.

Write $\mathcal{D}$ as the group of (physical) rotations of the dice, which we will describe now. A rotation of the dice will always have an *axis*, which can go through the dice in three possible ways:

(1) opposite faces,

(2) centers of opposite edges,

(3) opposite corners (i.e. *diagonal axis*).

In the case (1), there are two possible 90° rotations around the axis, and as we have three possible face pairs, this produces in total 6 possible 90° rotations. We can also go 180° rotation in these 3 cases, and in which case both ways would produce the same rotation. Thus in total we have 9 rotations in the case (1).

In the case (2), it is only possible to rotate by 180°, so we have in total 6 rotations in the case (2) as there are 6 possible pairs of edges on the opposite sides of the dice.

In the case (3), if we fix one such axis, the only possible rotations here are 120° one- or the other way. As there are four possible pairs of corners on the opposite side of the dice, we have 8 possible rotations.

These describe the $\mathcal{D}$ and show also that the number of elements in $\mathcal{D}$, $|\mathcal{D}| = 24$. This description also helps to see why actually $\mathcal{D}$ can be identified with the symmetric group S_4 (i.e. in the language of group theory, they are *isomorphic*). We do not formally prove this here, the proof is done e.g. in [8, Theorem 7.4], but we give now the basic idea. First of all, consider the set of 4 diagonals d_1, d_2, d_3, d_4 that go through the cube. When you now rotate

a cube using any of the rotations (1), (2) or (3) (i.e. an element $r \in \mathcal{D}$), then r will permute the diagonals $d_1, \ldots, d_4$ into some new order (take for example a dice in your hand and see what happens to the diagonals when you rotate, or Figure 1.8), so any r gives rise to some permutation $\sigma_r \in S_4$. It turns out that this identification is $1-1$: any permutation $\sigma \in S_4$ (i.e. a permutation of the diagonals $d_1, \ldots, d_4$) corresponds to one of the three rotations (1), (2) or (3). This can be done by first showing two simple rotations of type (1) (see Figure 1.8) correspond to two permutations $\alpha = (1234)$ and $\beta = (1423)$ in cycle notation of permutations, which then can be used to form a larger subgroup of S_4.

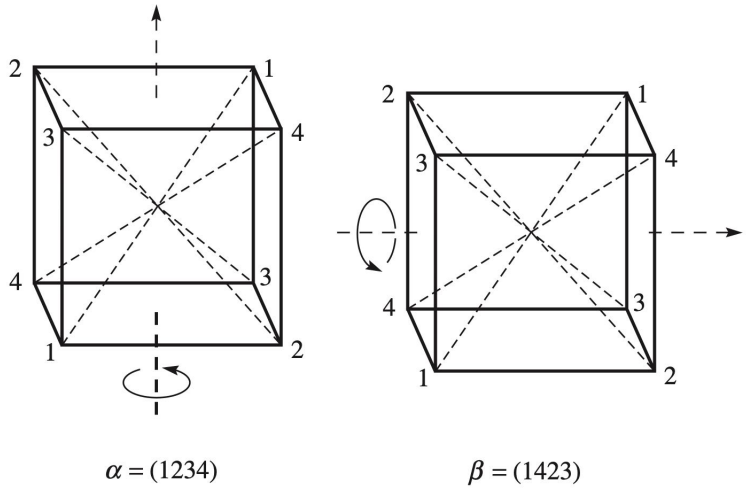

Figure 1.8: Two rotations α and β that will be used to identify $\mathcal{D}$ with S_4, image from [8, Figure 7.3].

This way we can think physical dice rolling as a random walk on the symmetric group S_4. One natural random walk here would be defined as follows, which we can answer these formally in the end using the tools of this course:

Questions 1.7

Consider the two rotations α and β from Figure 1.8 and the rotation corresponding to permutation $\gamma = (123)$. Then randomly roll dice applying the rotation α with probability 1/3 and β with probability 1/3, and γ with 1/3.

Q1. When we keep on applying these three rotations randomly, is the distribution of the dice "close to random"?

Q2. If yes, then how many rotations we need to be close to random?

Q3. After how many rotations of the dice we can reach a given state of the dice with high probability?

Q4. After how many rotations of the dice we have reached every state of the dice with high probability?

1.5 Pass the broccoli

In this section we will provide a simple setting/example, where we can prove analogous results as we asked in the case of card shuffling, Rubik's cube scrambling and dice rolling with not much background assumed from the reader. Suppose there is a dinner gathering with p people sitting in a round table such as the King Arthur and the Knights of the Round Table. Let us assign to each person sitting at the round table a label from $\{0, 1, \ldots, p-1\}$.

Figure 1.9: King Arthur and the Knights of the Round Table.

Suppose now that that **King Arthur** is sitting on the chair labeled 0 and has a bowl of broccoli, which they would like to pass to either to the knight to their left or to their right so that they can get their share of vitamins. However, King Arthur does not know to which one they would like to pass the bowl, so they will flip a coin. If the outcome is heads, they will pass is to the knight sitting to the right (label 1) or to the knight sitting to their left (label $p-1$). After this pass, the knight who received the bowl of broccoli (either knight 1 or knight $p-1$) will do the same: they will flip a coin and pass the bowl of broccoli either to the knight sitting to their right or to their left.

Figure 1.10: Passing the broccoli process with the Knights of the Round Table: the broccoli is first with the knight sitting on the chair 0, and then gives with probability 1/2 the broccoli either to the knight sitting to their left or to their right.

This forms a "random walk" of the bowl of broccoli around the table. Depending on your preferences, you may or may not want to taste the broccoli. Hence you could ask for example

what is the best place to sit on the table to avoid the bowl of broccoli for the longest possible time? Or is there a place where we could avoid it completely? Unfortunately, the answer to the second question is no: **there is no escape from the broccoli!** As we can see in Figure 1.11, the people the bowl of broccoli has visited begins to "spread" around the table (starting from the King Arthur sitting at 0).

Figure 1.11: The "orbit" of the broccoli: as time evolves, the broccoli has visited larger and larger arc around 0 eventually reaching to the other end of the table. And, **there is no escape!** Eventually the broccolis become uniformly distributed around the table.

However, we can still ask formally how fast is the spreading, or what is the distribution of it? For example, consider the following questions:

Questions 1.8

Q1. How many passes does it take for the bowl of broccoli to reach a given person?

Q2. How many passes does it take for the bowl of broccoli to reach every person?

Q3. How many passes do we need to take that the distribution of the people who may hold the bowl of broccoli is "close to random"?

It turns out that the answer to all of these questions is roughly p^2, where p is the number of people dining. We can prove these by formalising the questions as a long-time behaviour of a **ergodic random walk** (Chapter 3) on the **additive cyclic group** $(\mathbb{Z}_p, \oplus)$ (Chapter 2) and use **harmonic analysis** (Chapter 4) to estimate the distribution of the broccoli after n steps (Chapter 5), which us to find the optimal **mixing times** (Chapter 5) of the random walk (i.e. the pass the broccoli process).

It is also possible that the person at 0 chooses to use some other method of passing the broccoli. Suppose for example the number of people p is even and assume the person 0 flips a coin and throws the bowl of broccoli to either the person right to the person right of 0 (i.e. person 2) or the one left the left of 0 (person $p-2$). Then the person who got the bowl of broccoli does the same, and so on. In this process it turns out that the bowl of broccoli can only be on those people on the table whose label is an even number, i.e. the labels

$$\Gamma = \{0, 2, 4, \ldots, p-2\}.$$

In the language of group theory (see Chapter 2), the set Γ forms a **proper subgroup** of the additive cyclic group $(\mathbb{Z}_p, \oplus)$. In Chapter 3 we see that the random walk formed by such process that concentrates on this type of proper subgroup cannot have similar mixing properties as the pass the broccoli process described above (it is **not ergodic**, see Chapter 3). In particular, the distribution fo the broccoli cannot be uniform in the whole table, but only within the subgroup Γ.

The course will mostly concentrate on the pass the broccoli process modelled by a random walk on $(\mathbb{Z}_p, \oplus)$. This may sound too "easy" for some readers but $\mathbb{Z}_p$ is simple enough to make most of the theory of random walks and harmonic analysis very short, but still contain most of the crucial ideas. In Chapter 6, we will formalise these ideas in more general groups G, such as the symmetric group discussed above and other more complicated groups that lack the pleasant properties of the group $(\mathbb{Z}_p, \oplus)$. We will see that the general idea is still more or less the same as in $\mathbb{Z}_p$ and the success of the method relies more on what is known about the algebraic structure of the group G.

As a final note, we also mention that many of the ideas presented for $\mathbb{Z}_p$ also can be adapted even in continuous setting, such as in the analysis of dynamical systems or random walks on $\mathbb{R}$, $\mathbb{R}^d$ or even hyperbolic spaces but requires then more abstract measure theory (such as Lebesgue integration and Haar measures) and Fourier analysis on Euclidean spaces or hyperbolic spaces.

1.6 Goals of the course and references

It will be helpful throughout the course to keep in mind the following Intended Learning Outcomes (ILOs), which are also available on the course's official website. The Chapter(s) below each ILO indicate the location where the content related to each ILO is taught in these notes.

Intended Learning Outcomes

On successful completion of this course unit students will be able to:

1. Model card shuffling as a random walk on the symmetric group
   (Chapter 1: Introduction & Chapter 6: Applying the ideas beyond $\mathbb{Z}_p$)
2. Define total variation distances between probability distributions on the discrete circle, group $\mathbb{Z}_p$ and calculate these distances for simple examples in $\mathbb{Z}_p$,
   (Chapter 2: Probability theory on the group $\mathbb{Z}_p$)
3. Define convolutions of probability distributions on $\mathbb{Z}_p$, model random walks as iterated convolutions and estimate probabilities of events using iterated convolutions,
   (Chapter 3: Dynamics)
4. Define Fourier transforms on the group $\mathbb{Z}_p$ and estimate Fourier transforms of probability distributions and their convolutions on $\mathbb{Z}_p$,
   (Chapter 4: Harmonic Analysis)
5. Outline the calculations of computing total variation distances of convolutions of probability distributions to the uniform distribution on $\mathbb{Z}_p$ and alter these proofs in other examples with different constants or parameters,
   (Chapter 3: Dynamics, Chapter 5: Finding the mixing time)
6. Explain the key ideas of the theorems and methods presented in the course and describe how each component (harmonic analysis, random walks and group theory) come into play,
   (Chapter 5: Finding the mixing time)
7. Apply the methods presented in the course and prove similar results for analogous contexts such as random walks on higher dimensional lattices, matrix groups, models for card shuffling, Rubik's cube scrambling or dice rolling.
   (Chapter 1: Introduction, Chapter 6: Applying the ideas beyond $\mathbb{Z}_p$)

The ILOs will be basis of the summative examination of the course (final exam) and we will build the teaching materials and assignments around them. They can be helpful to keep track on what is your learning level and where there might still be things to improve.

This course does not have a fixed source and much of the material has been taken from various scattered sources. However, the key sources for the $\mathbb{Z}_p$ part and modelling card shuffling come from the following two books:

(1) **P. Diaconis**: *Group Representations in Probability and Statistics*, IMS Lecture Series volume 11, Institute of Mathematical Statistics, Hayward, California, 1988

(2) **F. Ceccherini-Silberstein, T. Scarabotti, F. Tolli**: *Harmonic Analysis on Finite Groups*. Cambridge University Press, New York, 2008.

The book (1) by Diaconis is the classical source and contains a vast amount of examples and goes very much beyond the scope of the course. The book (2) by Ceccherini-Silberstein, Scarabotti and Tolli has a far more followable Section 2: "*Two basic examples on abelian groups*" which discuss the group $\mathbb{Z}_p$ and also the torus $\mathbb{Z}_p^d$, which we will go through in this course more in detail.

Another useful source that goes more into the Harmonic Analysis side of the course is the book by Stein and Shakarchi:

(3) **E. Stein, R. Shakarchi**: *Fourier Analysis: An Introduction* (Princeton Lectures in Analysis), 2011

This book by Stein and Shakarchi goes again beyond the scope of this course but can provide helpful support for surrounding material in Fourier analysis and further example.

Finally more on the probability side is the following book by Lyons and Peres:

(4) **R. Lyons, Y. Peres**: *Probability on Trees and Networks*, Cambridge Series in Statistical and Probabilistic Mathematics, 2017

This book goes far close to applications such as on the theory of trees and networks, but can be helpful to provide a better background on probabilistic notions.

Finally, on the notions such as ergodicity and mixing of dynamical systems, a recommended introduction is given by the following book by Walters:

(5) **P. Walters**: *An Introduction to Ergodic Theory*, Springer, 1982

To summarise the core aims of the course: we will

- first concentrate on the **additive cyclic group** $(\mathbb{Z}_p, \oplus)$,
- develop **random walks** and **harmonic analysis** in $\mathbb{Z}_p$,
- prove **quantitative mixing rates** for random walks in $\mathbb{Z}_p$ using harmonic analysis,
- formalise these ideas for **card shuffling** models by using the symmetric group S_{52},
- apply the ideas to **more general groups**.

Good luck with the course!

1.7 Preliminary notations/definitions for the course

We will now give some preliminary notations and definitions we will use in the course. They should be familiar from basic courses on algebra, analysis, complex numbers and probability, but we will give them here for reference.

On the **analysis side**, we assume basic familiarity with concepts of analysis in the fields of real and complex numbers.

Definition 1.9 (Complex conjugate and modulus)

Complex numbers are denoted by

$$\mathbb{C} = \{z = x + iy : x \in \mathbb{R}, y \in \mathbb{R}\},$$

where $i^2 = -1$.

(1) The **complex conjugate** of a complex number $z = x + iy \in C$ is denoted by

$$\overline{z} = x - iy.$$

(2) The **modulus** of a complex number $z = x + iy$ is denoted by

$$|z| = \sqrt{x^2 + y^2}.$$

(3) The **exponential map** is defined by

$$e^{ix} := \cos(x) + i\sin(x), \quad x \in \mathbb{R},$$

which are complex numbers on the unit circle in $\mathbb{C}$.

Definition 1.10 (Limits)

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of complex numbers. Then $a \in \mathbb{C}$ is **limit** of the sequence, denoted by,

$$a = \lim_{n\to\infty} a_n$$

if for any $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ we have

$$|a_n - a| < \varepsilon.$$

On the **probability side** we will assume some basic familiarity with the probabilistic language, which we will give below. We will only do probability theory here in finite sets, so we do not need to assume any language on algebras or σ-algebras of sets.

Definition 1.11 (Sample space, events, probabilities and random variables)

(1) A **sample space** is any finite set Ω. Elements $\omega \in \Omega$ are called **outcomes**.

(2) Given a sample space Ω, we call any subset $A \subset \Omega$ an **event**.

(3) Given an outcome $\omega \in \Omega$, we associate to each ω a **probability** $\mathbb{P}(\omega) \in [0, 1]$ such that their sum over all the possible outcomes is 1 (full probability):

$$\sum_{\omega \in \Omega} \mathbb{P}(\omega).$$

Then we define the **probability of an event** $A \subset \Omega$ by

$$\mathbb{P}(A) := \sum_{\omega \in A} \mathbb{P}(\omega)$$

with the convention $\mathbb{P}(\varnothing) = 0$ for the empty set $\varnothing$.

(4) A S-valued **random variable** (for some set S) is a map $X : \Omega \to S$.

Example 1.12 (Coin tossing)

In the random trial of coin tossing, the outcomes are either heads or tails. Thus the sample space is

$$\Omega = \{\text{heads}, \text{tails}\}.$$

An event is a subset $A \subset \Omega$, so for example in a random trial getting heads is the singleton

$$A = \{\text{heads}\},$$

or getting heads or tails is the union

$$\{\text{heads}\} \cup \{\text{tails}\} = \Omega.$$

Then we can associate a probability $\mathbb{P}$ to events $A \subset \Omega$ by defining

$$\mathbb{P}(\text{heads}) = \frac{1}{2} \quad \text{and} \quad \mathbb{P}(\text{tails}) = \frac{1}{2}.$$

Then for example

$$\mathbb{P}(\text{we get heads or tails}) = \mathbb{P}(\Omega) = 1$$

or

$$\mathbb{P}(\text{we get heads and tails}) = \mathbb{P}(\{\text{heads}\} \cap \{\text{tails}\}) = \mathbb{P}(\varnothing) = 0$$

On the **algebraic side** this course assumes some familiarity with basic group theory, mainly the notations and some examples.

Definition 1.13 (Group)

A pair $(G, \cdot)$ is called a **group** if the **binary operation** $\cdot : G \times G \to G$ satisfies the axioms:

(1) **Closure**: If $a, b \in G$, then
$$a \cdot b \in G.$$

(2) **Associativity**: For all $a, b, c \in G$ we have
$$a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$

(3) **Identity element**: There exists an element $e \in G$ such that
$$e \cdot a = a \cdot e = a$$
for all $a \in G$.

(4) **Inverse element**: For each $a \in G$ there exists $a^{-1} \in G$ such that
$$a \cdot a^{-1} = a^{-1} \cdot a = e.$$

Definition 1.14 (Abelian group)

A group $(G, \cdot)$ is called a **Abelian group** if it satisfies the axiom:

(5) **Commutativity**: For each $a, b \in G$ we have
$$a \cdot b = b \cdot a.$$

Definition 1.15 (Subgroups)

Let $(G, \cdot)$ be a group. A subset $\Gamma \subset G$ is a **subgroup** if $(\Gamma, \cdot)$ is also a group. We sometimes write $\Gamma < G$ is Γ is a subgroup.

Definition 1.16 (Generators of groups)

Let $(G, \cdot)$ be a group and $A \subset G$ any subset. The **subgroup generated by** A is the smallest subgroup $\Gamma < G$ containing A. Equivalently, $\langle A \rangle$ set of all $a \in G$ that can be written as finite products of elements (or their inverses) from A:
$$\langle A \rangle = \{a_1 \cdot a_2 \cdot \cdots \cdot a_n : a_j \in A \text{ or } a_j^{-1} \in A, j = 1, \ldots, n, n \in \mathbb{N}\}.$$

Chapter 2

Probability theory on the group $\mathbb{Z}_p$

2.1 Group $\mathbb{Z}_p$

The vast majority of the course will concentrate on the additive cyclic group $(\mathbb{Z}_p, \oplus)$ for some $p \geqslant 2$, which is formally just the integers $0, 1, \ldots, p-1$ placed on the unit circle with equal distance to each other equipped with the binary operation between each points is just addition modulo p. Note that p is just an integer, so it does not have to be prime for example.

Definition 2.1 (Group $(\mathbb{Z}_p, \oplus)$)

Let $p \in \mathbb{N}$, $p \geqslant 2$, be an integer. We write formally

$$\mathbb{Z}_p = \{0, 1, \ldots, p-1\}.$$

We equip $\mathbb{Z}_p$ with the following operation: for $t, s \in \mathbb{Z}_p$, we define

$$t \oplus s := \begin{cases} t+s, & \text{if } 0 \leqslant t+s \leqslant p-1; \\ t+s-p, & \text{if } t+s \geqslant p. \end{cases}$$

Visually we can think about the group $(\mathbb{Z}_p, \oplus)$ as a circle with p points.

Figure 2.1: Visual representation of the group $\mathbb{Z}_{21}$: if we take $t \in \mathbb{Z}_{21}$ and add, say, 3, we will move 3 steps counterclockwise. Adding an inverse of an element $-3 = 21 - 3 = 18$, say, 3 means we move clockwise 3 steps from t.

Remark 2.2

This is the definition of $\mod p$ you may have seen in fundamental mathematics courses, and so $t \oplus s$ is with this definition the same as

$$t \oplus s := t + s \mod p.$$

The operation $t \oplus s$ means that we move s steps right on the circle $\mathbb{Z}_p$ from t.
For this purpose we sometimes also use the following "minus notation" to denote the movement to the left instead of right: if $t, s \in \mathbb{Z}_p$, we define

$$t \ominus s := t - s \mod p.$$

This means that we move s steps left from the point t on the circle $\mathbb{Z}_p$.

Theorem 2.3

$(\mathbb{Z}_p, \oplus)$ is a **group**, that is, it satisfies the group axioms

(1) **Closure**: If $t, s \in \mathbb{Z}_p$, then
$$t \oplus s \in \mathbb{Z}_p.$$

(2) **Associativity**: For all $t, s, r \in \mathbb{Z}_p$ we have
$$t \oplus (s \oplus r) = (t \oplus s) \oplus r.$$

(3) **Identity element**: There exists an element $0 \in \mathbb{Z}_p$ such that
$$0 \oplus t = t \oplus 0 = t$$
for all $t \in \mathbb{Z}_p$.

(4) **Inverse element**: For each $t \in \mathbb{Z}_p$ there exists $-t \in \mathbb{Z}_p$ such that
$$t \oplus -t = -t \oplus t = 0.$$
Formally the inverse is, by definition, for $t \in \mathbb{Z}_p$ the number
$$-t \mod p = -t + p = p - t \in \mathbb{Z}_p.$$

Proof
Exercise. □

Theorem 2.4

$(\mathbb{Z}_p, \oplus)$ is also **Abelian group**, that is, it also satisfies the commutativity axiom

(5) **Commutativity**: For each $t, s \in \mathbb{Z}_p$ we have

$$t \oplus s = s \oplus t.$$

Proof
Exercise. □

Example 2.5

(1) In $\mathbb{Z}_3$ we have $2 \oplus 3 = (2+3) - 3 = 2 = (3+2) - 3 = 3 \oplus 2$.

(2) In $\mathbb{Z}_4$ we have $2 \oplus 2 = (2+2) - 4 = 0$ so 2 is the inverse of 2 in $\mathbb{Z}_4$.

Exercise 2.6

(1) In $\mathbb{Z}_6$ find the inverse of 4.

(2) In $\mathbb{Z}_7$ find the inverse of 4.

An important notion in the theory of random walks on groups and also in Fourier analysis are the notions of **subgroups**, which are algebraic structures within $\mathbb{Z}_p$. Random walks could get trapped into these spaces and the analysis might need to be reduced to these cases separately.

Definition 2.7 (Subgroups)

(1) A subset $\Gamma \subset \mathbb{Z}_p$ is a **subgroup** if $(\Gamma, \oplus)$ is also a group.

(2) A typical way to construct a subgroup is to take the **subgroup generated by** $A \subset \mathbb{Z}_p$:

$$\langle A \rangle = \{t_1 \oplus \cdots \oplus t_n : t_i \in A \text{ or } -t_i \in A, i = 1, \ldots, n, n \in \mathbb{N}\}.$$

In other words, every element in $\langle A \rangle$ can be constructed as a finite sum of the elements in A.

Example 2.8

Suppose $p \in \mathbb{N}$ is even. Then all even numbers $\Gamma \subset \{0, 1, \ldots, p-1\}$ form a subgroup Γ in $(\mathbb{Z}_p, \oplus)$ and the subgroup Γ is generated by 2:

$$\Gamma = \langle 2 \rangle.$$

Indeed, if $t, s \in \mathbb{N}$ are even, then their sum $t + s$ is even so as p is even so is $t + s \mod p$.

Subgroups of $\mathbb{Z}_p$ depend heavily on the properties of the integer p, and we have the following:

Theorem 2.9 (Subgroups of $\mathbb{Z}_p$)

(1) If p is prime, then $\mathbb{Z}_p$ has only the "trivial" subgroups $\{0\}$ and $\mathbb{Z}_p$.

(2) If p is not prime, then all the subgroups of $\mathbb{Z}_p$ are $\{0\}$, $\mathbb{Z}_p$ and the generated subgroups

$$\langle t \rangle,$$

for those $t \in \{1, \ldots, p-1\}$, which divide p.

Proof
Left as an exercise. □

2.2 Probability distributions on $\mathbb{Z}_p$

The aim of this section is to define probability distributions on $\mathbb{Z}_p$ in order to formalise the notion of a *Random Walk*. Here we let $[0,1]$ to be the closed unit interval in $\mathbb{R}$.

We want to formalise the idea of choosing a point $t \in \mathbb{Z}_p$ "at random". For this purpose functions $\mu : \mathbb{Z}_p \to [0,1]$ that assign each point $t \in \mathbb{Z}_p$ a value $\mu(t) \in [0,1]$ such that they sum to 1 gives us formally the *probability* $\mu(t)$ of a point $t \in \mathbb{Z}_p$ to be chosen in a random trial.

Definition 2.10 (Probability distribution)

A function $\mu : \mathbb{Z}_p \to [0,1]$ is called a **probability distribution** if it satisfies

$$\sum_{t=0}^{p-1} \mu(t) = 1.$$

There are two fundamental probability distributions that we will often use in our analysis:

Definition 2.11 (Uniform and singular distributions)

(1) The **uniform (or Lebesgue) distribution** on $\mathbb{Z}_p$ is the function $\lambda : \mathbb{Z}_p \to [0,1]$ defined by

$$\lambda(t) := \frac{1}{p}, \quad t \in \mathbb{Z}_p.$$

(2) Fix some $s \in \mathbb{Z}_p$. Then **singular (or Dirac) distribution** at s is the function $\delta_s : \mathbb{Z}_p \to [0,1]$ defined by

$$\delta_s(t) := \begin{cases} 1, & \text{if } t = s; \\ 0, & \text{if } t \neq s. \end{cases}$$

Example 2.12

(1) The uniform distribution λ is a probability distribution:

$$\sum_{t=0}^{p-1} \lambda(t) = \sum_{t=0}^{p-1} \frac{1}{p} = p/p = 1.$$

(2) The singular distribution δ_s at $s \in \mathbb{Z}_p$ is a probability distribution:

$$\sum_{t=0}^{p-1} \delta_s(t) = \delta_s(s) = 1.$$

In the case of uniform distribution λ all $t \in \mathbb{Z}_p$ have equal chance of being chosen: each have probability $1/p$. However, in the case of singular distribution δ_s at $s \in \mathbb{Z}_p$ it will be with probability 1 that we choose s and with probability 0 that we choose any other $t \in \mathbb{Z}_p$.

Uniform and singular distributions are important as they give the two natural distributions often defined and in later sections we will see how they are related to *uncertainty*, *uniformity* and *entropy* of a probability distribution μ, that describe "how random" the choices we made according to μ are.

In the **pass the broccoli** process we will assign a natural probability distribution to describe the evolution of the broccoli, which we will given in the following exercise:

Exercise 2.13 (Pass the broccoli distribution)

Define a function $\mu : \mathbb{Z}_p \to [0, 1]$ by

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

Prove that μ is a probability distribution. Note that -1 is the inverse of 1 in $\mathbb{Z}_p$, that is, $p - 1$.

We will see later how this μ is related to the pass the broccoli process, but we can see already that if $t \in \mathbb{Z}_p$ is chosen randomly according to μ in Exercise 2.13, then either $t = 1$ with 50% probability or $t = -1$ with 50% probability. Thus this describes the location of the broccoli after the first step when the King Arthur (person at $t = 0$) has given the broccoli either the knight on their right $0 \oplus 1$ or left $0 \ominus 1$.

In probability theory, we often encounter the word **event** and **probability of an event**. Events are formally just subsets $A \subset \mathbb{Z}_p$ of the space of outcomes $\mathbb{Z}_p$ and probability is the sum of the weights $\mu(t)$ on each $t \in A$. The following definition makes these notions precise:

Definition 2.14 (Events and measures)

(1) An **event** is any subset $A \subset \mathbb{Z}_p$.

(2) Given a probability distribution $\mu : \mathbb{Z}_p \to [0, 1]$ and an event $A \subset \mathbb{Z}_p$ we define the **probability of the event** A as the finite sum:

$$\mu(A) := \sum_{t \in A} \mu(t).$$

Moreover, for $A = \varnothing$, we just define $\mu(\varnothing) := 0$. Thus this extends the definition of μ to all subsets of $A \subset \mathbb{Z}_p$ and we have for singletons $\{t\} \subset \mathbb{Z}_p$ for $t \in \mathbb{Z}_p$ that

$$\mu(\{t\}) = \mu(t).$$

In probability theory one says that μ, when defined on all events $A \subset \mathbb{Z}_p$, is a **probability measure** which satisfies the axioms of a (finitely additive) probability measure:

Theorem 2.15

Given a probability distribution $\mu : \mathbb{Z}_p \to [0,1]$, the quantity $\mu(A)$ defined on subsets $A \subset \mathbb{Z}_p$ satisfies

(1) **monotonicity**: If $A \subset B \subset \mathbb{Z}_p$, then

$$\mu(A) \leqslant \mu(B).$$

(2) **additivity**: if $A_1, \ldots, A_n \subset \mathbb{Z}_p$ are disjoint (that is, $A_k \cap A_\ell = \varnothing$ for $k \neq \ell$), then

$$\mu\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mu(A_k).$$

(3) $\mathbb{Z}_p$ **has probability** 1: we have

$$\mu(\mathbb{Z}_p) = 1$$

Proof
Left as an exercise. □

In literature a **probability measure** is often called a function satisfying the conditions (1), (2) and (3) of Theorem 2.15 but the condition (2) on additivity is replaced by **σ-additivity**: if $A_1, A_2, \cdots \subset \mathbb{Z}_p$ are disjoint, then

$$\mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mu(A_k).$$

In general additivity in more complicated spaces does not imply σ-additivity (such examples can be found from the literature on the field of *Measure Theory*). However, in our setting of $\mathbb{Z}_p$ this is true:

Theorem 2.16

If $\mu : \mathbb{Z}_p \to [0,1]$ is a probability distribution, then the associated quantity $\mu(A)$ on subsets $A \subset \mathbb{Z}_p$ is σ-additive.

Proof
Left as an exercise. □

When we have a probability distribution μ, we often have some function $f : \mathbb{Z}_p \to \mathbb{C}$ whose values we would like to "observe" with respect to μ. For example, we could take the following oscillating function:

$$f(t) = \begin{cases} 1; & t \text{ is even;} \\ -1; & t \text{ is odd.} \end{cases}$$

Then if we try to observe the value of f at random with respect to μ, we would ask for the *expected value* of f given the random choice of μ. If p is even and $\mu = \lambda$, the uniform distribution, then we will see that the expected value of f is 0 as there are equal number of odd and even numbers in $\{1, 2, \ldots, p-1\}$.

This leads to statistical concepts such as expectation/integral which measure the average value of $f(t)$ when we choose t randomly with respect to μ.

Definition 2.17 (Integral/expectation $\mu(f)$)

Let $f : \mathbb{Z}_p \to \mathbb{C}$ be a function. Then the **integral** (or **expectation**) of f with respect to a probability distribution μ on $\mathbb{Z}_p$ is given by the value

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

In literature notations

$$\int f \, d\mu \quad \text{or} \quad \mathbb{E}_\mu(f)$$

are used for these.

Example 2.18

Note that if $f = \mathbf{1}_A$, the indicator function of a set $A \subset \mathbb{Z}_p$, that is

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

Then the integral of $\mathbf{1}_A$ is the measure $\mu(A)$:

$$\mu(\mathbf{1}_A) = \int \mathbf{1}_A \, d\mu = \mathbb{E}(\mathbf{1}_A) = \mu(A) = \sum_{t \in A} \mu(t).$$

Example 2.19

Consider the function

$$f(t) = \begin{cases} 1; & t \text{ is even}; \\ -1; & t \text{ is odd}. \end{cases}$$

If p is even, then there are $p/2$ even and $p/2$ odd numbers in $\{0, 1, \ldots, p-1\}$. Thus

$$\lambda(f) = \sum_{t \in \mathbb{Z}_p} f(t)\lambda(t) = \frac{p}{2} \cdot \frac{1}{p} - \frac{p}{2} \cdot \frac{1}{p} = 0$$

A good way to construct new probability distributions from a given set of probability distributions is by taking their convex combinations:

Theorem 2.20 (Convex combinations)

Let $\mu_1, \ldots, \mu_n : \mathbb{Z}_p \to [0,1]$ be probability distributions and let $\alpha_1, \ldots, \alpha_n \in [0,1]$ be real numbers summing to one:

$$\sum_{j=1}^{n} \alpha_j = 1.$$

Then the function

$$\mu(t) = \sum_{j=1}^{n} \alpha_j \mu_j(t), \quad t \in \mathbb{Z}_p,$$

is a probability distribution.

Example 2.21 (Pass the broccoli as a convex combination)

The pass the broccoli process given by the probability distribution $\mu : \mathbb{Z}_p \to [0,1]$ can be written as a convex combination of the singular distributions at 1 and -1 as follows:

$$\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}.$$

Example 2.22 (Biased passing the broccoli)

Let $0 < \alpha < 1$ and define the following probability distribution:

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

Then μ_α describes a variation of the pass the broccoli process but where we are biased to one direction. For example, if $\alpha < 1/2$, then it is more likely we give the broccoli to the clockwise (i.e. use -1).

Exercise 2.23

What is the integral of

$$f(t) = \begin{cases} 1; & t \text{ is even;} \\ -1; & t \text{ is odd.} \end{cases}$$

with respect to the biased passing the broccoli distribution

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

2.3 Formalising uncertainty

The aim of this section is to formalise the notion of *uncertainty* within a probability distribution $\mu : \mathbb{Z}_p \to [0,1]$. The uncertainty means here that if we choose $t \in \mathbb{Z}_p$ randomly according to μ, then under much "uncertainty" it will be hard to predict the value of t. An extreme example is the uniform distribution $\lambda(t) = 1/p$, $t \in \mathbb{Z}_p$, see Figure 2.2.

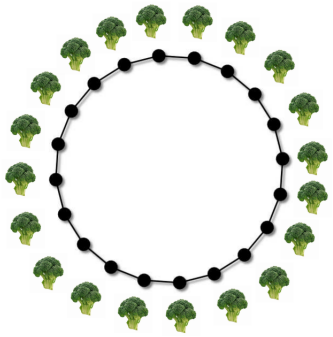

Figure 2.2: Uniform distribution on $\mathbb{Z}_p$: the location of the broccoli is with equal probability $1/p$ at a place $t \in \mathbb{Z}_p$. Therefore, the "uncertainty" is maximal: there is no information we can use to say t would be, say, more likely be at 0 than, say, 7.

If the distribution μ is not equal to λ, then we have some extra information about the location of $t \in \mathbb{Z}_p$. For example, if $\mu = \delta_0$, the singular distribution at 0, then we know with 100% certainty that $t = 0$. However, typically the distributions considered are not such but of something in between singular and uniform, see Figure 2.3

Figure 2.3: An example of a probability distribution $\mu : \mathbb{Z}_p \to [0,1]$ with the height of each "broccoli tower" telling us where it is more likely to find the broccoli. Note that some $t \in \mathbb{Z}_p$ have $\mu(t) = 0$, so we know with 100% certainty a randomly chosen $t \in \mathbb{Z}_p$ is not any of those. There seems to be concentration around one side of the circle $\mathbb{Z}_p$.

There are many ways to define "uncertainty" formally. One clear way to do this is to just simply measure the distance of μ to uniform λ using the *total variation distance*

Definition 2.24 (Total variation distance)

Let $\mu, \nu : \mathbb{Z}_p \to [0,1]$ be probability distributions. The **total variation distance** between μ and ν is defined by the maximal distances of the probabilities $\mu(A)$ and $\nu(A)$ over all the events $A \subset \mathbb{Z}_p$, that is,

$$d(\mu, \nu) := \max\Big\{|\mu(A) - \nu(A)| : A \subset \mathbb{Z}_p\Big\}.$$

Now total variation distance can be used to measure how uniform (or singular) a distribution μ is. In particular, we will be interested of the distance to uniform:

$$d(\mu, \lambda) = \max\Big\{|\mu(A) - \lambda(A)| : A \subset \mathbb{Z}_p\Big\}.$$

Recall that $\lambda(t) = 1/p$ for all $t \in \mathbb{Z}_p$.

For those who are familiar with metric spaces, it can be checked that the total variation distance provides a natural notion of *metric* in the set of all probability distributions:

Exercise 2.25 (Total variation distance is a metric)

Prove that total variation distance between two probability distributions satisfies the following properties of a **metric**: if μ, ν, τ are probability distributions, then

(1) They satisfy the **triangle inequality**:

$$d(\mu, \nu) \leqslant d(\mu, \tau) + d(\tau, \nu);$$

(2) Symmetry: $d(\mu, \nu) = d(\nu, \mu)$; and

(3) Equality: $d(\mu, \nu) = 0$ if and only if $\mu = \nu$.

Computing total variation distance directly using its definition of finding the measures $\mu(A)$ for all sets $A \subset \mathbb{Z}_p$ can be hard and it is helpful to use the following identity that links the total variation distance to the so called "L^1-distance":

Theorem 2.26 (Total variation distance and L^1 distance)

We have the following formula for the total variation distance

$$d(\mu, \nu) = \frac{1}{2} \sum_{t=0}^{p-1} |\mu(t) - \nu(t)|$$

Proof

Consider the set

$$B = \{t \in \mathbb{Z}_p : \mu(t) \geqslant \nu(t)\}.$$

(1) Firstly, we have that

$$\mu(B) - \nu(B) = \nu(B^c) - \mu(B^c).$$

Indeed, by definition $\mu(B) \geqslant \nu(B)$ so in particular

$$|\mu(B) - \nu(B)| = \mu(B) - \nu(B).$$

On the other hand, by the additivity of μ and ν, we have:

$$\begin{aligned}
&(\mu(B) - \nu(B)) - (\nu(B^c) - \mu(B^c)) \\
&= (\mu(B) + \mu(B^c)) - (\nu(B) + \nu(B^c)) \\
&= \mu(\mathbb{Z}_p) - \nu(\mathbb{Z}_p) \\
&= 1 - 1 \\
&= 0.
\end{aligned}$$

(2) Secondly, we see that the set B maximises the total variation distance:

$$d(\mu, \nu) = |\mu(B) - \nu(B)|.$$

Indeed, fix any set $A \subset \mathbb{Z}_p$. We use additivity of μ and ν to write

$$\mu(A) - \nu(A) = \sum_{t \in A \cap B} (\mu(t) - \nu(t)) + \sum_{t \in A \backslash B} (\mu(t) - \nu(t))$$

By definition, we have for all $t \in A \backslash B$ we have $\mu(t) - \nu(t) < 0$. Hence

$$\sum_{t \in A \backslash B} (\mu(t) - \nu(t)) < 0.$$

Hence

$$\mu(A) - \nu(A) \leqslant \sum_{t \in A \cap B} (\mu(t) - \nu(t)) = \mu(A \cap B) - \nu(A \cap B) \leqslant \mu(B) - \nu(B). \tag{2.1}$$

A symmetric argument (Exercise!) shows that

$$\nu(A) - \mu(A) \leqslant \nu(B^c) - \mu(B^c). \tag{2.2}$$

By **(1)** we thus have

$$\nu(A) - \mu(A) \leqslant \mu(B) - \nu(B).$$

This proves that

$$|\mu(A) - \nu(A)| \leqslant |\mu(B) - \nu(B)| \leqslant d(\mu, \nu)$$

so as $A \subset \mathbb{Z}_p$ is arbitrary we have

$$d(\mu, \nu) = |\mu(B) - \nu(B)|.$$

(3) Thirdly, we have that

$$|\mu(B) - \nu(B)| = \frac{1}{2} \sum_{t=0}^{p-1} |\mu(t) - \nu(t)|.$$

Indeed by **(2)**, we have

$$|\mu(B) - \nu(B)| = \frac{1}{2}|\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)|$$

By definition

$$\mu(B) - \nu(B) = \sum_{t\in\mathbb{Z}_p, \mu(t)\geqslant\nu(t)} \mu(t) - \nu(t) = \sum_{t\in\mathbb{Z}_p, \mu(t)\geqslant\nu(t)} |\mu(t) - \nu(t)|$$

and

$$\nu(B^c) - \mu(B^c) = \sum_{t\in\mathbb{Z}_p, \mu(t)<\nu(t)} -(\mu(t) - \nu(t)) = \sum_{t\in\mathbb{Z}_p, \mu(t)<\nu(t)} |\mu(t) - \nu(t)|$$

Hence

$$\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c) = \sum_{t=0}^{p-1} |\mu(t) - \nu(t)|.$$

(4) Finally, combining **(2)** and **(3)** then gives us the claim:

$$d(\mu, \nu) = |\mu(B) - \nu(B)| = \frac{1}{2}\sum_{t=0}^{p-1} |\mu(t) - \nu(t)|.$$

□

Theorem 2.26 introduces the L^1 distance between two probability distributions, and we can give a notation for it using the so called L^1 norms:

Definition 2.27 (L^1 norm)

Define the L^1 norm of a function $f : \mathbb{Z}_p \to \mathbb{R}$ by

$$\|f\|_1 = \sum_{t\in\mathbb{Z}_p} |f(t)|.$$

The difference of two probability distributions $\mu - \nu$ is a function that is defined by $t \in \mathbb{Z}_p$ by

$$(\mu - \nu)(t) = \mu(t) - \nu(t).$$

Hence Theorem 2.26 says the following:

$$d(\mu, \nu) = \frac{1}{2}\|\mu - \nu\|_1.$$

Other useful lemma, which we can use to bound total variation distance from below is the following that allows use "integrals"/"expectation" with respect to probability distributions. Recall Definition 2.17 for the definition. Here we need the notion of the L^∞ norm:

Definition 2.28 (L^∞ norm)

Define the L^∞ norm of a function $f : \mathbb{Z}_p \to \mathbb{R}$ by

$$\|f\|_\infty = \max\{|f(t) : t \in \mathbb{Z}_p\}.$$

Theorem 2.29 (Variational formula)

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

Proof
Exercise. □

Exercise 2.30

(1) Compute the total variation distance

$$d(\lambda, \delta_0)$$

between the uniform distribution λ and the singular distribution δ_0 at 0.

(2) Give a formula for the total variation distance

$$d(\delta_s, \delta_r)$$

in terms of $s, r \in \mathbb{Z}_p$.

(3) Define $\mu : \mathbb{Z}_p \to [0,1]$ by

$$\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}.$$

What is the total variation distance

$$d(\mu, \lambda)?$$

(4) Define $\mu : \mathbb{Z}_p \to [0,1]$ by

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

What is the total variation distance

$$d(\mu_\alpha, \mu_\beta)$$

for $0 < \alpha < \beta < 1$?

(5) Given two probability distributions $\mu, \nu : \mathbb{Z}_p \to [0,1]$ and $0 < \alpha < 1$, define their convex combination

$$\tau_\alpha := \alpha\mu + (1-\alpha)\nu.$$

Prove that the mapping $\alpha \mapsto d(\mu_\alpha, \mu)$, $\alpha \in [0,1]$, is continuous.

2.4 Information and entropy

Another way to measure uncertainty comes from information theory, which was formalised and popularised by Shannon inspired by ideas from statistical mechanics and Maxwell's equations. The basic concept here is *entropy*, which roughly speaking tells us information about how much uncertainty a random choice $t \in \mathbb{Z}_p$ with probability $\mu(t)$ has.

Definition 2.31 (Entropy)

The **entropy** of a probability distribution $\mu : \mathbb{Z}_p \to [0,1]$ is given by

$$H(\mu) = -\sum_{t=0}^{p-1} \mu(t) \log \mu(t).$$

Here log is taken in base e, that is, log = ln.

Entropy is formally the expected **information** $I_\mu(t)$ of μ at $t \in \mathbb{Z}_p$, which is defined by:

$$I_\mu(t) := -\log \mu(t), \quad t \in \mathbb{Z}_p.$$

Indeed, using the expected value / integral notation, we have that the entropy of μ is

$$H(\mu) = \mu(I_\mu) = \mathbb{E}_\mu(I_\mu).$$

Example 2.32

(1) In $\mathbb{Z}_p$ the uniform distribution λ has entropy

$$H(\lambda) = \log p.$$

Indeed, by definition

$$H(\lambda) = -\sum_{t=0}^{p-1} p^{-1} \log p^{-1} = (\log p) \sum_{t=0}^{p-1} \frac{1}{p} = \log p.$$

(2) In $\mathbb{Z}_p$, given $t_0 \in \mathbb{Z}_p$, the singular distribution δ_{t_0} at t_0 has entropy

$$H(\delta_{t_0}) = 0.$$

Indeed, as $\delta_{t_0}(t_0) = 1$ and 0 for $t \neq t_0$, we have

$$H(\delta_{t_0}) = -\sum_{t=0}^{p-1} \delta_{t_0}(t) \log \delta_{t_0}(t) = \log 1 = 0.$$

Exercise 2.33

For $0 < \alpha < 1$ define the probability distribution

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

Express the entropy

$$H(\mu_\alpha)$$

as a function of α. Compute also the total variation distances

$$d(\mu_\alpha, \delta_1) \quad \text{and} \quad d(\mu_\alpha, \delta_{-1})$$

as a function of α and compare the results.

2.5 Entropy and total variation distance

The entropy relates naturally to the total variation distance through **Pinsker's inquality** (not examinable, but the idea is useful):

Theorem 2.34 (Pinsker's inequality)

Let $\mu : \mathbb{Z}_p \to [0,1]$ be a probability distribution. Then the distance to uniform satisfies the following comparison to entropy distances

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

Recall that the entropy $H(\lambda) = \log p$, recall Example 2.32.

Proof

Let us first prove

$$\boxed{d(\mu,\lambda) \leqslant \sqrt{2|H(\mu) - H(\lambda)|}}$$

Define

$$r(t) := p\mu(t) - 1, \quad t \in \mathbb{Z}_p.$$

Recall that $H(\lambda) = \log p$ and $0 \leqslant H(\mu) \leqslant \log p$, so

$$|H(\mu) - H(\lambda)| = \log p - H(\mu).$$

Using the definition of $r(t)$ as $\sum_{t\in\mathbb{Z}_p} \mu(t) = 1$ we can write

$$\log p - H(\mu) = \frac{1}{p}\sum_{t\in\mathbb{Z}_p}[(1+r(t))\log(1+r(t)) - r(t)].$$

Logarithm satisfies the following equality for $x \geqslant -1$:

$$(1+x)\log(1+x) - x \geqslant \frac{1}{2} \cdot \frac{x^2}{1+x/3}.$$

Thus

$$\frac{1}{p}\sum_{t\in\mathbb{Z}_p}[(1+r(t))\log(1+r(t)) - r(t)] \geqslant \frac{1}{2p}\sum_{t\in\mathbb{Z}_p}\frac{r(t)^2}{1+r(t)/3}.$$

Since

$$\sum_{s\in\mathbb{Z}_p}(1+r(s)/3) = p + \frac{1}{3}\sum_{s\in\mathbb{Z}_p}(p\mu(s)-1) = p,$$

the right-hand side is equal to

$$\frac{1}{2p^2}\sum_{t\in\mathbb{Z}_p}\frac{r(t)^2}{1+r(t)/3}\sum_{s\in\mathbb{Z}_p}(1+r(s)/3).$$

Define

$$f(t) = \sqrt{r(t)^2/(1+r(t)/3)}, \quad \text{and} \quad g(t) = \sqrt{1+r(t)/3}.$$

Then

$$\frac{1}{2}\sum_{t\in\mathbb{Z}_p}\frac{r(t)^2}{1+r(t)/3}\sum_{s\in\mathbb{Z}_p}(1+r(s)/3) = \frac{1}{2}\sum_{t\in\mathbb{Z}_p}f(t)^2\sum_{s\in\mathbb{Z}_p}g(s)^2$$

Cauchy-Schwartz inequality for sums (see Theorem 4.13 in the later sections) gives that

$$\Big(\sum_{t\in\mathbb{Z}_p}f(t)g(t)\Big)^2 \leqslant \sum_{t\in\mathbb{Z}_p}f(t)^2\sum_{t\in\mathbb{Z}_p}g(s)^2$$

and we see that

$$f(t)g(t) = |r(t)|$$

so as $\lambda(t) = 1/p$ we have

$$\sum_{t\in\mathbb{Z}_p}f(t)g(t) = \sum_{t\in\mathbb{Z}_p}|r(t)| = \sum_{t\in\mathbb{Z}_p}|p\mu(t)-1| = p\sum_{t\in\mathbb{Z}_p}|\mu(t)-\lambda(t)| = p\|\mu-\lambda\|_1 = pd(\mu,\lambda).$$

Thus we have proved

$$\frac{1}{2}p^2 d(\mu,\lambda)^2 \leqslant p^2|H(\mu)-H(\lambda)|,$$

which gives the claim after dividing by p^2.

Now let us prove the other inequality

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

Write

$$B = \{t\in\mathbb{Z}_p : \mu(t) \geqslant \lambda(t)\} = \{t\in\mathbb{Z}_p : \mu(t) \geqslant 1/p\}.$$

Since $H(\lambda) = \log p$ we have

$$|H(\mu)-H(\lambda)| = \log p - \Big(-\sum_{t\in\mathbb{Z}_p}\mu(t)\log\mu(t)\Big) = \sum_{t\in\mathbb{Z}_p}\mu(t)\log\frac{\mu(t)}{\lambda(t)}.$$

and as $\log \frac{\mu(t)}{\lambda(t)} < 0$ for $t \notin B$, we have

$$\sum_{t\in\mathbb{Z}_p} \mu(t) \log \frac{\mu(t)}{\lambda(t)} \leqslant \sum_{t\in B} \mu(t) \log \frac{\mu(t)}{\lambda(t)}.$$

Add and substract:

$$\sum_{t\in B} \mu(t) \log \frac{\mu(t)}{\lambda(t)} = \sum_{t\in B} (\mu(t) - \lambda(t)) \log \frac{\mu(t)}{\lambda(t)} + \sum_{t\in B} \lambda(t) \log \frac{\mu(t)}{\lambda(t)} \tag{2.3}$$

Firstly as $\mu(t) \leqslant 1$ and $\lambda(t) = 1/p$, we have

$$\log \frac{\mu(t)}{\lambda(t)} \leqslant \log p,$$

which gives for the first term in (2.3) the bound

$$\sum_{t\in B} (\mu(t) - \lambda(t)) \log \frac{\mu(t)}{\lambda(t)} \leqslant (\log p) \sum_{t\in B} (\mu(t) - \lambda(t)).$$

For the second term in (2.3), we write $\mu(t)/\lambda(t) = 1 + (\mu(t) - \lambda(t))/\lambda(t)$ and obtain

$$\sum_{t\in B} \lambda(t) \log \frac{\mu(t)}{\lambda(t)} = \sum_{t\in B} \lambda(t) \log \Big(1 + \frac{\mu(t) - \lambda(t)}{\lambda(t)}\Big).$$

Applying $\log(1 + x) \leqslant x$ with $x = \frac{\mu(t)-\lambda(t)}{\lambda(t)}$ we obtain

$$\sum_{t\in B} \lambda(t) \log \Big(1 + \frac{\mu(t) - \lambda(t)}{\lambda(t)}\Big) \leqslant \sum_{t\in B} (\mu(t) - \lambda(t)).$$

Hence by (2.3) we have

$$\sum_{t\in B} (\mu(t) - \lambda(t)) \log \frac{\mu(t)}{\lambda(t)} \leqslant (\log p + 1) \sum_{t\in B} (\mu(t) - \lambda(t)).$$

Finally, the sum

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

which is equal to $d(\mu, \lambda)$ by Theorem 2.26. Hence as $H(\lambda) = \log p$ we have proved

$$|H(\mu) - H(\lambda)| \leqslant 2(H(\lambda) + 1) d(\mu, \lambda).$$

□

Chapter 3

Dynamics

3.1 Convolution on $\mathbb{Z}_p$

In the previous section we talked about uncertainty and randomness of the location of $t \in \mathbb{Z}_p$. For example, in the passing the broccoli process, we knew that the broccoli was at location $t = 0$ and then we randomly chose $t \in \mathbb{Z}_p$ according to the probability distribution μ given by

$$\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}.$$

In other words, we choose $t = 1$ with probability $1/2$ and $t = -1 = p-1$ with probability $1/2$, see Figure 3.1. However, the idea is to continue at each step passing the broccoli either to the person on their right/left. Hence we would like to talk about the *evolution* of the process.

Figure 3.1: Passing the broccoli process: the broccoli is first with the person sitting on the chair 0, and then gives with probability $1/2$ the broccoli either to the person sitting to the left or the right.

Assume we have started with the broccoli being with the person sitting at $t = 0$ and then they have given the broccoli either to their left or their right. This point is $t = 1$ or $t = -1$. Now, assume that at the next step we do the same: the person either at $t = 1$ or -1 gives the broccoli to the person at their left or their right. In the case $t = 1$, this would be $t = 0$ or $t = 2$, and in the case of $t = -1 = p-1$ this would be $t = 0$ or $t = p-2$. Is there a probability

distribution $\tilde{\mu} : \mathbb{Z}_p \to [0,1]$ that would give us the location of the broccoli after the second step?

We would need to define a new distribution $\tilde{\mu} : \mathbb{Z}_p \to [0,1]$ that would take into account the choice where we landed if choosing the location of the broccoli randomly with respect to μ. We can see that in the process we can only go once left or right, so we know that whatever $\tilde{\mu}$ is, it can only give positive values to the arc $\{p-2, p-1, 0, 1, 2\} \subset \mathbb{Z}_p$ since after two steps the broccoli could have only travelled at most 2 steps right or left. However, after the second step it is impossible $t = 1$ or $t = -1$ because we always give the broccoli to their left or right from $t = 1$ or $t = -1$ and those points are $p-2$, 0 and 2. Hence the distribution after second step should be concentrated on $\{p-2, 0, 2\}$.

This distribution $\tilde{\mu} : \mathbb{Z}_p \to [0,1]$ is the so called **convolution** $\mu * \mu$ of μ with itself.

Definition 3.1 (Convolution)

Let $f, g : \mathbb{Z}_p \to \mathbb{C}$ be functions. The **convolution** $f * g : \mathbb{Z}_p \to \mathbb{C}$ of f and g is

$$f * g(t) = \sum_{s=0}^{p-1} f(t \ominus s) g(s), \quad t \in \mathbb{Z}_p.$$

Recall that $t \ominus s = t \oplus (-s) = t - s \mod p$.

In the case of $f = g = \mu$ for a probability distribution $\mu : \mathbb{Z}_p \to [0,1]$ we see that

$$\mu * \mu(t) = \sum_{s=0}^{p-1} \mu(t \ominus s)\mu(s).$$

Recalling the notation of "expectation" from the previous chapter, we see that

$$\mu * \mu(t) = \mathbb{E}_\mu(f_t)$$

for $f_t : \mathbb{Z}_p \to [0,1]$ defined by

$$f_t(s) = \mu(t \ominus s), \quad s \in \mathbb{Z}_p.$$

Hence formally convolution $\mu * \mu(t)$ describes the expected value of the probability $\mu(t \ominus s)$ when $s \in \mathbb{Z}_p$ is chosen randomly according to μ. In the case of $\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$, we thus have

$$\begin{aligned} \mu * \mu(t) &= \sum_{s=0}^{p-1} \mu(t \ominus s)\mu(s) \\ &= \mu(t \ominus 1)/2 + \mu(t \oplus 1)/2 \\ &= \delta_2(t)/4 + \delta_0(t)/4 + \delta_0(t)/4 + \delta_{-2}(t)/4 \\ &= \delta_2(t)/4 + \delta_0(t)/2 + \delta_{-2}(t)/4, \end{aligned}$$

where in the last line we used the identities (for $t \in \mathbb{Z}_p$):

$$\delta_1(t \ominus 1) = \delta_2(t);$$

$$\delta_{-1}(t \ominus 1) = \delta_0(t);$$
$$\delta_1(t \oplus 1) = \delta_0(t);$$
$$\delta_{-1}(t \oplus 1) = \delta_{-2}(t).$$

Hence if we choose $t \in \mathbb{Z}_p$ randomly according to $\mu * \mu$, we have $t = 2$ with probability $1/2$, $t = 0$ with probability $1/2$ and $t = -2$ with probability $1/4$. Thus the resulting distribution $\mu * \mu$ is also a probability distribution, which is true in general:

Theorem 3.2

Suppose μ and ν are probability distributions on $\mathbb{Z}_p$. Prove that $\mu * \nu$ is a probability distribution $\mathbb{Z}_p$.

Proof
Exercise. □

Probabilistically, we can think about the convolution as an evolution of a random walk on $\mathbb{Z}_p$, where the transition is given by "transition kernel"

$$P(t, s) = \mu(t \ominus s),$$

which is the probability of the walk to transition from the state s to the state t. The transition kernel $P(t, s)$ when ordered $t, s \in \mathbb{Z}_p$ gives a matrix $p \times p$ matrix P with entries given by $P(t, s)$. This matrix is very important in the study of the evolution of the random walk and it is sometimes called the "*transfer operator*" of the process. In this course we will not pursue much into this, but for those interested in graph theory or dynamical systems might encounter this more.

Remark 3.3 (Probability vs convolution)

One good way to understand convolution is through the probabilistic idea: if $t_1 \in \mathbb{Z}_p$ has distribution μ and $t_2 \in \mathbb{Z}_p$ has distribution ν, then the sum $t_1 \oplus t_2$ has distribution $\mu * \nu$. Indeed, for any $t \in \mathbb{Z}_p$ we can write

$$\mathbb{P}(t_1 \oplus t_2 = t) = \mathbb{E}(\delta_t(t_1 \oplus t_2))$$

for the two dimensional Dirac mass function $\delta_t : \mathbb{Z}_p \times \mathbb{Z}_p \to [0,1]$, which is 1 when $t_1 \oplus t_2 = t$ and 0 otherwise. On the other hand as t_1 has distribution μ and t_2 has distribution ν, we have

$$\mathbb{E}(\delta_t(t_1 \oplus t_2)) = \int\int \delta_t(t_1 \oplus t_2)\, d\mu(t_1)\, d\nu(t_2)$$

and as $t_1 \oplus t_2 = t$ if and only if $t_1 = t \ominus t_2$, we obtain

$$\int\int \delta_t(t_1 \oplus t_2)\, d\mu(t_1)\, d\nu(t_2) = \int \mu(t \ominus t_2)\, d\nu(t_2) = \sum_{t_2 \in \mathbb{Z}_p} \mu(t \ominus t_2)\nu(t_2) = \mu * \nu(t).$$

Here formally the expectations are taken with respect to the **product distribution** $\mu \times \nu : \mathbb{Z}_p \times \mathbb{Z}_p \to [0,1]$, defined by $\mu \times \nu(t,s) = \mu(t)\nu(s)$, $(t,s) \in \mathbb{Z}_p \times \mathbb{Z}_p$.

Convolution also enjoys the following useful identities:

Theorem 3.4

For all $f, g, h : \mathbb{Z}_p \to \mathbb{C}$ we have

(a) **Commutativity:** $f * g = g * f$

(b) **Associativity:** $f * (g * h) = (f * g) * h$

(c) **Linearity**: if $\alpha, \beta \in \mathbb{C}$, then $f * (\alpha g + \beta h) = \alpha f * g + \beta f * h$

Proof
Exercise. □

Geometrically or information theoretically, convolution should be thought as "smoothening" operation. If we convolve, say, two probability distributions μ and ν, then the resulting convolution $\mu * \nu$ should somehow be more smooth than the original. Information theoretically this means that the "uncertainty" of choosing a point $t \in \mathbb{Z}_p$ according to $\mu * \nu$ increases. This can be formally done in the following theorem:

Theorem 3.5 (Entropy grows under convolutions)

If $\mu, \nu : \mathbb{Z}_p \to [0,1]$ are probability distributions, then the entropy

$$H(\mu * \nu) \geqslant \max\{H(\mu), H(\nu)\}.$$

Proof
Exercise. □

In the case of uniform distribution $\lambda(t) = 1/p$, $t \in \mathbb{Z}_p$, the convolution with any other $\mu : \mathbb{Z}_p \to [0,1]$ does not change the situation at all:

Theorem 3.6

For the uniform $\lambda(t) = 1/p$, $t \in \mathbb{Z}_p$, we have

$$\mu * \lambda = \lambda.$$

Proof
Exercise. □

If we convolve with a singular distribution $\delta_s : \mathbb{Z}_p \to [0,1]$ at some $s \in \mathbb{Z}_p$, then convolution works as a translation with $s \in \mathbb{Z}_p$:

Theorem 3.7

For the singular distribution δ_s at $s \in \mathbb{Z}_p$ we have

$$\delta_s * \mu(t) = \mu(t \ominus s).$$

Proof
Exercise. □

Exercise 3.8

For $0 < \alpha < 1$

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

compute $\mu * \mu$ and $\mu * \mu * \mu$.

Exercise 3.9

Prove that if $\mu, \nu : \mathbb{Z}_p \to [0,1]$ are probability distributions, then

$$H(\mu * \nu) \leqslant H(\mu) + H(\nu).$$

3.2 Sumsets in $\mathbb{Z}_p$ and relation to convolutions

Convolution is closely related to an additive combinatoric notion of a **sumset**:

Definition 3.10 (Sumset)

Let $A, B \subset \mathbb{Z}_p$. Then their **sumset** is the formal sum

$$A \oplus B = \{t \oplus s : t \in A, s \in B\}.$$

If $A \subset \mathbb{Z}_p$, let us write $|A|$ as the cardinality of A. That is, if $A = \{a_1, \ldots, a_n\}$, then $|A| = n$. Additive combinatorics concerns the relationship between the structure of A and B and the cardinality of their sumset $A \oplus B$.

Exercise 3.11

(1) Prove that

$$\max\{|A|, |B|\} \leqslant |A \oplus B| \leqslant |A||B|.$$

(2) Give an example of sets $A, B \subset \mathbb{Z}_p$ such that

$$|A \oplus B| = \max\{|A|, |B|\}.$$

(3) Give an example of sets $A, B \subset \mathbb{Z}_p$ which are not $\mathbb{Z}_p$ such that

$$|A \oplus B| = |A||B|.$$

If p is prime, then $\mathbb{Z}_p$ has no nontrivial subgroups (i.e. the only subgroups are $\{0\}$ and $\mathbb{Z}_p$.) Thus it is hard to have

$$|A \oplus B| = \max\{|A|, |B|\}$$

achieved. The Cauchy-Davenport inequality gives the common lower bound in these cases:

Exercise 3.12 (Cauchy-Davenport inequality)

If p is prime, then for all $A, B \subset \mathbb{Z}_p$ we have

$$|A \oplus B| \geqslant \min\{|A| + |B| - 1, p\}.$$

See [14, Proposition 5.4] for a proof.

The connection to sumsets comes from the notion of *support* of the convolution.

Definition 3.13 (Support of a probability distribution)

Let $\mu : \mathbb{Z}_p \to [0,1]$ be a probability distribution. The subset of $\mathbb{Z}_p$

$$\mathrm{spt}(\mu) = \{t \in \mathbb{Z}_p : \mu(t) > 0\}$$

is defined to be the **support** of μ.

If we consider the support of the convolution $\mu * \nu$, it reveals that the support is contained in the **sumset** of the supports $\mathrm{spt}(\mu)$ and $\mathrm{spt}(\nu)$. Thus this creates a link between the group theoretic properties of $\mathbb{Z}_p$ and the probabilistic heuristics behind convolution and we formalise it in the following theorem:

Theorem 3.14 (Support of convolutions and the sumset of supports)

Let $\mu, \nu : \mathbb{Z}_p \to [0,1]$ be probability distributions. Then the support

$$\mathrm{spt}(\mu * \nu) = \mathrm{spt}(\mu) \oplus \mathrm{spt}(\nu).$$

Proof

$\boxed{\subset}$ Let $r \in \mathrm{spt}(\mu * \nu)$. Then $\mu * \nu(r) > 0$. We need to prove that $r = t \oplus s$ for some $t \in \mathrm{spt}(\mu)$ and $s \in \mathrm{spt}(\nu)$. By the definition of the convolution

$$\mu * \nu(r) = \sum_{s \in \mathbb{Z}_p} \mu(r \ominus s)\nu(s)$$

so if this is positive then we know that

$$\mu(r \ominus s)\nu(s) > 0$$

for some $s \in \mathbb{Z}_p$. This is only possible if $\mu(r \ominus s) > 0$ and $\nu(s) > 0$ so $s \in \mathrm{spt}(\nu)$. Define

$$t := r \ominus s.$$

Then $\mu(t) > 0$ so $t \in \mathrm{spt}(\mu)$ and

$$r = t \oplus s.$$

Thus as $r \in \mathrm{spt}(\mu * \nu)$ is arbitrary, we have

$$\mathrm{spt}(\mu * \nu) \subset \mathrm{spt}(\mu) \oplus \mathrm{spt}(\nu).$$

$\boxed{\supset}$ Take $t \in \mathrm{spt}(\mu)$ and $s \in \mathrm{spt}(\nu)$. We want to prove that $\mu * \nu(t \oplus s) > 0$. By the definition of convolution

$$\mu * \nu(t \oplus s) = \sum_{r \in \mathbb{Z}_p} \mu(t \oplus s \ominus r)\nu(r).$$

If now $\mu(t \oplus s) = 0$, then $\mu(t \oplus s \ominus r)\nu(r) = 0$ for some $r \in \mathbb{Z}_p$, which means that

$$\mu(t \oplus s \ominus r) = 0$$

or

$$\nu(r) = 0.$$

Suppose $r = s$. Then

$$t \oplus s \ominus r = t \oplus s \ominus s = t.$$

If the first case happens, then

$$\mu(t) = \mu(t \oplus s \ominus r) = 0,$$

which is a contradiction with $t \in \mathrm{spt}(\mu)$. If the second case happens, then

$$\nu(s) = \nu(r) = 0,$$

which is also a contradiction with $s \in \mathrm{spt}(\nu)$. Hence $\mu * \nu(t \oplus s) > 0$ so $t \oplus s \in \mathrm{spt}(\mu * \nu)$. In particular, as $t \in \mathrm{spt}(\mu)$ and $s \in \mathrm{spt}(\nu)$ are arbitrary, we have

$$\mathrm{spt}(\mu * \nu) \supset \mathrm{spt}(\mu) \oplus \mathrm{spt}(\nu).$$

□

Exercise 3.15

For $0 < \alpha < 1$, let

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

Compute $\mathrm{spt}(\mu * \mu * \mu)$.

3.3 Convolutions model a random walk on $\mathbb{Z}_p$

The aim of this section is to introduce formally **random walks on the group** $\mathbb{Z}_p$, give notations for it and link it to convolutions.

The core idea behind a random walk is that we have a sequence of probability distributions $\mu_1, \mu_2, \cdots : \mathbb{Z}_p \to [0,1]$ and at step one, we choose a random point t_1 from $\mathbb{Z}_p$ according to μ_1, then add (using $\oplus$ operation) a random point t_2 to t_1 obtaining $t_1 \oplus t_2$. The distribution of this random variable will be $\mu_1 * \mu_2$. If we continue this process we obtain a random point $t_1 \oplus t_2 \oplus \cdots \oplus t_n \in \mathbb{Z}_p$ with distribution $\mu_1 * \mu_2 * \cdots * \mu_n$.

For example, the pass the broccoli process we have the distribution $\mu = \mu_1 = \mu_2 = \dots$ defined by

$$\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}.$$

Then if we choose a random point t_1 according to μ, we obtain either -1 or 1 with probability $1/2$, and then the next point t_2 is again chosen according to μ and added to the value t_1, which gives us $t_1 \oplus t_2$, and so on.

Commonly we deal with a single distribution $\mu = \mu_1 = \mu_2 = \dots$ as in the pass the broccoli process, and we will mostly concentrate on that case. The case of different distributions gives arise to a more complicated theory.

To formalise what we said above, let us introduce some notation.

Definition 3.16 (Iterated convolution)

Let μ be a probability distribution and $n \geqslant 1$. Then the *n-iterated convolution* is defined recursively by

$$\mu^{*n} = \mu^{*(n-1)} * \mu$$

with $\mu^{*0} := \delta_0$.

Thus we have

$$\mu^{*0} = \delta_0, \quad \mu^{*1} = \mu, \quad \mu^{*2} = \mu * \mu, \quad \mu^{*3} = \mu * \mu * \mu, \quad \text{and so on.}$$

Example 3.17

In the case of the passing the broccoli, the probability distribution

$$\mu^{*n} : \mathbb{Z}_p \to [0,1]$$

tells us the distribution of the broccoli after n steps: we know first that the Broccoli is with the person sitting on the chair 0, and then with probability $1/2$ we pass the Broccoli either left or right and iterate this n times.

We can use convolutions to define random walks on the group $\mathbb{Z}_p$ formally.

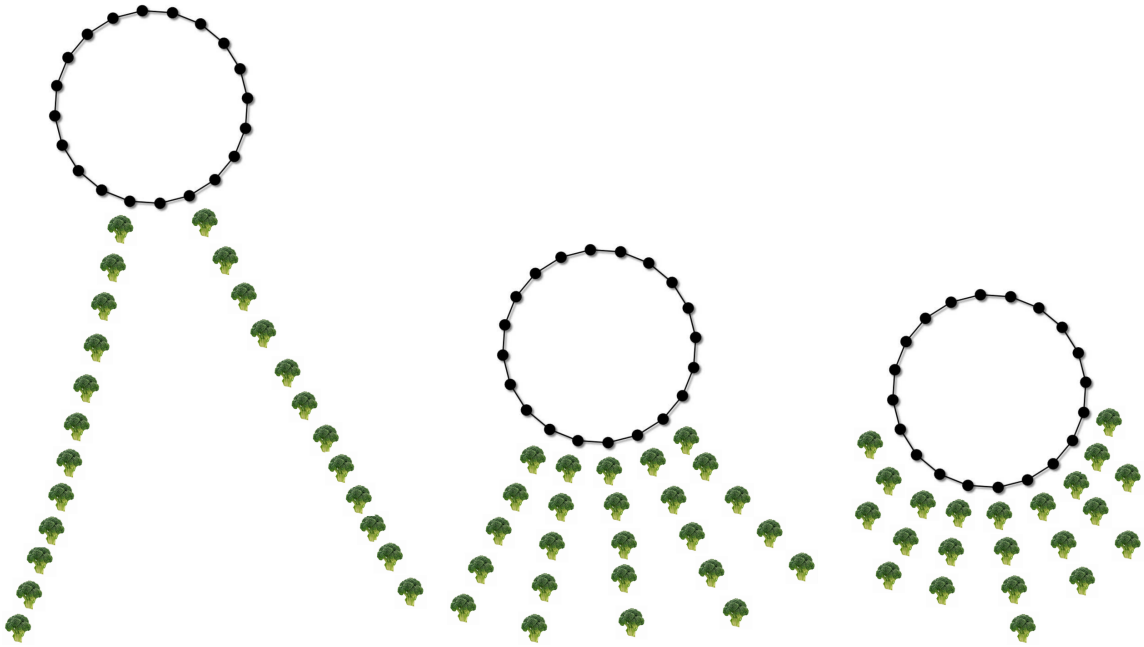

Figure 3.2: Visualisation of the iterated convolution. We first have the distribution $\mu^{*0} = \delta_0$, we know that he broccoli begins at the location 0. Then, in the first step, we take $\mu^{*1} = \mu$, that is, with equal probability the broccoli is located at 1 or -1. The height of the broccoli towers describe how likely is the broccoli found at $t \in \mathbb{Z}_p$ is. Initially they are in 1 and -1. Then the picture in the middle is the iterated convolution $\mu^{*2} = \mu * \mu$, which shows the broccoli starting to spread around the table, and finally $\mu^{*3} = \mu * \mu * \mu$. We observe "flattening" of the distribution from relatively singular μ to more uniformly distributed μ^{*3}.

Definition 3.18 (Random walk on $\mathbb{Z}_p$)

Fix some probability distribution $\mu : \mathbb{Z}_p \to [0, 1]$ and let $t_1, \ldots, t_n$ be $\mathbb{Z}_p$ valued random variables with for each $j = 1, 2, \ldots, n$ that

$$\mathbb{P}(t_j = t) = \mu(t), \quad t \in \mathbb{Z}_p.$$

We define a **random walk on** $\mathbb{Z}_p$ by the random variable

$$X_n := t_1 \oplus \cdots \oplus t_n \in \mathbb{Z}_p.$$

This new random variable X_n on $\mathbb{Z}_p$ has distribution μ^{*n}, that is,

$$\mathbb{P}(X_n = t) = \mu^{*n}(t), \quad t \in \mathbb{Z}_p.$$

We say that $X_1, X_2, \ldots$ is then **driven** by the probability distribution μ.

Here we see that for example

$$\mathbb{P}(X_1 = t) = \mu(t)$$

and if $A \subset \mathbb{Z}_p$, we have

$$\mathbb{P}(X_1 \in A) = \mu(A).$$

A common theme in random walks is to understand the return probabilities of random walks or probabilities we reach from one state to other. In the case of $\mathbb{Z}_p$, this could mean that what is the probability that $X_1 = s$ (we begin from the state $s \in \mathbb{Z}_p$) **and** after n steps we reach state $t \in \mathbb{Z}_p$, that is, $X_n = t$?

There are two ways to approach this. First of all, notice that the values of X_1 and X_n **statistically independent** of each other, that is, $X_1 = s$ and $X_n = t$ do not influence each other. This is because we have

$$X_1 := t_1'$$

for some $t_1' \in \mathbb{Z}_p$ with distribution μ and

$$X_n := t_1 \oplus \cdots \oplus t_n$$

for some $t_1, \ldots, t_n$ with distribution μ, but here the random variables $t_1, t_1' \in \mathbb{Z}_p$ may not be the same, they just have the same distribution μ.

Thus we can define the probability of the event that $X_1 = s$ and $X_n = t$ as follows:

Definition 3.19 (Probability of the event $X_1 = s, X_n = t$)

If $s, t \in \mathbb{Z}_p$, then define

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

Remark 3.20 (Fixing initial $t_1 = s$ in X_n)

If we would like to consider the probability when assuming the first element in X_n is s, that is, then we are fixing $t_1 = s$ in X_n, then there is dependence: we would be asking the probability of the event

$$\mathbb{P}(s \oplus t_2 \oplus \cdots \oplus t_n = t),$$

which is by definition as $s \oplus t_2 \oplus \cdots \oplus t_n$ has distribution $\delta_s * \mu^{*(n-1)}$ given by

$$\delta_s * \mu^{*(n-1)}(t) = \mu^{*(n-1)}(t \ominus s).$$

Remark 3.21 (Conditioning $X_1 = s$)

We need to emphasise here that $\mathbb{P}(X_1 = s, X_n = s)$ does not mean **conditional probability** you may have seen in a probability course. If we write

$$\mathbb{P}(X_n = t | X_1 = s) \tag{3.1}$$

we mean the probability of the event that the random walk $X_1, X_2, X_3, \ldots$ in $\mathbb{Z}_p$ driven by a probability distribution $\mu : \mathbb{Z}_p \to [0, 1]$ satisfies $X_n = t$ **assuming** that the same walk started with $X_1 = s$. This is more information about the location and it is possible that $X_n = t$ might be harder/easier to achieve if we have $X_1 = s$. Formally one defines

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

which as $\mathbb{P}(X_1 = s) = \mu(t)$ is equal to

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

Example 3.22

Define

$$\mu = \frac{1}{2}(\delta_1 + \delta_{-1}),$$

i.e. the pass the broccoli random walk. Then

$$\begin{aligned}
\mathbb{P}(X_1 = 1, X_2 = 1) &= \mu(1)\mu * \mu(1) \\
&= \frac{1}{2}\mu * \mu(1) \\
&= \frac{1}{2}\sum_{s \in \mathbb{Z}_p} \mu(1 \ominus s)\mu(s) \\
&= \frac{1}{2}(\mu(1 \ominus 1)\mu(1) + \mu(1 \oplus 1)\mu(-1)) \\
&= 0
\end{aligned}$$

However, if we assume the first summand in X_2 is 1, then we are computing the probability

$$\mathbb{P}(1 \oplus t_2 = 1) = \mathbb{P}(t_2 = 0) = \mu(0) = 0.$$

Finally, the conditional probability

$$\mathbb{P}(X_2 = 1 | X_1 = 1) = \frac{\mathbb{P}(X_1 = 1, X_2 = 1)}{\mathbb{P}(X_1 = 1)} = 0.$$

Example 3.23

Define

$$\mu = \frac{1}{2}(\delta_1 + \delta_{-1}),$$

i.e. the pass the broccoli random walk. Then

$$\begin{aligned}
\mathbb{P}(X_1 = 1, X_2 = 2) &= \mu(1)\mu * \mu(2) \\
&= \frac{1}{2}\mu * \mu(2) \\
&= \frac{1}{2}\sum_{s\in\mathbb{Z}_p} \mu(2 \ominus s)\mu(s) \\
&= \frac{1}{2}(\mu(2 \ominus 1)\mu(1) + \mu(2 \oplus 1)\mu(-1)) \\
&= \frac{1}{2}(\mu(1)\mu(1) + \mu(3)\mu(-1)) \\
&= \frac{1}{2} \cdot (\frac{1}{2} \cdot \frac{1}{2} + 0) \\
&= \frac{1}{8}.
\end{aligned}$$

Exercise 3.24

Let $X_1, X_2, \ldots,$ be the random walk driven by $\mu_\alpha = \alpha\delta_1 + (1-\alpha)\delta_{-1}$. Compute the probabilities

(1) $\mathbb{P}(X_1 = 1, X_2 = 2)$.

(2) $\mathbb{P}(X_1 = 0, X_2 = -2)$.

(3) $\mathbb{P}(X_1 = 1, X_4 = 1)$.

Compare these to the probabilities

(1) $\mathbb{P}(1 \oplus t_2 = 2)$.

(2) $\mathbb{P}(0 \oplus t_2 = -2)$.

(3) $\mathbb{P}(1 \oplus t_4 = 1)$.

3.4 Ergodic theory and subgroups

Having defined iteration, we can now talk about the long-term asymptotics of the process. For example, is there a way to associate a *limit* μ_∞ probability distribution to the iterated convolutions μ^{*n} as $n \to \infty$? This limit should give us some information of the long-term asymptotics of the random walk on $\mathbb{Z}_p$ with initial distribution given by μ.

For this purpose, let us define formally "limits" of sequences of probability distributions:

Definition 3.25 (Limits of probability distributions)

Let $\mu_1, \mu_2, \cdots : \mathbb{Z}_p \to [0,1]$ be a sequence of probability distributions. Then they have a **limit** $\mu_\infty : \mathbb{Z}_p \to [0,1]$ if for every $t \in \mathbb{Z}_p$ the limit exists:

$$\lim_{n\to\infty} \mu_n(t) = \mu_\infty(t).$$

The limit μ^∞ is a probability distribution (Exercise!).

We can characterise them using the total variation distance:

Theorem 3.26

Let $\mu_1, \mu_2, \cdots : \mathbb{Z}_p \to [0,1]$ be a sequence of probability distributions. Then they have a limit $\mu_\infty : \mathbb{Z}_p \to [0,1]$ if and only if

$$\lim_{n\to\infty} d(\mu_n, \mu_\infty) = 0.$$

Proof
Exercise. □

Ergodicity of a random walk is a fundamental notion of *chaos*. The basic idea of an ergodic random walk that it will forget the initial state and the limit will be independent of where we began. In the case of non-ergodic random walks the initial state (or distribution) will completely determine the range of the random walk.

Definition 3.27 (Ergodicity)

A probability distribution $\mu : \mathbb{Z}_p \to [0,1]$ is **ergodic** if the limit of the iterated convolutions is the uniform distribution λ, that is,

$$\lim_{n\to\infty} \mu^{*n}(t) = \lambda(t), \quad t \in \mathbb{Z}_p.$$

Recall that λ is the uniform distribution $\lambda(t) = 1/p$ for $t \in \mathbb{Z}_p$.

In other words, recalling the notations from Definition 3.18, if $t_1, t_2 \dots$ is the associated random walk on $\mathbb{Z}_p$ with $t_1, t_2, \dots,$ have distribution μ, then the distributions of

$$X_n = t_1 \oplus t_2 \cdots \oplus t_n \in \mathbb{Z}_p$$

converge in total variation distance to the uniform distribution λ on $\mathbb{Z}_p$. For those more familiar with probability theory, random walks and Markov chains, one can equivalently write ergodicity using the sites the random walk generated by $\mu : \mathbb{Z}_p \to [0,1]$ attains. We will discuss this in more detaill at the end of the section.

Recall from the language of group theory:

Definition 3.28

Let $p \geqslant 2$.

(1) A subset $\Gamma \subset \mathbb{Z}_p$ is a **subgroup** if

$$t, s \in \mathbb{Z}_p \quad \Longrightarrow \quad t \oplus s \in \mathbb{Z}_p.$$

We write then $\Gamma < \mathbb{Z}_p$.

(2) Subgroup $\Gamma < \mathbb{Z}_p$ is **trivial** if $\Gamma = \{0\}$ and proper if $\Gamma \neq \mathbb{Z}_p$.

(3) A subset $A \subset \mathbb{Z}_p$ **generates** a subgroup Γ if

$$\langle A \rangle = \Gamma.$$

(4) A subset $A \subset \mathbb{Z}_p$ is a **coset** of a subgroup Γ if there exists $t \in \mathbb{Z}_p$ such that

$$A = \Gamma \oplus t,$$

where

$$\Gamma \oplus t = \{a \oplus t : a \in \Gamma\}.$$

We will need the following simple lemma that tells us that not a set $A \subset \mathbb{Z}_p$ not being concentrated on proper subgroups implies that taking large enough sumsets

$$A^{\oplus n} := A^{\oplus(n-1)} \oplus A$$

for $n \geqslant 1$ with $A^{*0} = \varnothing$, become the whole group $\mathbb{Z}_p$.

Lemma 3.29 (Non-concentration and sumsets)

Let $A \subset \mathbb{Z}_p$. If A is not contained in a coset of a proper subgroup of $\mathbb{Z}_p$, then there exists $n \in \mathbb{N}$ such that

$$A^{\oplus n} = \mathbb{Z}_p.$$

Proof
(Thanks Borys Kuca for the proof, see also an alternative proof in the book by Tao and Vu [14, Proposition 2.2])

Let us first prove the following property: $(*)$ *If $A, B \subset \mathbb{Z}_p$ are non-empty and $|A \oplus B| = |A| = |B|$, then A, B are cosets of the same subgroup $\Gamma < \mathbb{Z}_p$.*

To prove this property $(*)$, fix some $t \in A$ and some $s \in B$, and define the translates

$$A' = A \ominus t, \quad \text{and} \quad B' = B \ominus s.$$

Then the sumset

$$A' \oplus B' = (A \oplus B) \oplus (t \oplus s)$$

so by the assumption on A and B we have

$$|A' \oplus B'| = |A'| \quad \text{and} \quad |A' \oplus B'| = |B'|.$$

Note that $0 = t \ominus t \in A'$ and $0 = s \ominus s \in B'$ so $A' \subset A' \oplus B'$ and $B' \subset A' \oplus B'$. Therefore we have from the cardinality equality that

$$A' = A' \oplus B' = B'.$$

Thus $A' \oplus A' = A'$. This implies A' is a subgroup. On the other hand, $A = A' \oplus t$ and $B = A' \oplus s$ so A, B are cosets of the same subgroup. Thus the property $(*)$ above is proved.

Now, let us look at the claim of the lemma. If A is not contained in a coset of a proper subgroup, then neither is $A^{\oplus n}$ for any $n \in \mathbb{N}$. Hence applying the property $(*)$ above we obtain

$$|A^{\oplus 2}| = |A \oplus A| > |A|, \quad |A^{\oplus 4}| = |A^{\oplus 2} \oplus A^{\oplus 2}| > |A^{\oplus 2}|, \quad |A^{\oplus 8}| = |A^{\oplus 4} \oplus A^{\oplus 4}| > |A^{\oplus 4}| \dots$$

so we have

$$|A^{\oplus 2^n}| > |A^{\oplus 2^{n-1}}|, \quad n \in \mathbb{N},$$

so the cardinality $|A^{\oplus 2^n}|$ grows as $n \to \infty$. However, we always have $|A^{\oplus 2^n}| \leqslant |\mathbb{Z}_p| = p$ for all $n \in \mathbb{N}$ so that means there exists $n \in \mathbb{N}$ such that $|A^{\oplus 2^n}| = |\mathbb{Z}_p|$, which implies $A^{\oplus n} = \mathbb{Z}_p$. □

Lemma 3.29 gives us the following fundamental result on supports of iterated convolutions and concentration on cosets of subgroups:

Theorem 3.30 (Non-concentration of convolutions on subgroups)

Let $\mu : \mathbb{Z}_p \to [0, 1]$ be a probability distribution. Then the support $\mathrm{spt}(\mu)$ is not contained in a coset of a proper subgroup of $\mathbb{Z}_p$ if and only if there exists $n \in \mathbb{N}$ with

$$\mathrm{spt}(\mu^{*n}) = \mathbb{Z}_p.$$

Proof
$\boxed{\Rightarrow}$ Write $A = \mathrm{spt}(\mu)$. Then Theorem 3.14 gives

$$\mathrm{spt}(\mu^{*n}) = A^{\oplus n}.$$

Since $A = \mathrm{spt}(\mu)$ is not contained in a coset of a proper subgroup of $\mathbb{Z}_p$, we can apply Lemma 3.29 with $A = \mathrm{spt}(\mu)$ to obtain

$$A^{\oplus n} = \mathbb{Z}_p$$

as claimed.

$\boxed{\Leftarrow}$ Suppose on the contrary that $\mathrm{spt}(\mu) \subset \Gamma \oplus s$ for some proper subgroup Γ of $\mathbb{Z}_p$, but there exists $n \in \mathbb{N}$ such that $\mu^{*n}(t) > 0$ for all $t \in \mathbb{Z}_p$. Using Theorem 3.14 we see that the support

$$\mathrm{spt}(\mu^{*n}) = (\Gamma \oplus s)^{\oplus n} = \Gamma^{\oplus n} \oplus (ns).$$

However, since Γ is a subgroup, we know that

$$\Gamma^{\oplus n} \subset \Gamma$$

so we have that $\mathrm{spt}(\mu^{*n})$ is also contained in a coset of a proper subgroup Γ of $\mathbb{Z}_p$. Since Γ is proper, that is, not equal to $\mathbb{Z}_p$, also $\mathrm{spt}(\mu^{*n}) \neq \mathbb{Z}_p$ so we know that there exists $t \in \mathbb{Z}_p$ such that

$$\mu^{*n}(t) = 0.$$

Contradiction. □

Theorem 3.30 relates closely to ergodicity:

Theorem 3.31 (Ergodic theorem)

A probability distribution μ is ergodic if and only if the support $\mathrm{spt}(\mu)$ is not contained in a coset of a proper subgroup of $\mathbb{Z}_p$.

Proof

If μ is ergodic, then $\mu^{*n}(t) \to \lambda(t) = 1/p$ for all $t \in \mathbb{Z}_p$. Hence there exists $n \in \mathbb{N}$ such that $\mu^{*n}(t) > 0$ for all $n \in \mathbb{N}$. Then by Theorem 3.30 the support $\mathrm{spt}(\mu)$ is not contained in a proper subgroup of $\mathbb{Z}_p$.

Thus we just need to check the other direction, where we assume that $\mathrm{spt}(\mu)$ generates $\mathbb{Z}_p$ and it is not contained in a proper subgroup of $\mathbb{Z}_p$. Let us prove the ergodicity of μ:

$$\lim_{k\to\infty} \mu^{*k}(t) = \lambda(t), \quad t \in \mathbb{Z}_p.$$

(1) Define the sequences

$$M_k = \max_{t\in\mathbb{Z}_p} \mu^{*k}(t)$$

and

$$m_k = \min_{t\in\mathbb{Z}_p} \mu^{*k}(t).$$

By Theorem 3.30, since $\mathrm{spt}(\mu)$ is not contained in a proper subgroup of $\mathbb{Z}_p$ we know that there exists $k_0 \in \mathbb{N}$ such that

$$\varepsilon := m_{k_0} \in (0, 1).$$

(2) It is enough to prove that M_k and m_k both converge to a common limit $\ell > 0$. Indeed, since by definition

$$m_k \leqslant \mu^{*k}(t) \leqslant M_k$$

we know that for all $t \in \mathbb{Z}_p$.

$$\lim_{k\to\infty} \mu^{*k}(t) = \ell.$$

The only possibility for such limit is $\ell = 1/p$ since by summing over $t \in \mathbb{Z}_p$ and using the fact that μ^{*k} is a probability distribution, we have

$$1 = \lim_{k\to\infty} \sum_{t\in\mathbb{Z}_p} \mu^{*k}(t) = \sum_{t\in\mathbb{Z}_p} \lim_{k\to\infty} \mu^{*k}(t) = \sum_{t\in\mathbb{Z}_p} \ell = p\ell,$$

which is claim.

(3) Let us now prove that M_k and m_k have a common limit. We know that (M_k) and (m_k) both converge to some limits M_∞ and m_∞ respectively as the sequences M_k and m_k are monotonic: for m_k we have that for all $t \in \mathbb{Z}_p$ we have

$$m_k = \sum_{s\in\mathbb{Z}_p} \mu(t \ominus s) m_k \leqslant \sum_{s\in\mathbb{Z}_p} \mu(t \ominus s)\mu^{*k}(s) = \mu^{*(k+1)}(t)$$

so taking minimum over $t \in \mathbb{Z}_p$ gives

$$m_k \leqslant m_{k+1}$$

and similarly for M_k we have

$$\mu^{*(k+1)}(t) = \sum_{s\in\mathbb{Z}_p} \mu(t \ominus s)\mu^{*k}(s) \leqslant \sum_{s\in\mathbb{Z}_p} \mu(t \ominus s) M_k = M_k$$

so by taking maximum over $t \in \mathbb{Z}_p$ gives

$$M_{k+1} \leqslant M_k.$$

Now we just need to prove that $M_\infty = m_\infty$.

(4) To prove $M_\infty = m_\infty$, it is enough to prove that for all $r \geqslant 0$ we have

$$M_{k_0+r} - m_{k_0+r} \leqslant (1-\varepsilon)(M_r - m_r). \tag{3.2}$$

Indeed, by iterating this inequality k times we obtain

$$M_{kk_0+r} - m_{kk_0+r} \leqslant (1-\varepsilon)^k (M_r - m_r),$$

which converges to 0 as $k \to \infty$ as $0 < 1 - \varepsilon < 1$. Since $M_k \to M_\infty$ and $m_k \to m_\infty$ as $k \to \infty$ the limits along these subsequences will be the same, so $M_\infty = m_\infty$.

(5) Let us now prove the final claim (3.2). We have

$$\begin{aligned}
\mu^{*(k_0+r)}(t) &= \sum_{s\in\mathbb{Z}_p} \mu^{*k_0}(t \ominus s)\mu^{*r}(s) \\
&= \sum_{s\in\mathbb{Z}_p} [\mu^{*k_0}(t \ominus s) - \varepsilon\mu^{*r}(-s)]\mu^{*r}(s) + \varepsilon \sum_{s\in\mathbb{Z}_p} \mu^{*r}(-s)\mu^{*r}(s) \\
&= \sum_{s\in\mathbb{Z}_p} [\mu^{*k_0}(t \ominus s) - \varepsilon\mu^{*r}(-s)]\mu^{*r}(s) + \varepsilon\mu^{*(2r)}(0) \\
&\geqslant \sum_{s\in\mathbb{Z}_p} [\mu^{*k_0}(t \ominus s) - \varepsilon\mu^{*r}(-s)] m_r + \varepsilon\mu^{*(2r)}(0) \\
&= (1-\varepsilon) m_r + \varepsilon\mu^{*(2r)}(0).
\end{aligned}$$

The inequality in the above chain follows from the fact that

$$\mu^{*k_0}(t \ominus s) - \varepsilon\mu^{*r}(-s) \geqslant 0$$

since by definition $\varepsilon = m_{k_0}$ we have $\varepsilon \leqslant \mu^{*k_0}(t \ominus s)$ so

$$\mu^{*k_0}(t \ominus s) - \varepsilon\mu^{*r}(-s) \geqslant \mu^{*k_0}(t \ominus s)(1 - \mu^{*r}(-s))$$

and here

$$\mu^{*k_0}(t \ominus s)(1 - \mu^{*r}(-s)) \geqslant 0.$$

We have proved for all $t \in \mathbb{Z}_p$ the inequality

$$\mu^{*(k_0+r)}(t) \geqslant (1 - \varepsilon)m_r + \varepsilon\mu^{*(2r)}(0).$$

Now taking the minimum over all $t \in \mathbb{Z}_p$ gives

$$m_{k_0+r} \geqslant (1 - \varepsilon)m_r + \varepsilon\mu^{*(2r)}(0).$$

A similar argument (left as an exercise) shows the upper bound

$$M_{k_0+r} \leqslant (1 - \varepsilon)M_r + \varepsilon\mu^{*(2r)}(0).$$

Combining these gives us the claimed inequality (3.2). □

Exercise 3.32

Let μ be a probability distribution on $\mathbb{Z}_p$ and assume that the support

$$\mathrm{spt}(\mu) = \{t \in \mathbb{Z}_p : \mu(t) > 0\}$$

is a proper subgroup of $\mathbb{Z}_p$. What is the limit

$$\lim_{t\to\infty} \mu^{*k}(t)?$$

3.5 Mixing

Having now found out that as long as the support $\mathrm{spt}(\mu)$ is not contained in a proper subgroup of $\mathbb{Z}_p$, then

$$\mu^{*n}(t) \to \lambda(t)$$

at every $t \in \mathbb{Z}_p$. Note that this is equivalent to

$$d(\mu^{*n}, \lambda) \to 0.$$

What is the rate of this? This is very relevant to us if we want to find out the number of card shuffles we would like to do to properly mix the deck, or say, how many moves of the Rubik's cube we would need to perform in order to have a state of the cube that is random enough. For this purpose, practically one could for example require to find the minimal $n \in \mathbb{N}$ such that

$$d(\mu^{*n}, \lambda) \leqslant \frac{1}{100},$$

which implies that all the probabilities $\mu^{*n}(A)$ for every event $A \subset \mathbb{Z}_p$ are very close to the uniform $\lambda(A) = 1/|A|$ up to an error of 1%.

This minimal n for which

$$d(\mu^{*n}, \lambda) \leqslant \frac{1}{100},$$

is called the **mixing time** of the random walk with threshold $\varepsilon = 1/100$.

Definition 3.33 (Mixing time)

Given a threshold $\varepsilon > 0$, we say that the random walk has **mixing time** $n_{\mathrm{mix}}(\varepsilon)$ if for all $n \geqslant n_{\mathrm{mix}}(\varepsilon)$

$$d(\nu * \mu^{*n}, \lambda) < \varepsilon.$$

To find the mixing time $n_{\mathrm{mix}}(\varepsilon)$ it is thus very important to know the quantitative rate at which

$$d(\mu^{*n}, \lambda) \to 0.$$

One usually called the **rate of mixing** of the random walk.

Definition 3.34 (Mixing)

We say that the random walk driven by $\mu : \mathbb{Z}_p \to [0, 1]$ is **mixing with a rate** $\phi(n) \to 0$ as $n \to \infty$ if for all initial distributions ν we have

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

for all $n \in \mathbb{N}$.

Commonly we see that the rate of mixing is exponential, and in the case of $\mathbb{Z}_p$, but to prove this we need harmonic analysis.

Chapter 4

Harmonic analysis

4.1 Introduction

Suppose we have a function $f : [0,1] \to \mathbb{C}$ that has a relatively messy looking graph. One could consider this as a sound signal with various high and low frequencies in it (like a tape of a music record), or a very fractal like function. In order to understand the behaviour of f, one needs to find ways to decompose the function into simpler pieces from which one could read properties of it. This can be useful in signal processing to find which high or low frequency sounds are contributing to the signal.

It was observed by Fourier that using sums of simple trigonometric functions (waves) $x \mapsto \cos(2\pi kx)$, if the function f has enough regularity (such as differentiability), it can be expressed as a sum of cosine waves:

$$f(x) = \sum_{k \in \mathbb{Z}} a_k \cos(2\pi kx)$$

for some coefficients a_k that represent the *amplitude* (or height) of the wave, and the integer k corresponds to a *frequency* (or number of oscillations) of the wave.

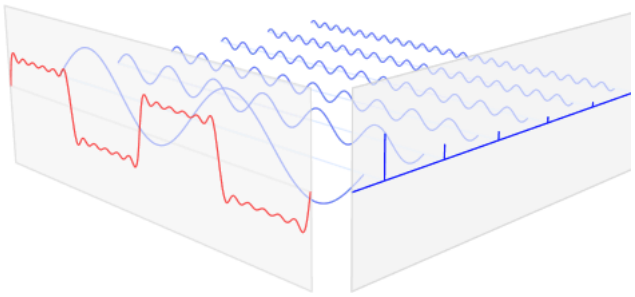

Figure 4.1: The core idea behind Fourier analysis: we would like to represent a complicated function/signal $f : [0,1] \to \mathbb{C}$ (*e.g. the red saw-tooth function*) as a sum of simpler cosine functions $x \mapsto a_k \cos(2\pi kx)$, $k \in \mathbb{Z}$ (*the blue waves*) with the coefficients ("amplitudes") a_k given by the Fourier transform of the function $a_k = \widehat{f}(k)$. The values of a_k thus tell how "high" the oscillations of $x \mapsto a_k \cos(2\pi kx)$ become and k tells the frequency of the waves $x \mapsto a_k \cos(2\pi kx)$. Picture copyright CC0.

The amplitudes a_k are in most cases given by the *cosine transform* $\widehat{f}^{\cos}(k)$ of the function

μ at a frequency k, which can be formally defined using integration:

$$a_k = \widehat{f}^{\cos}(k) = \int_0^1 f(x)\cos(-2\pi kx)\,dx.$$

This number represents a kind of expected value of the wave. If it is close to 0, then one expects not to have many oscillations in high frequencies.

However, due to theoretical advantages (and also connections to quantum mechanics, which we will not explore in this course), it is more beneficial to use complex valued waves

$$e^{2\pi kix} = \cos(2\pi kx) + i\sin(2\pi kx), \quad x \in [0,1],$$

when defining the waves. The advantage of this is that we have many useful theoretical formula at our disposal (like convolution theorem and Plancherel theorem below) and one can still recover similar intuition. In this case, the *Fourier transform* of $f : [0,1] \to \mathbb{C}$ at frequency $k \in \mathbb{Z}$ is defined by

$$\widehat{f}(k) = \int_0^1 f(x)e^{-2\pi kix}\,dx$$

and one looks for a representation of f as a *Fourier series*:

$$f(x) = \sum_{k\in\mathbb{Z}} \widehat{f}(k)e^{2\pi kix}, \quad x \in [0,1].$$

In this course, we will not go into this "continuous" side of Fourier analysis, but more consider the case for functions $f : \mathbb{Z}_p \to \mathbb{C}$ instead. However, we emphasise that many of the ideas presented in the simpler case $\mathbb{Z}_p$ carry to the continuous setting with sums replaced by integrals. The proofs in the case $\mathbb{Z}_p$ are just much simpler and do not require to take into account the intricacies of real numbers.

4.2 Fourier transform in $\mathbb{Z}_p$

In the case of the group $\mathbb{Z}_p$ where we study our random walks, we can also study similar representations of functions $f : \mathbb{Z}_p \to \mathbb{R}$ using waves. In literature such Fourier transforms are also known as *Discrete Fourier Transforms* (DFT).

Definition 4.1 (Fourier transform in $\mathbb{Z}_p$)

The **Fourier transform** of $f : \mathbb{Z}_p \to \mathbb{C}$ in the frequency $k \in \mathbb{Z}_p$ is given by

$$\widehat{f}(k) = \sum_{t=0}^{p-1} f(t) e^{-2\pi i k t/p}. \tag{4.1}$$

The functions $t \mapsto e^{-2\pi i k t/p}, t \in \mathbb{Z}_p$ for a given frequency $k \in \mathbb{Z}_p$ are called **characters** or **stationary waves**, and they have a notation: $\chi_k : \mathbb{Z}_p \to \mathbb{C}$, defined by

$$\chi_k(t) := e^{-2\pi i k t/p}.$$

Thus the Fourier transform of $f : \mathbb{Z}_p \to \mathbb{C}$ can be written as

$$\widehat{f}(k) = \sum_{t=0}^{p-1} f(t)\chi_k(t).$$

In this chapter we will consider mostly functions $f : \mathbb{Z}_p \to \mathbb{C}$ but we mostly apply the Fourier transforms to probability distributions $\mu : \mathbb{Z}_p \to [0, 1]$, which are a special case of this theory.

An useful way to think Fourier transforms of, say, a probability distribution $\mu : \mathbb{Z}_p \to [0, 1]$ is that each we choose each complex number $\chi_k(t)$ with probability $\mu(t)$ and sum over them. For computing such resulting exponential sums, the exponential sum formula will be useful:

Theorem 4.2 (Exponential sum formula)

As long as $\theta \neq 0$, we have

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

Proof
Exercise. □

To demonstrate the use of Theorem 4.2 we can use it to directly compute the Fourier transform of the uniform distribution:

Example 4.3

For the uniform distribution $\lambda(t) = 1/p$, $t \in \mathbb{Z}_p$, we have for $k \neq 0$ by the geometric sum formula with $\theta = -2\pi k/p \neq 0$ that

$$\widehat{\lambda}(k) = \sum_{t=0}^{p-1} \frac{1}{p} e^{-2\pi i k t/p} = \frac{1}{p} \frac{1 - e^{ip\theta}}{1 - e^{i\theta}} = \frac{1}{p} \frac{1 - e^{-2\pi k i}}{1 - e^{-2\pi i k/p}} = 0$$

and for $k = 0$ we have

$$\widehat{\lambda}(0) = \sum_{t=0}^{p-1} \frac{1}{p} e^{-2\pi i 0 t/p} = \sum_{t=0}^{p-1} \frac{1}{p} = 1$$

For the singular distribution a complete opposite happens: the Fourier transform has constant modulus 1 everywhere!

Example 4.4

For the singular distribution δ_s at $s \in \mathbb{Z}_p$, we have

$$\widehat{\delta_s}(k) = e^{-2\pi i k s/p}$$

for all $k \in \mathbb{Z}_p$. Note that in particular $|\widehat{\delta_s}(k)| = 1$ for all $k \in \mathbb{Z}_p$ and for $s = 0$ we have

$$\widehat{\delta_0}(k) = 1.$$

Suppose now we have a general probability distribution $\mu : \mathbb{Z}_p \to [0, 1]$. What does $\widehat{\mu}(k)$ tell us about μ? We see that for the uniform distribution $\widehat{\lambda}$ being 0 at every $k \neq 0$ and for the singular distribution $\widehat{\delta_0}$ being 1 shows that the Fourier coefficients $\widehat{\mu}(k)$ being "large" for most $k \in \mathbb{Z}_p$ should mean μ is close to being singular and $\widehat{\mu}(k)$ being "small" for most $k \in \mathbb{Z}_p$ should mean μ is close to being uniform. Let us look at the case of the pass the broccoli distribution.

Example 4.5 (Fourier transform of the pass the broccoli distribution)

Recall the initial distribution $\mu : \mathbb{Z}_p \to [0, 1]$ defining the passing the broccoli process was given by $\mu(1) = \mu(-1) = \frac{1}{2}$ and $\mu(t) = 0$ otherwise. The Fourier transform of μ is then given by

$$\widehat{\mu}(k) = \sum_{t=0}^{p-1} \mu(t) e^{-2\pi i k t/p} = \frac{1}{2} e^{-2\pi i k/p} + \frac{1}{2} e^{2\pi i k/p} = \cos(2\pi k/p).$$

Thus depending on k, we see that $\widehat{\mu}(k)$ can attain large and small values, which means it is neither uniform or singular.

Fourier coefficients allow us to represent any function $f : \mathbb{Z}_p \to \mathbb{C}$ as trigonometric series with coefficients given by $\widehat{f}(k)$ as follows:

Theorem 4.6 ("Fourier series")

Any function $f : \mathbb{Z}_p \to \mathbb{C}$ has the following **Fourier expansion**:

$$f(t) = \frac{1}{p} \sum_{k=0}^{p-1} \widehat{f}(k) e^{2\pi i k t/p}, \quad t \in \mathbb{Z}_p.$$

Proof
By definition of Fourier coefficients

$$\begin{aligned}
\sum_{k=0}^{p-1} \widehat{f}(k) e^{2\pi i k t} &= \sum_{k=0}^{p-1} \sum_{s=0}^{p-1} f(s) e^{-2\pi i k s/p} e^{2\pi i k t/p} \\
&= \sum_{s=0}^{p-1} \sum_{k=0}^{p-1} f(s) e^{-2\pi i k s/p} e^{2\pi i k t/p} \\
&= \sum_{s=0}^{p-1} f(s) \sum_{k=0}^{p-1} e^{-2\pi i k s/p} e^{2\pi i k t/p} \\
&= \sum_{s=0}^{p-1} f(s) \sum_{k=0}^{p-1} e^{-2\pi i k (s-t)/p}
\end{aligned}$$

Now we see that

$$\sum_{k=0}^{p-1} e^{-2\pi i k (t-s)/p} = \begin{cases} p, & t = s; \\ 0, & t \neq s. \end{cases} \tag{4.2}$$

Let us argue (4.4) this in the following two cases.

$\boxed{t = s}$ We are summing 1 in total p times, so the sum is p.

$\boxed{t \neq s}$ Write $\theta = -2\pi(t-s)/p \neq 0$. Then by the geometric sum formula

$$\sum_{k=0}^{p-1} e^{k\theta i} = \frac{1 - e^{p\theta i}}{1 - e^{\theta i}}.$$

However, since $t - s$ is an integer we have that

$$e^{p\theta i} = e^{-2\pi(t-s)i} = 1,$$

as the value of e^{ix} at any integer multiple of 2π is 1. Thus

$$\frac{1 - e^{p\theta i}}{1 - e^{\theta i}} = 0$$

as claimed in (4.4).

Continuing now first computation, we see that

$$\sum_{s=0}^{p-1} f(s) \sum_{k=0}^{p-1} e^{-2\pi ik(s-t)/p} = f(t)p,$$

which gives the claim. □

Definition 4.7 (Inverse Fourier transform)

The formula for the Fourier series is also called the **Inverse Fourier transform**, denoted for $g : \mathbb{Z}_p \to \mathbb{C}$ by

$$\check{g}(t) := \frac{1}{p} \sum_{k=0}^{p-1} g(k) e^{2\pi ikt/p}, \quad t \in \mathbb{Z}_p.$$

Notice that by definition

$$\check{g}(t) = \frac{1}{p} \widehat{g}(-t), \quad t \in \mathbb{Z}_p.$$

Inverse Fourier transform returns from Fourier transform back the function:

Theorem 4.8

If $f : \mathbb{Z}_p \to \mathbb{C}$, then

$$\check{\widehat{f}} = f.$$

Proof
This is precisely Theorem 4.6. □

4.3 L^2 theory

One of the fundamental properties of Fourier transform is that it forms an isometry with respect to the so called "L^2 norm" on the space of functions $f : \mathbb{Z}_p \to \mathbb{C}$. This property helps us to transfer questions for probability distributions $\mu : \mathbb{Z}_p \to [0, 1]$ to their Fourier transforms $\widehat{\mu}(k)$, prove something for them, and then transfer this information back to the probability distribution μ.

Definition 4.9 (Inner product and L^2 norm)

(1) The **inner product** between two functions $f, g : \mathbb{Z}_p \to \mathbb{C}$ is by

$$\langle f, g \rangle = \sum_{t=0}^{p-1} f(t)\overline{g(t)},$$

where $\overline{z}$ is the complex conjugate of $z \in \mathbb{C}$.

(2) The L^2 **norm** of $f : \mathbb{Z}_p \to \mathbb{C}$ is given by

$$\|f\|_2 := \sqrt{\langle f, f \rangle} = \sqrt{\sum_{t=0}^{p-1} |f(t)|^2}.$$

Recall that for $k \in \mathbb{Z}_p$ the **character** $\chi_k : \mathbb{Z}_p \to \mathbb{C}$ was defined by

$$\chi_k(x) := e^{-2\pi i x k/p}, \quad x \in \mathbb{Z}_p$$

Recall that by the Fourier series representation we can write every $f : \mathbb{Z}_p \to \mathbb{C}$ as a sum of these characters times the Fourier coefficients. A bit more is true: these functions form an orthonormal basis:

Exercise 4.10 (Orthonormality of the characters)

Show that the characters χ_k are **orthonormal** (with respect to the inner product above) to each other, that is,

$$\langle \chi_k, \chi_\ell \rangle = \begin{cases} 0, & k \neq \ell; \\ 1, & k = \ell. \end{cases}$$

A fundamental inequality we will often need in the analysis is the Cauchy-Schwartz inequality that links the inner product to the L^2 norms:

Theorem 4.11 (Cauchy-Schwartz inequality)

Let $f, g : \mathbb{Z}_p \to \mathbb{C}$ be any functions. Then

$$|\langle f, g\rangle| \leqslant \|f\|_2 \|g\|_2.$$

Proof

There are many ways to prove this. We will give just one example proof here.

Firstly we may assume both $\|f\|_2 > 0$ and $\|g\|_2 > 0$. If one of them is 0, say, $\|f\|_2 = 0$, then $f(t) = 0$ for all $t \in \mathbb{Z}_p$ so also $|\langle f, g\rangle| = 0$ no matter what g is.

Let $a, b \geqslant 0$. Then $0 \leqslant (a-b)^2 = a^2 - 2ab + b^2$ so

$$ab \leqslant \frac{a^2}{2} + \frac{b^2}{2}. \tag{4.3}$$

Fix $t \in \mathbb{Z}_p$. Apply (4.3) to the non-negative real numbers

$$a = \frac{|f(t)|}{\|f\|_2} \quad \text{and} \quad b = \frac{|g(t)|}{\|g\|_2}$$

to obtain

$$\frac{|f(t)|}{\|f\|_2}\frac{|g(t)|}{\|g\|_2} \leqslant \frac{|f(t)|^2}{2\|f\|_2^2} + \frac{|g(t)|^2}{2\|g\|_2^2}$$

Sum both sides over $t \in \mathbb{Z}_p$ so we obtain

$$\sum_{t\in\mathbb{Z}_p} \frac{|f(t)|}{\|f\|_2}\frac{|g(t)|}{\|g\|_2} \leqslant \sum_{t\in\mathbb{Z}_p} \frac{|f(t)|^2}{2\|f\|_2^2} + \sum_{t\in\mathbb{Z}_p} \frac{|g(t)|^2}{2\|g\|_2^2} = 1.$$

Multiply both sides by $\|f\|_2\|g\|_2$ to obtain

$$\sum_{t\in\mathbb{Z}_p} |f(t)||g(t)| \leqslant \|f\|_2\|g\|_2.$$

Finally, by the triangle inequality (as $|\overline{g(t)}| = |g(t)|$) we have

$$|\langle f, g\rangle| = \Big|\sum_{t\in\mathbb{Z}_p} f(t)\overline{g(t)}\Big| \leqslant \sum_{t\in\mathbb{Z}_p} |f(t)||g(t)|,$$

so the proof is complete. □

There is also an analogue of this for higher order moments, the so called L^p norms:

Definition 4.12 (L^p norms)

Let $f : \mathbb{Z}_p \to \mathbb{C}$. For $1 < p < 2$ define the L^p norm of f by

$$\|f\|_p = \Big(\sum_{t\in\mathbb{Z}_p} |f(t)|^p\Big)^{1/p}.$$

Theorem 4.13 (Hölder's inequality)

Let $f, g : \mathbb{Z}_p \to \mathbb{C}$ be any functions. Suppose $1 < p, q < \infty$ satisfy the relation $1/p + 1/q = 1$. Then

$$|\langle f, g\rangle| \leqslant \|f\|_p \|g\|_q.$$

Proof

Exercise. *Hint: Use Young's inequality for products: if $a, b \geqslant 0$ and $1/p + 1/q = 1$, then*

$$ab \leqslant a^p/p + b^q/q.$$

□

A key property for Fourier analysis is the so called Parseval's identity, which says that, up to a normalisation by $p^{-1/2}$, the Fourier transform operator

$$f \mapsto \widehat{f}$$

is an isometry.

Theorem 4.14 (Plancherel's Theorem / Parseval's Identity)

If $f, g : \mathbb{Z}_p \to \mathbb{C}$, then

$$\langle f, g\rangle = \frac{1}{p}\langle \widehat{f}, \widehat{g}\rangle.$$

In particular, the L^2 norm

$$\|f\|_2 = \frac{1}{\sqrt{p}}\|\widehat{f}\|_2.$$

Proof

By the definition of the inner product and Fourier transform, we have, after changing the order of summation that

$$\begin{aligned}
\langle \widehat{f}, \widehat{g}\rangle &= \sum_{k\in\mathbb{Z}_p} \widehat{f}(k)\overline{\widehat{g}(k)} \\
&= \sum_{k\in\mathbb{Z}_p} \sum_{t=0}^{p-1} f(t)e^{-2\pi ikt/p} \overline{\sum_{s=0}^{p-1} g(s)e^{-2\pi iks/p}} \\
&= \sum_{k\in\mathbb{Z}_p} \sum_{t=0}^{p-1} f(t)e^{-2\pi ikt/p} \sum_{s=0}^{p-1} \overline{g(s)}e^{2\pi iks/p} \\
&= \sum_{t=0}^{p-1} f(t) \sum_{s=0}^{p-1} \overline{g(s)} \sum_{k\in\mathbb{Z}_p} e^{-2\pi ik(t-s)/p}
\end{aligned}$$

Now we see that

$$\sum_{k\in\mathbb{Z}_p} e^{-2\pi ik(t-s)/p} = \sum_{k=0}^{p-1} e^{-2\pi ik(t-s)/p} = \begin{cases} p, & t = s; \\ 0, & t \neq s. \end{cases} \tag{4.4}$$

Let us argue (4.4) this in the following two cases.

$\boxed{t = s}$ We are summing 1 in total p times, so the sum is p.

$\boxed{t \neq s}$ Write $\theta = -2\pi(t-s)/p \neq 0$. Then by the geometric sum formula

$$\sum_{k=0}^{p-1} e^{k\theta i} = \frac{1 - e^{p\theta i}}{1 - e^{\theta i}}.$$

However, since $t - s$ is an integer we have that

$$e^{p\theta i} = e^{-2\pi(t-s)i} = 1,$$

as the value of e^{ix} at any integer multiple of 2π is 1. Thus

$$\frac{1 - e^{p\theta i}}{1 - e^{\theta i}} = 0$$

as claimed in (4.4).

Using (4.4) we see that for any $t \in \mathbb{Z}_p$ we have

$$\sum_{s=0}^{p-1} \overline{g(s)} \sum_{k\in\mathbb{Z}_p} e^{-2\pi ik(t-s)/p} = p\overline{g(t)}.$$

Hence

$$\sum_{t=0}^{p-1} f(t) \sum_{s=0}^{p-1} \overline{g(s)} \sum_{k\in\mathbb{Z}_p} e^{-2\pi ik(t-s)/p} = p \sum_{t=0}^{p-1} f(t)\overline{g(t)} = p\langle f, g\rangle,$$

which gives the claim. □

4.4 Convolution Theorem

The Convolution Theorem is a very basic identity but it is very powerful and enables us to translate convolutions of probability distributions (which describes dynamics of a random walk) into products of their Fourier transforms. Recall that the convolution of two functions $f, g : \mathbb{Z}_p \to \mathbb{C}$ is defined by

$$f * g(t) = \sum_{s \in \mathbb{Z}_p} f(t \ominus s) g(s).$$

Theorem 4.15 (Convolution Theorem)

If $f, g : \mathbb{Z}_p \to \mathbb{C}$, then

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

Proof
Firstly, we have the following invariance for summations: for every $h : \mathbb{Z}_p \to \mathbb{C}$ and $s \in \mathbb{Z}_p$ we have

$$\sum_{t \in \mathbb{Z}_p} h(t) = \sum_{t \in \mathbb{Z}_p} h(t \ominus s). \tag{4.5}$$

This is just a reparametrisation: the map $t \mapsto t \ominus s$ is a bijection $\mathbb{Z}_p \to \mathbb{Z}_p$ so we will count each value in both sums in (6.1) exactly once.

Fix now $k \in \mathbb{Z}_p$. Then after changing the order of summation and using $t = (t \ominus s) \oplus s$, we have

$$\begin{aligned}
\widehat{f * g}(k) &= \sum_{t \in \mathbb{Z}_p} f * g(t) e^{-2\pi i k t/p} \\
&= \sum_{t \in \mathbb{Z}_p} \Big(\sum_{s \in \mathbb{Z}_p} f(t \ominus s) g(s) \Big) e^{-2\pi i k t/p} \\
&= \sum_{t \in \mathbb{Z}_p} \sum_{s \in \mathbb{Z}_p} f(t \ominus s) g(s) e^{-2\pi i k t/p} \\
&= \sum_{s \in \mathbb{Z}_p} \sum_{t \in \mathbb{Z}_p} f(t \ominus s) g(s) e^{-2\pi i k t/p} \\
&= \sum_{s \in \mathbb{Z}_p} \sum_{t \in \mathbb{Z}_p} f(t \ominus s) g(s) e^{-2\pi i k (t \ominus s)/p} e^{-2\pi i k s/p} \\
&= \sum_{s \in \mathbb{Z}_p} g(s) e^{-2\pi i k s/p} \sum_{t \in \mathbb{Z}_p} f(t \ominus s) e^{-2\pi i k (t \ominus s)/p} \\
&= \sum_{s \in \mathbb{Z}_p} g(s) e^{-2\pi i k s/p} \sum_{t \in \mathbb{Z}_p} f(t) e^{-2\pi i k t/p} \\
&= \widehat{f}(k) \widehat{g}(k).
\end{aligned}$$

In the second last equality we applied (6.1) for $h(t) = f(t) e^{-2\pi i k t/p}$, $t \in \mathbb{Z}_p$. □

Example 4.16

Let $\mu(1) = \mu(-1) = 1/2$ and $\mu(t) = 0$ for other $t \in \mathbb{Z}_p$. That is, μ is the driving distribution for the pass the broccoli process. Recall that

$$\widehat{\mu}(k) = \cos(2\pi k/p).$$

Therefore by the convolution theorem

$$\widehat{\mu * \mu}(k) = \cos(2\pi k/p)^2.$$

4.5 Heisenberg Uncertainty Principle in $\mathbb{Z}_p$

We want to finish the harmonic analysis section by a fundamental consequence of the Plancherel's theorem that relates functions $f : \mathbb{Z}_p \to \mathbb{C}$ to their Fourier transforms $\widehat{f} : \mathbb{Z}_p \to \mathbb{C}$. Using a bit of terminology from quantum mechanics, if we take $f : \mathbb{Z}_p \to \mathbb{C}$ with $\|f\|_2 = 1$ (a "wave function"), then the function $\mu_f : \mathbb{Z}_p \to [0, 1]$ defined by

$$\mu_f(t) := |f(t)|^2, \quad t \in \mathbb{Z}_p,$$

is by definition a probability distribution: $\sum_{t\in\mathbb{Z}_p} \mu_f(t) = \|f\|_2^2 = 1$. In quantum physics one thinks then the values $\mu_f(t)$ measuring the probability of finding a particle a **position** $t \in \mathbb{Z}_p$ given the quantum state f of a particle. Now by Plancherel's theorem, the Fourier transform $\widehat{f}$ also satisfies $\|\widehat{f}\|_2 = \|f\|_2 = 1$, so

$$\mu_{\widehat{f}}(k) := |\widehat{f}(k)|^2, \quad k \in \mathbb{Z}_p,$$

is also a probability distribution $\mathbb{Z}_p \to \mathbb{C}$ where $\mu_f(k)$ measures the probability of the particle at quantum state f having **velocity** $k \in \mathbb{C}$.

Heisenberg Uncertainty Principle fundamentally limits what information we can say about the position and velocity of a particle simultanously: essentially, it is impossible to measure the position μ_f and $\mu_{\widehat{f}}$ with equal accuracy. One can write this formally using entropy in the following:

Theorem 4.17 (Entropic Heisenberg Uncertainty Principle in $\mathbb{Z}_p$)

There exists $C_p > 0$ such that if $f : \mathbb{Z}_p \to \mathbb{C}$ and $\|f\|_2 = 1$, then

$$H(\mu_f) + H(\mu_{\widehat{f}}) \geqslant C_p.$$

In other words, it is impossible for entropies of μ_f and $\mu_{\widehat{f}}$ be small simultaneously, which would mean, using the heuristics of entropy being the expected amount of information, that it is impossible to have very accurate information on the μ_f random position $t \in \mathbb{Z}_p$ and the $\mu_{\widehat{f}}$ random velocity $k \in \mathbb{Z}_p$.

We do not prove entropic Heisenberg Uncertainty Principle here, and we leave it as an exercise (find out e.g. what is the optimal C_p?). Instead, we prove a weaker version, which is slightly easier to prove, and that involves only the supports of μ_f and $\mu_{\widehat{f}}$ or equivalently the supports of f and $\widehat{f}$. Recall that the support of $f : \mathbb{Z}_p \to \mathbb{C}$ is given by $\operatorname{spt}(f) = \{t \in \mathbb{Z}_p : f(t) \neq 0\}$:

Theorem 4.18 (Heisenberg Uncertainty Principle in $\mathbb{Z}_p$ for supports)

Let $f : \mathbb{Z}_p \to \mathbb{C}$ with $f \neq 0$. Then

$$|\operatorname{spt}(f)||\operatorname{spt}(\widehat{f})| \geqslant p.$$

Thus if $\|f\|_2 = 1$, then

$$|\operatorname{spt}(\mu_f)||\operatorname{spt}(\mu_{\widehat{f}})| \geqslant p.$$

Proof
Define the indicator function of $\operatorname{spt}(\widehat{f})$ by:

$$\mathbf{1}_{\operatorname{spt}(\widehat{f})}(k) := \begin{cases} 1, & k \in \operatorname{spt}(\widehat{f}) \\ 0, & k \notin \operatorname{spt}(\widehat{f}). \end{cases}$$

Fix $t \in \mathbb{Z}_p$. Then by Fourier inversion:

$$f(t) = \frac{1}{p} \sum_{k \in \mathbb{Z}_p} \widehat{f}(k) e^{2\pi i k t/p} = \frac{1}{p} \sum_{k \in \mathbb{Z}_p} \mathbf{1}_{\operatorname{spt}(\widehat{f})}(k) \widehat{f}(k) e^{2\pi i k t/p}.$$

Applying Cauchy-Schwartz inequality to the functions

$$k \mapsto \mathbf{1}_{\operatorname{spt}(\widehat{f})}(k) \quad \text{and} \quad k \mapsto \widehat{f}(k) e^{2\pi i k t/p},$$

we obtain

$$\begin{aligned} |f(t)|^2 &= \frac{1}{p^2} \Big| \sum_{k \in \mathbb{Z}_p} \mathbf{1}_{\operatorname{spt}(\widehat{f})}(k) \widehat{f}(k) e^{2\pi i k t/p} \Big|^2 \\ &\leqslant_{CS} \frac{1}{p^2} \sum_{k \in \mathbb{Z}_p} |\mathbf{1}_{\operatorname{spt}(\widehat{f})}(k)|^2 \sum_{k \in \mathbb{Z}_p} |\widehat{f}(k) e^{2\pi i k t/p}|^2 \\ &= \frac{1}{p^2} |\operatorname{spt}(\widehat{f})| \sum_{k \in \mathbb{Z}_p} |\widehat{f}(k)|^2 \\ &= \frac{1}{p^2} |\operatorname{spt}(\widehat{f})| \|\widehat{f}\|_2^2 \end{aligned}$$

so taking max over all $t \in \mathbb{Z}_p$ we obtain

$$\|f\|_\infty^2 \leqslant \frac{1}{p^2} |\operatorname{spt}(\widehat{f})| \|\widehat{f}\|_2^2.$$

By Plancherel theorem

$$\|\widehat{f}\|_2^2 = p\|f\|_2^2$$

and here the L^2 norm is, after summing over the support, bounded by

$$\|f\|_2^2 = \sum_{s\in\mathbb{Z}_p} |f(s)|^2 = \sum_{s\in\mathrm{spt}(f)} |f(s)|^2 \leqslant |\,\mathrm{spt}(f)|\|f\|_\infty^2$$

so we have

$$\|f\|_\infty^2 \leqslant \frac{1}{p}|\,\mathrm{spt}(\widehat{f})||\,\mathrm{spt}(f)|\|f\|_\infty^2$$

Thus as the support of f is non-empty, we know $\|f\|_\infty^2 = \max\{|f(t)| : t \in \mathbb{Z}_p\} > 0$ so we can divide by it and obtain

$$|\,\mathrm{spt}(f)||\,\mathrm{spt}(\widehat{f})| \geqslant p$$

as claimed.

The case of μ_f and $\mu_{\widehat{f}}$ follows since $\mathrm{spt}(f) = \mathrm{spt}(\mu_f)$ and $\mathrm{spt}(\widehat{f}) = \mathrm{spt}(\mu_{\widehat{f}})$. □

Heisenberg Uncertainty Principle also has an inverse that relates to the algebraic structure of the wave function f:

Theorem 4.19 (Inverse the Heisenberg Uncertainty Principle in $\mathbb{Z}_p$ for supports)

If $f : \mathbb{Z}_p \to \mathbb{C}$ with $0 \in \mathrm{spt}(f)$ satisfies the equality:

$$|\mathrm{spt}(f)||\mathrm{spt}(\widehat{f})| = p,$$

then $\mathrm{spt}(f)$ is a subgroup of $\mathbb{Z}_p$.

Proof
Left as an exercise. □

Note that for prime p the only subgroups of $\mathbb{Z}_p$ are $\{0\}$ or $\mathbb{Z}_p$. Here Heisenberg Uncertainty Principle for supports can be improved from products to sums by the work of T. Tao [13], who proved:

Theorem 4.20 (Improved Heisenberg Uncertainty Principle in $\mathbb{Z}_p$ for supports)

Suppose p is a prime number. Let $f : \mathbb{Z}_p \to \mathbb{C}$ with $f \neq 0$. Then

$$|\,\mathrm{spt}(f)| + |\,\mathrm{spt}(\widehat{f})| \geqslant p + 1.$$

Moreover, this inequality is sharp.

This result is useful as it can be used to give a short proof of the fundamental Cauchy-Davenport inequality from additive combinatorics, we refer to [13] for details.

Chapter 5

Finding the mixing time

5.1 Distance to uniform and Fourier transform

This section is the culmination of all the ideas presented in the previous sections. Here we will present the *Upper Bound Lemma* proved by Diaconis and Shashahani, which in the case of the group $\mathbb{Z}_p$ can be done simply with Fourier transforms. It allows one to estimate quantitatively the rate of convergence of the convolution μ^{*n} to the uniform distribution in terms of the Fourier coefficients of μ, which may be easier to compute than the actual weights.

Theorem 5.1 ("Upper Bound Lemma")

Let $\mu : \mathbb{Z}_p \to [0,1]$ be a probability distribution. Then for all $n \in \mathbb{N}$ we have

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

Proof
By Theorem 2.26 connecting total variation distance to L^1 distance, we have

$$4d(\mu^{*n}, \lambda)^2 = \Big(\sum_{t=0}^{p-1} |\mu^{*n}(t) - \lambda(t)|\Big)^2.$$

Since $\lambda(t) = 1/p$ for all $t \in \mathbb{Z}_p$, we have

$$\Big(\sum_{t=0}^{p-1} |\mu^{*n}(t) - \lambda(t)|\Big)^2 = p^2\Big(\sum_{t=0}^{p-1} \lambda(t)|\mu^{*n}(t) - \lambda(t)|\Big)^2.$$

Using the definition of the inner product for the functions

$$f(t) := \lambda(t), \quad \text{and} \quad g(t) := |\mu^{*n}(t) - \lambda(t)|, \quad t \in \mathbb{Z}_p,$$

and Cauchy-Schwartz Inequality (Theorem 4.13) we obtain

$$\Big(\sum_{t=0}^{p-1} \lambda(t)|\mu^{*n}(t) - \lambda(t)|\Big)^2 = |\langle f, g\rangle|^2 \leqslant \|f\|_2^2 \|g\|_2^2.$$

The L^2 norms here are

$$\|f\|_2^2 = \sum_{t\in\mathbb{Z}_p} \lambda(t)^2 = \sum_{t\in\mathbb{Z}_p} p^{-2} = p^{-1}$$

and by definition of g:

$$\|g\|_2^2 = \sum_{t\in\mathbb{Z}_p} |\mu^{*n}(t) - \lambda(t)|^2.$$

Hence we have proved

$$4d(\mu^{*n}, \lambda)^2 \leqslant p \sum_{t\in\mathbb{Z}_p} |\mu^{*n}(t) - \lambda(t)|^2 = p\|\mu^{*n} - \lambda\|_2^2$$

By Plancherel's Theorem (Theorem 4.14), we have that

$$p\|\mu^{*n} - \lambda\|_2^2 = \|\widehat{\mu^{*n} - \lambda}\|_2^2 = \|\widehat{\mu^{*n}} - \widehat{\lambda}\|_2^2 = \sum_{k=0}^{p-1} |\widehat{\mu^{*n}}(k) - \widehat{\lambda}(k)|^2.$$

Recall that we already established that

$$\widehat{\lambda}(k) = \begin{cases} 1, & k = 0; \\ 0, & k \neq 0. \end{cases}$$

On the other hand, as μ^{*n} is a probability distribution, the Fourier transform

$$\widehat{\mu^{*n}}(0) = \sum_{t\in\mathbb{Z}_p} \mu^{*n}(t) = 1.$$

Hence the difference

$$\widehat{\mu^{*n}}(k) - \widehat{\lambda}(k) = \begin{cases} 0, & k = 0; \\ \widehat{\mu^{*n}}(k), & k \neq 0. \end{cases}$$

Moreover, by the Convolution Theorem (Theorem 4.15) we have

$$\widehat{\mu^{*n}}(k) = \widehat{\mu}(k)^n.$$

Thus

$$\sum_{k=0}^{p-1} |\widehat{\mu^{*n}}(k) - \widehat{\lambda}(k)|^2 = \sum_{k\in\mathbb{Z}_p\setminus\{0\}} |\widehat{\mu}(k)|^{2n}.$$

Dividing by 4 and taking square roots from both sides gives the claim. □

There is also a converse to the the upper bound lemma, which we will leave as an exercise:

Theorem 5.2 ("Lower Bound Lemma")

Let $\mu : \mathbb{Z}_p \to [0, 1]$ be a probability distribution. Then for all $n \in \mathbb{N}$ we have

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

Proof
Exercise. □

There is also a generalisation of the Upper Bound Lemma for general sequence of measures:

Theorem 5.3

Let $\mu_1, \mu_2, \cdots : \mathbb{Z}_p \to [0,1]$ be a sequence of probability distributions. Then for all $n \in \mathbb{N}$ we have

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

Proof
Exercise. □

The Upper Bound Lemma is a very useful lemma to also answer to our questions on the growth of entropy for the random walks on $\mathbb{Z}_p$. It implies the following growth bound:

Theorem 5.4 (Entropy growth under convolutions)

Let $\mu : \mathbb{Z}_p \to [0,1]$ be a probability distribution. Then

$$H(\mu^{*n}) \geqslant \log p - (\log p + 1)\sqrt{\sum_{k \in \mathbb{Z}_p \setminus \{0\}} |\widehat{\mu}(k)|^{2n}}.$$

Proof
Pinsker's inequality gives

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

so the claim follows from the Upper Bound Lemma as $H(\lambda) = \log p$. □

5.2 Spectral gap, ergodicity and mixing

Having the Upper Bound Lemma at our disposal, we can now apply it to prove ergodicity and mixing (of exponential rate) of μ assuming the Fourier coefficients $\widehat{\mu}(k)$, when $k \neq 0$, are all strictly less than 1 in modulus. In this case μ is said to have a **spectral gap**:

Definition 5.5 (Spectral gap)

A probability distribution μ on $\mathbb{Z}_p$ has a **spectral gap** if

$$|\widehat{\mu}(k)| < 1$$

for all $k \in \mathbb{Z}_p \backslash \{0\}$.

Notice that for $k = 0$ we always have

$$\widehat{\mu}(0) = \sum_{t \in \mathbb{Z}_p} \mu(t) e^{-2\pi i 0 t/p} = \sum_{t \in \mathbb{Z}_p} \mu(t) = 1$$

as μ is a probability distribution, and for other k, we have by the triangle inequality

$$|\widehat{\mu}(k)| \leqslant \sum_{t \in \mathbb{Z}_p} |\mu(t) e^{-2\pi i k t/p}| = \sum_{t \in \mathbb{Z}_p} \mu(t) = 1$$

since $|e^{ix}| = 1$ for all $x \in \mathbb{R}$.

Recall that by mixing with rate function $\phi(n) \to 0$, as $n \to \infty$, we mean that

$$d(\mu^{*n}, \lambda) \leqslant \phi(n)$$

for all $n \in \mathbb{N}$. We say that the mixing is **exponential** if for some constant $C > 0$ and $0 \leqslant \theta < 1$ we have

$$\phi(n) \leqslant C\theta^n$$

for all $n \in \mathbb{N}$. The Upper Bound Lemma implies that spectral gap is enough to prove exponential mixing:

Theorem 5.6 (Spectral gap implies exponential mixing)

If a probability distribution μ on $\mathbb{Z}_p$ has a spectral gap, then μ is exponentially mixing.

Proof
Set

$$\theta = \max\{|\widehat{\mu}(k)| : k \in \mathbb{Z}_p \backslash \{0\}\}.$$

Since μ has a spectral gap, then $0 \leqslant \theta < 1$. By the Upper Bound Lemma (recall Theorem 5.1) we have for all $n \in \mathbb{N}$ that

$$d(\mu^{*n}, \lambda) \leqslant \frac{1}{2}\sqrt{\sum_{k \in \mathbb{Z}_p \setminus \{0\}} |\widehat{\mu}(k)|^{2n}} \leqslant \frac{1}{2}\sqrt{\sum_{k \in \mathbb{Z}_p \setminus \{0\}} \theta^{2n}} = \frac{\sqrt{p-1}}{2}\theta^n.$$

Thus, by setting

$$\phi(n) = \frac{\sqrt{p-1}}{2}\theta^n$$

we have that μ is mixing with the rate $\phi(n)$ and as $0 \leqslant \theta < 1$ we know the rate is exponential with $C = \frac{\sqrt{p-1}}{2}$. □

Notice that Theorem 5.6 in particular implies that μ is ergodic if μ has a spectral gap. Using the lower bound lemma (recall Theorem 5.2) this can be made into a characterisation:

Theorem 5.7 (Spectral gap is equivalent to ergodicity)

If a probability distribution μ on $\mathbb{Z}_p$ has a spectral gap if and only if μ is ergodic.

Proof

We just need to prove the direction that if μ is ergodic, then μ has a spectral gap, as the other direction follows from Theorem 5.6. Assume on the contrary that μ does not have a spectral gap. Then we can find $\ell \neq 0$ such that $|\widehat{\mu}(\ell)| = 1$. By the Lower Bound Lemma (Theorem 5.2) we have for all $n \in \mathbb{N}$ that

$$d(\mu^{*n}, \lambda) \geqslant \frac{1}{2}\sqrt{\frac{1}{p}\sum_{k \in \mathbb{Z}_p \setminus \{0\}} |\widehat{\mu}(k)|^{2n}} \geqslant \frac{1}{2}\sqrt{\frac{1}{p}|\widehat{\mu}(\ell)|^{2n}} = \frac{1}{2\sqrt{p}}.$$

On the other hand, we assumed μ is ergodic, so by definition

$$d(\mu^{*n}, \lambda) \to 0, \quad n \to \infty.$$

Thus we can find $n \in \mathbb{N}$ such that

$$d(\mu^{*n}, \lambda) < \frac{1}{2\sqrt{p}}.$$

Contradiction. □

Let us now look at this result in a specific example of the "passing the broccoli process", which is driven by

$$\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}.$$

Now, depending on p, we will see that μ typically has a spectral gap. Hence we should be able to compute explicit estimates using the maximal Fourier coefficient of μ (recall the proof of Theorem 5.6.

Recall the Questions 1.8 from the introduction:

Questions 5.8

Q1. How many passes does it take for the broccoli to reach a given person?

Q2. How many passes does it take for the broccoli to reach every person?

Q3. How many passes do we need to take that the distribution of the broccoli is "close to random"?

We can now answer to all of these questions using the following quantitative estimates following from the Upper Bound Lemma and the definition of total variation distance. Here it depends on the mixing time we set, that is, which ε we put for the mixing time. In any case, if we set, say, $\varepsilon = 1/100$, then for $n \geqslant Cp^2$ we will have

$$d(\mu^{*n}, \lambda) \leqslant \frac{1}{100}$$

using the following explicit estimate (which follows the same idea as Theorem 5.6 above by exploiting the spectral gap of μ).

Theorem 5.9

Suppose $p \geqslant 7$ is odd and let

$$\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}.$$

Then for all $n \geqslant p^2$ we have

$$d(\mu^{*n}, \lambda) \leqslant e^{-\alpha n/p^2}$$

for $\alpha = \pi^2/2$. Moreover, for any $p \geqslant 7$ and for any $n \in \mathbb{N}$ we have a lower bound

$$d(\mu^{*n}, \lambda) \geqslant \frac{1}{2\sqrt{p}} e^{-\alpha n/p^2 - \beta n/p^4}$$

with $\beta = \pi^4/11$.

Proof

In Example 4.5 we computed the Fourier transforms:

$$\widehat{\mu}(k) = \cos(2\pi k/p).$$

Hence by the Upper Bound Lemma (Theorem 5.1) we have

$$d(\mu^{*n}, \lambda)^2 \leqslant \frac{1}{4} \sum_{k \in \mathbb{Z}_p \setminus \{0\}} |\cos(2\pi k/p)|^{2n} = \frac{1}{4} \sum_{k=1}^{p-1} |\cos(2\pi k/p)|^{2n}.$$

Note that as p is odd, the number $(p-1)/2 \in \mathbb{Z}$. Reordering summation gives us (exercise!)

$$\sum_{k=1}^{p-1} |\cos(2\pi k/p)|^{2n} = 2 \sum_{k=1}^{(p-1)/2} |\cos(\pi k/p)|^{2n}.$$

Hence

$$d(\mu^{*n}, \lambda)^2 \leqslant \frac{1}{2} \sum_{k=1}^{(p-1)/2} |\cos(\pi k/p)|^{2n}$$

We know that if $x \in [0, \pi/2]$, then

$$\cos x \leqslant e^{-x^2/2}$$

(again, an exercise, or using Taylor series of cosine around 0). This gives by the geometric series formula

$$\begin{aligned} d(\delta_0 * \mu^{*n}, \lambda)^2 &\leqslant \frac{1}{2} \sum_{k=1}^{(p-1)/2} |\cos(\pi k/p)|^{2n} \\ &\leqslant \frac{1}{2} \sum_{k=1}^{(p-1)/2} e^{-\pi^2 k^2 n/p^2} \\ &\leqslant \frac{1}{2} e^{-\pi^2 n/p^2} \sum_{k=1}^{\infty} e^{-\pi^2 (k^2-1) n/p^2} \\ &\leqslant \frac{1}{2} e^{-\pi^2 n/p^2} \sum_{k=1}^{\infty} e^{-3\pi^2 k n/p^2} \\ &= \frac{1}{2} e^{-\pi^2 n/p^2} \cdot \frac{1}{1 - e^{-3\pi^2 n/p^2}}. \end{aligned}$$

If we now assume $n \geqslant p^2$, then we know that the coefficient

$$\frac{1}{2(1 - e^{-3\pi^2 n/p^2})} < 1.$$

Hence for these n we have

$$d(\delta_0 * \mu^{*n}, \lambda)^2 \leqslant e^{-\pi^2 n/p^2}$$

as claimed.

As for the lower bound, we can see that in the sum the main contribution comes from the term $k_0 = (p-1)/2$, that is, the term

$$\widehat{\mu}(k_0) = \cos(2\pi k_0/p) = \cos(\pi - \pi/p) = -\cos(\pi/p).$$

Then, using the Lower Bound Lemma (Theorem 5.2), we obtain

$$d(\mu^{*n}, \lambda) \geqslant \frac{1}{2} \sqrt{\frac{1}{p} \sum_{k \in \mathbb{Z}_p \setminus \{0\}} |\widehat{\mu}(k)|^{2n}} \geqslant \frac{1}{2} \sqrt{\frac{1}{p} |\widehat{\mu}(k_0)|^{2n}} = \frac{1}{2\sqrt{p}} |\cos(\pi/p)|^n.$$

When $x \leqslant 1/2$, then $\cos(x) \geqslant e^{-x^2/2 - x^4/11}$ (again exercise using approximation of cosine around 0), which gives the desired lower bound. □

Chapter 6

Applying the ideas beyond $\mathbb{Z}_p$

6.1 Random walks on general finite groups G

As we have mentioned in the introduction, the methods presented in $\mathbb{Z}_p$ are possible to generalise into very general settings. However, due to the pleasant algebraic properties of $\mathbb{Z}_p$ (such as Abelian), complications will arise in particular in the "harmonic analysis" part, which we will discuss in a later section.

We will define now the concepts of probability distributions and random walks on a general finite group G. One could do this in an infinite topological groups (such as Lie groups) or other more general settings but then one requires theory from those settings (such as *Haar measure*), which we not assume the reader to necessary have.

From now on, we will assume G is some finite group, an example could be the symmetric group S_n and its subgroups like the Rubik's cube group $\mathcal{R}$. We will think about G being a **multiplicative group** in the notation in the sense that we write xy as the group operation of $x \in G$ and $y \in G$. iIf $x \in G$, we will write $x^{-1} \in G$ as its inverse and let $1 \in G$ be the identity/neutral element, that is,

$$x1 = 1x = x.$$

Definition 6.1 (Probability distributions on G)

A function $\mu : G \to [0, 1]$ is a **probability distribution** if

$$\sum_{x \in G} \mu(x) = 1.$$

The key examples of probability distributions on G are the uniform and singular distributions:

Definition 6.2 (Uniform and singular distributions)

The **uniform distribution** λ on G is defined by

$$\lambda(x) = \frac{1}{|G|}.$$

The **singular distribution** δ_y at $y \in G$ is defined by

$$\delta_y(x) = \begin{cases} 1, & x = y; \\ 0, & x \neq y. \end{cases}$$

We can extend every probability distribution $\mu : G \to [0,1]$ to all subsets $A \subset G$ as we did in $\mathbb{Z}_p$:

$$\mu(A) := \sum_{x \in G} \mu(x).$$

Now, similarly to the case of $\mathbb{Z}_p$, we can define the **total variation distance** between two probability distributions $\mu, \nu : G \to [0,1]$ by

$$d(\mu, \nu) = \max\{|\mu(A) - \nu(A)| : A \subset G\},$$

which can be proven to have a similar L^1 formula (exercise):

$$d(\mu, \nu) = \frac{1}{2} \sum_{x \in G} |\mu(x) - \nu(x)|.$$

Next, we can also define convolutions in general finite groups

Definition 6.3 (Convolutions on G)

Let $f, g : G \to \mathbb{C}$ be functions. Then the **left convolution** $f *_L g$ is defined by

$$f *_L g(x) = \sum_{y \in G} f(x^{-1}y)g(y).$$

We could also define the **right convolution** $f *_R g$ is then defined by

$$f *_R g(x) = \sum_{y \in G} f(xy^{-1})g(y).$$

Note that $f *_L g$ is not necessarily the same as $f *_R g$. If G is Abelian, then $*_L = *_R$. From now on, as in the case of $\mathbb{Z}_p$, we will concentrate on the definition of the **right convolution** $*_R$ throughout the rest of the analysis and just simply write $* = *_R$.

If we are dealing with a non-Abelian group (like the symmetric group S_n), then to define a random walk on S_n we need to choose a preference which convolution we use. Commonly one uses left convolution in literature to model a walk.

Definition 6.4 (Iterated convolutions on G)

Let $\mu : G \to [0,1]$ be a probability distribution. Then the n-**iterated convolution** is defined by

$$\mu^{*n} = \mu^{*(n-1)} * \mu,$$

for $n \geqslant 1$ with $\mu^{*0} = \delta_1$.

A **random walk** on G is defined by an i.i.d. sequence of G-valued random variables $x_1, x_2, \ldots$ with driving distribution μ, that is,

$$\mathbb{P}(x_j = x) = \mu(x), \quad \forall x \in G.$$

In particular then the "product"

$$X_n := x_1 \ldots x_n \in G$$

has distribution μ^{*n}:

$$\mathbb{P}(X_n = x) = \mu^{*n}(x), \quad \forall x \in G.$$

Example 6.5 (Gilbert-Shannon-Reeds riffle shuffle)

Take $G = S_{52}$, the symmetric group of permutations of $\{0, 1, \ldots, 51\}$. Say, we consider the riffle shuffle model defined by Gilbert-Shannon-Reeds. We say that a permutation $\sigma \in S_{52}$ is a **riffle shuffle** if σ has exactly two **rising sequences**. A rising sequence of a permutation $\sigma \in S_{52}$ is a maximal set of consecutive values that occur in the correct relative order in σ. Then it can be checked that the probability distribution $\mu : S_{52} \to [0,1]$ defined by Gilbert-Shannon-Reeds has the formula:

$$\mu(\sigma) = \begin{cases} 53 \cdot 2^{-52}, & \sigma = e; \\ 2^{-52}, & \sigma \text{ is a riffle shuffle}; \\ 0, & \text{otherwise}, \end{cases}$$

where e is the identity permutation. Then μ^{*n} models the state of random state of the deck after n riffle shuffles.

Example 6.6 (Random transpositions)

Recall that the random transposition is driven by the probability distribution $\mu : S_{52} \to [0,1]$ defined by

$$\mu(\sigma) = \begin{cases} \frac{1}{52}, & \text{if } \sigma = e; \\ \frac{2}{52^2}, & \text{if } \sigma \text{ is a transposition} \\ 0, & \text{otherwise .} \end{cases}$$

Now, as in the case of $\mathbb{Z}_p$, we can talk about the dynamics of the random walk generated by a probability distribution $\mu : G \to [0,1]$. For this purpose, we can define ergodicity as in $\mathbb{Z}_p$ by the convergence to uniform:

Definition 6.7 (Ergodicity)

We say that a probability distribution $\mu : G \to [0,1]$ is **ergodic** if

$$\mu^{*n}(x) \to \lambda(x), \quad n \to \infty,$$

for all $a \in G$, where $\lambda(x) = 1/|G|$, $a \in G$, is the uniform distribution on G.

We can again characterise ergodicity using the subgroups of G as follows (left as an exercise, the proof is very similar to $\mathbb{Z}_p$ version, but be careful as G may not be Abelian):

Theorem 6.8

A probability distribution μ is ergodic if and only if the support

$$\mathrm{spt}(\mu) := \{x \in G : \mu(x) > 0\}$$

is not contained in a coset of a proper subgroup of G.

The support assumption here rules out μ being a Dirac mass at some $y \in G$, but also not concentrated in a coset of a large subgroup. Then the question comes that if $\mathrm{spt}(\mu)$ is not contained in a coset of a proper subgroup of G, which means μ is ergodic, then how fast does

$$\mu^{*n}(x) \to \lambda(x), \quad n \to \infty?$$

As in $\mathbb{Z}_p$, we can see that this is equivalent to the total variation distance converging to 0:

$$d(\mu^{*n}, \lambda) \to 0$$

so we would like to know a rate of convergence and how many iterations it may take for $d(\mu^{*n}, \lambda)$ to become sufficiently small that the $a \in G$ chosen according to μ^{*n} is close to being very uncertain. For example, in the case of $G = S_{52}$ and riffle shuffles, having $d(\mu^{*n}, \lambda)$ small enough means the state of the deck of cards is very close to being very hard to predict.

The way to do this is to introduce harmonic analysis in the group G, but this goes beyond the scope of this course, but in the final section we will attempt to do this. However, in some cases, we can write some of the theory of harmonic analysis such as $\mathbb{Z}_2^d$ and the torus $\mathbb{Z}_p^d$, which we will do in the next sections.

6.2 Random walks on the d-torus $(\mathbb{Z}_p^d, \oplus)$

A case where we can establish bounds for mixing times without venturing into the representation theory is the d-torus $\mathbb{Z}_p^d$ equipped with the natural sum of coordinates:

Definition 6.9 (Discrete torus $\mathbb{Z}_p^d$)

Let $d \in \mathbb{N}$ be a dimension and $p \in \mathbb{N}$. Write

$$\mathbb{Z}_p^d = \{(t_1, \ldots, t_d) : t_j \in \mathbb{Z}_p, j = 1, 2, \ldots, d\}$$

Thus the elements of $\mathbb{Z}_p^d$ are **vectors** with d entries from the group $\mathbb{Z}_p$. Equipping $\mathbb{Z}_p^d$ with the binary operation

$$\mathbf{t} \oplus \mathbf{s} = (t_1 \oplus s_1, \ldots, t_d \oplus s_d),$$

where

$$\mathbf{t} = (t_1, \ldots, t_d) \in \mathbb{Z}_p^d \quad \text{and} \quad \mathbf{s} = (s_1, \ldots, s_d) \in \mathbb{Z}_p^d$$

makes $(\mathbb{Z}_p^d, \oplus)$ into an Abelian group (exercise!)

Visually $\mathbb{Z}_p^d$ could be considered as a d dimensional discrete torus, see for example Figure 6.1.

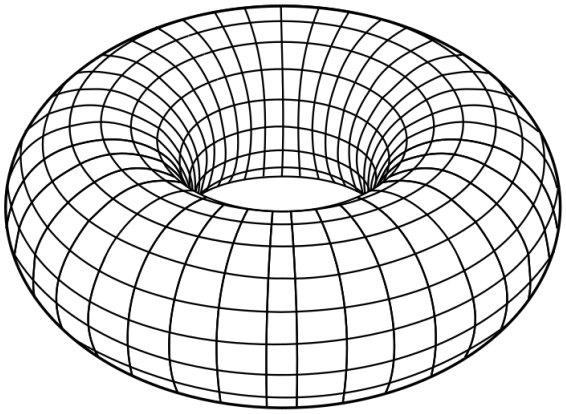

Figure 6.1: Discrete 2-torus $\mathbb{Z}_p^2$ with some value $p \in \mathbb{N}$. The $\mod p$ on both coordinates mean that we will identify 0 and $p-1$ on each coordinate (set $\mathbb{Z}_p$) so topologically we could think $\mathbb{Z}_p^2$ as a subset of the 2-torus in $\mathbb{R}^3$.

For simplicity we will consider just the case $\mathbb{Z}_2^d$, that is, $p = 2$ but a similar analysis can be carried over for general $p \geqslant 2$. The set $\mathbb{Z}_2^d$ could be regarded as a d-dimensional cube. The case $\mathbb{Z}_2^d$ has also other motivation. It can be used to model the so called *Ehrenfest's urn model* from statistical mechanics:

Ehrenfest's urn model

Suppose d balls are distributed in two "urns". Then one of the balls is chosen uniformly randomly and moved to the other urn. This process is then repeated and we would like to know what is the long-time asymptotic behaviour of this process?

We can solve Ehrenfest's urn model problem by realising it as a random walk on the group $\mathbb{Z}_2^d$ as follows. Enumerate all the balls from $1, 2, \ldots, d$. Then we can define a vector

$$\mathbf{t} = (t_1, t_2, \ldots, t_d) \in \mathbb{Z}_2^d$$

with entry $t_j = 1$ if the j:th ball is in the right urn and $t_j = 0$ if the j:th ball is the left urn. Then if we move a ball from one urn to the other, this can be described as a uniformly random choice of first a ball $j \in \{1, 2, \ldots, d\}$ (i.e. a coordinate of $\mathbf{t}$) and then swapping t_j to the opposite (e.g. if $t_j = 0$ it is changed to 1 and vice versa). This gives us a random $\mathbf{t}' \in \mathbb{Z}_2^d$, which gives a new order of the balls.

To get into the language we have used in this course, we will use the "standard basis vectors" $\mathbf{e}^j$, $j = 1, 2, \ldots, d$, to model this process. Here $\mathbf{e}^j$ is the j:th coordinate vector in $\mathbb{Z}_2^d$ defined by $\mathbf{e}^j_k = 1$ only when $j = k$ and 0 elsewhere. The link to above is that when starting from some configuration of d-balls in left and right urns, that is, we have some $\mathbf{t} \in \mathbb{Z}_2^d$, then the vector $\mathbf{t} \oplus \mathbf{e}^j$ gives a new vector in $\mathbb{Z}_p^2$ where the j:th coordinate has changed from either 0 to 1 or 1 to 0.

Thus by defining a probability distribution $\mu : \mathbb{Z}_2^d \to [0, 1]$ for all $\mathbf{t} = (t_1, \ldots, t_d) \in \mathbb{Z}_2^d$ by

$$\mu(\mathbf{t}) = \begin{cases} \frac{1}{d}, & \text{if } \mathbf{t} = \mathbf{e}^j \text{ for some } 1 \leqslant j \leqslant d; \\ 0, & \text{otherwise.} \end{cases}$$

We can model the first step of the random choice by considering the convolution

$$\mu * \delta_{\mathbf{t}}.$$

By computing this at $\mathbf{s} \in \mathbb{Z}_2^d$, we see that

$$\mu * \delta_{\mathbf{t}}(\mathbf{s}) = \mu(\mathbf{s} \ominus \mathbf{t}).$$

Thus

$$\mu * \delta_{\mathbf{t}}(\mathbf{s}) = 0$$

if $\mathbf{t}$ is not obtained from $\mathbf{s}$ by adding one of the coordinate vectors $\mathbf{e}^j$ and otherwise

$$\mu * \delta_{\mathbf{t}}(\mathbf{s}) = \frac{1}{d}.$$

Thus $\mu * \delta_{\mathbf{t}}$ tells the distribution of the d-balls after the first move. Iterating this we see that $\mu^{*n} * \delta_t$ tells us the distribution of the d-balls in the two urns after n iterations.

We can use the similar ideas as we did in the case of $\mathbb{Z}_p$ to prove the following quantitative rate of mixing for the Ehrenfest Urn model, which is effective when we choose a specific $c > 0$:

Theorem 6.10

For the Ehrenfest urn model probability distribution μ in $\mathbb{Z}_2^d$ defined above, we have for all $c > 0$ and $n \geqslant d(\log d + c)/4$ that

$$d(\mu^{*n} * \delta_{\mathbf{t}}, \lambda) \leqslant \frac{1}{\sqrt{2}}\sqrt{e^{e^{-c}} - 1}$$

for all initial configurations of d-balls $\mathbf{t} \in \mathbb{Z}_2^d$.

We will now sketch the idea of the proof. First of all, we need to define harmonic analysis on $\mathbb{Z}_2^d$. Here we can define Fourier transform using the following definition

Definition 6.11 (Fourier transform in $\mathbb{Z}_2^d$)

The **Fourier transform** of $f : \mathbb{Z}_2^d \to \mathbb{C}$ at $\mathbf{k} \in \mathbb{Z}_2^d$ is defined by

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

where $\mathbf{k} \cdot \mathbf{t}$ is the dot product

$$\mathbf{k} \cdot \mathbf{t} = k_1 t_1 + \cdots + k_d t_d.$$

This definition of Fourier transform has the same theory as the one in $\mathbb{Z}_p$, in particular, all the L^2 theory (Plancherel's theorem) and convolution theorem. Then the same strategy as we did in $\mathbb{Z}_p$ can be used to prove the following Upper Bound Lemma:

Theorem 6.12 (Upper bound lemma)

Let $\mu : \mathbb{Z}_2^d \to [0, 1]$ be a probability distribution. Then for all $n \in \mathbb{N}$ we have

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

Proof

Exercise. □

Thus to understand how fast μ^{*n} converges to uniform, we need to just understand the Fourier coefficients $\widehat{\mu}(\mathbf{k})$. In the case of the Ehrenfest Urn Model this is not hard to see:

Lemma 6.13

Consider the probability distribution μ on $\mathbb{Z}_2^d$ defined by

$$\mu(\mathbf{t}) = \begin{cases} \frac{1}{d}, & \text{if } \mathbf{t} = \mathbf{e}^j \text{ for some } 1 \leqslant j \leqslant d; \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\widehat{\mu}(\mathbf{k}) = 1 - \frac{2}{d} \cdot \sharp\{1 \leqslant j \leqslant d : k_j = 1\}.$$

Proof
Write

$$w(\mathbf{k}) = \sharp\{1 \leqslant j \leqslant d : k_j = 1\}.$$

Then

$$\begin{aligned} \widehat{\mu}(\mathbf{k}) &= \sum_{\mathbf{t} \in \mathbb{Z}_2^d} \mu(\mathbf{t})(-1)^{\mathbf{k}\cdot\mathbf{t}} \\ &= \frac{1}{d} \sum_{j=1}^{d} (-1)^{\mathbf{k}\cdot\mathbf{e}^j} \\ &= \frac{1}{d} \sum_{j=1}^{d} (-1)^{k_j} \\ &= \frac{1}{d} \Big(\sum_{k_j=1}^{d} (-1) + \sum_{k_j=0} 1 \Big) \\ &= \frac{1}{d} \Big(- w(\mathbf{k}) + (d - w(\mathbf{k})) \Big) \\ &= 1 - \frac{2}{d} w(\mathbf{k}) \end{aligned}$$

as claimed. □

Using Lemma 6.13 with the Upper Bound Lemma (Theorem 6.12) we obtain for the Ehrenfest Urn model probability distribution μ the following bound:

$$d(\mu^{*n}, \lambda)^2 \leqslant \frac{1}{4} \sum_{j=1}^{d} \binom{d}{j} \Big(1 - \frac{2j}{d}\Big)^{2n}.$$

Now a calculation shows that if $c > 0$ and $n \geqslant d(\log d + c)/4$, then

$$\frac{1}{4} \sum_{j=1}^{d} \binom{d}{j} \Big(1 - \frac{2j}{d}\Big)^{2n} \leqslant \frac{1}{2} \Big(e^{e^{-c}} - 1 \Big).$$

This completes the proof of Theorem 6.10. Note that here we only take μ^{*n} and not $\mu^{*n} * \delta_{\mathbf{t}}$, but in terms of total variation distance the distance to uniform remains unchanged.

6.3 Dual group $\widehat{G}$ and Fourier transform in G

Let us now go back to the card shuffling questions. In order thus to continue here, we would need to define "harmonic analysis" in the symmetric group S_{52}. Here we need to replace harmonic analysis by an abstract notion of **representation theory** of the symmetric group. Here the basic idea is to use the symmetries within the group to form (unitary) *representations* we can use to decompose functions $f : G \to \mathbb{R}$ as we did in the Harmonic analysis section for $\mathbb{Z}_p$. We refer to the book by Diaconis [4] for more details.

Definition 6.14 (Representations and subrepresentations)

(1) A **representation** of a finite group G is a map

$$\rho : G \to \mathrm{GL}(V_\rho),$$

which assigns to each $x \in G$ an invertible linear map $\rho(x) : V_\rho \to V_\rho$ such that

$$\rho(xy) = \rho(x)\rho(y), \quad x, y \in G.$$

Here V_ρ is some finite dimensional complex vector space depending on ρ with an inner product (V_ρ is formed with complex scalars $\mathbb{C}$) of dimension $\dim V_\rho \in \mathbb{N}$ (known as the **dimension** of the representation ρ) and $\mathrm{GL}(V_\rho)$ is the set of all invertible linear maps $L : V_\rho \to V_\rho$ (e.g. if $V_\rho = \mathbb{C}^d$, then $\mathrm{GL}(\mathbb{C}^d)$ is the set of invertible complex $d \times d$ matrices and $d = \dim V_\rho$.).

(2) An **unitary representation** of a finite group G is a representation $\rho : G \to \mathrm{GL}(V_\rho)$ such that each $\rho(x)$ is a unitary matrix, that is, the inverse $\rho(x)^{-1}$ equals to the adjoint: $\rho(x)^{-1} = \rho(x)^*$, recall that the **adjoint** A^* of A is defined by $\langle A^*v, w\rangle_{V_\rho} = \langle v, Aw\rangle_{V_\rho}$ for all $v, w \in V_\rho$. Writing $U(V_\rho)$ as the set of all unitary matrices of V_ρ, we have that a representation ρ is unitary if and only if ρ maps G to $U(V_\rho)$, that is, $\rho : G \to U(V_\rho)$. It is possible to change the inner product of V_ρ such that ρ becomes unitary in V_ρ, so in general one could assume all representations are unitary.

(2) If $\rho : G \to U(V_\rho)$ is a unitary representation and W is a subspace of V_ρ, which is **ρ-invariant**, that is, W is invariant under all the linear maps $\rho(x)$, $x \in G$:

$$\rho(x)W \subset W,$$

then the restriction $\rho|_W : G \to U(W)$ is a called a **subrepresentation**

An important example of a representation is the trivial representation:

Definition 6.15 (Trivial representations)

Given any finite dimensional complex vector space with an inner product V, then the associated **trivial representation** $\mathrm{id}_V : G \to U(V)$ is the map that acts as an identity:

$$\mathrm{id}_V(x)v = v$$

for all $x \in G$ and $v \in V$. That is, $\mathrm{id}_V(x)$ is the identity matrix of V for all $x \in G$.

Moreover, other important examples come from the irreducible representations:

Definition 6.16 (Irreducible representations)

A representation $\rho : G \to U(V_\rho)$ is **irreducible** if the only invariant subspace for $\rho(x)$ is either $\{0\}$ or the whole space V_ρ, that is, if W is a subspace of V_ρ and

$$\rho(x)(W) = \{\rho(x)w : w \in W\} \subset W,$$

then $W = \{0\}$ or $W = V_\rho$.

An important concept to analyse representations is to study whether they are isomorphic or not.

Definition 6.17 (Morphisms and isomorphisms)

Given two representations $\rho_1 : G \to U(V_{\rho_1})$ and $\rho_2 : G \to U(V_{\rho_2})$, then a linear map $\phi : V_{\rho_1} \to V_{\rho_2}$ is called **morphism** if

$$\phi \circ \rho_1(x) = \rho_2(x) \circ \phi$$

for all $x \in G$. We say that ρ_1 and ρ_2 are **isomorphic** if there exists a morphism $\rho : V_{\rho_1} \to V_{\rho_2}$ such that ρ^{-1} is invertible and is also a morphism $\phi^{-1} : V_{\rho_2} \to V_{\rho_1}$.

Now Schur's lemma gives us a powerful way to understand the structure of irreducible representations that are isomorphic:

Lemma 6.18 (Schur's lemma)

(1) Let $\rho_1 : G \to \mathrm{GL}(V_1)$ and $\rho_2 : G \to \mathrm{GL}(V_2)$ be irreducible representations of a finite group G and a morphism $\phi : V_1 \to V_2$. If ρ_1 and ρ_2 are not isomorphic, then $\phi(v) = 0$ for all $v \in V_1$.

(2) Let $\rho : G \to \mathrm{GL}(V_\rho)$ be an irreducible representation of a finite group G and a morphism $\phi : V_\rho \to V_\rho$. Then

$$\phi = \kappa_\rho I$$

for some constant κ.

Proof
The proof of (1) is similar to (2). For the proof of (2), write $V = V_\rho$. Let $\lambda \in \mathbb{C}$ be an eigenvalue of ϕ, that is, there is non-zero $v_0 \in V$ such that $\phi v_0 = \lambda v_0$. Write $E_\lambda = \{v \in V : \phi v = \lambda v\}$ the eigenspace associated to λ. Then as $\phi : V \to V$ is a morphism we have for all $v \in E_\lambda$ that

$$(\phi - \lambda)\rho(a)v = (\phi\rho(a) - \lambda\rho(a))v = (\rho(a)\phi - \lambda\rho(a))v = \rho(a)(\phi - \lambda)v = 0.$$

Thus E_λ is invariant $\phi(a)$:

$$\rho(g)E_\lambda \subset E_\lambda.$$

Since $v_0 \in E_\lambda$ is non-zero, we know that $E_\lambda \neq \{0\}$. Hence, as ϕ is irreducible, we must have $E_\lambda = V$. The only way this can happen when $\phi = \lambda I$. This completes the proof. □

The reason we talk about representations in the context of random walks on groups is that they give us the building blocks for Fourier analysis. First let us define the notion of **dual group** that acts out analogue of the frequencies $k \in \mathbb{Z}_p$ we define the Fourier transform in $\mathbb{Z}_p$. Note that it is not in general a group, but it can be proved to be a finite set as we see later.

Definition 6.19 (Dual group $\widehat{G}$)

The **dual group** of G, denoted by $\widehat{G}$, indexes all the irreducible unitary representations of G up to an isomorphism. In other words, for any $\xi \in \widehat{G}$, there exists an irreducible representation $\rho_\xi : G \to U(V_{\rho_\xi})$ and every irreducible representation of G is isomorphic to one and only one ρ_ξ. In other words $\widehat{G}$ is the set of all equivalence classes of irreducible unitary representation with the equivalence relation given by the isomorphism. Furthermore, we define $1 \in \widehat{G}$ to correspond to the class of trivial representations of G up to an isomorphism.

From now on we write $V_\xi = V_{\rho_\xi}$, when $\xi \in \widehat{G}$.

Definition 6.20 (Fourier transforms with respect to a representation)

The **Fourier transform** of $f : G \to \mathbb{C}$ at $\xi \in \widehat{G}$ is

$$\widehat{f}(\xi) := \sum_{x \in G} f(x)\rho_\xi(x).$$

At trivial representations we can always compute the Fourier transforms of probability distributions.

Example 6.21

Let $\mu : G \to [0, 1]$ be a probability distribution, then

$$\widehat{\mu}(1) = I_1$$

for the identity map of V_1: $I_1 v = v$ for all $v \in V_1$.
Proof: Indeed, as $\rho_1(x)v = v$ for all $v \in V_1$ and $x \in G$, we have that $\rho_1(x) = I_1$ is the identity map of V_1, identified with I_1, so we have

$$\widehat{\mu}(1) = \sum_{x \in G} \mu(x)\rho_1(x) = \sum_{x \in G} \mu(x)I_1 = I_1.$$

In the case $G = \mathbb{Z}_p$, compare this to the identity $\widehat{\mu}(0) = 1$.

Moreover, for the uniform distribution, we have:

Example 6.22

Let $\lambda : G \to [0,1]$ be the uniform distribution $\lambda(x) = 1/|G|$ we have for all $\xi \in \widehat{G}$ that

$$\widehat{\lambda}(\xi) = \begin{cases} I_1, & \xi = 1; \\ 0, & \xi \neq 1, \end{cases}$$

where 0 is the zero-representation: $0(x) = 0$ the zero matrix for all $x \in G$.

Proof: The case $\xi = 1$ was done in the previous example. Suppose $\xi \neq 1$ so ρ_ξ is non-trivial irreducible representation. In particular there exists $x_0 \in G$ such that $\rho_\xi(x_0) \neq I_\xi$, the identity matrix of V_ξ. By computing now the Fourier transform, we see

$$\widehat{\lambda}(\xi) = \sum_{x \in G} \lambda(x)\rho_\xi(x) = \frac{1}{|G|} \sum_{x \in G} \rho_\xi(x).$$

Define now the set

$$W := \sum_{x \in G} \rho_\xi(x)V_\xi = \{\sum_{x \in G} \rho_\xi(x)v : v \in V_\xi\}$$

Then $W \subset V_\xi$ since every $\rho_\xi(x)v \in V_\xi$ for all $v \in V_\xi$ and V_ξ is a vector space. Moreover, for since $\rho_\xi : G \to U(V_\xi)$ is a homomorphism any $x, y \in G$, as the map $x \mapsto yx$ is a bijection, we have

$$\rho_\xi(y) \sum_{x \in G} \rho_\xi(x)v = \sum_{x \in G} \rho_\xi(yx)v = \sum_{x \in G} \rho_\xi(x)v \in W.$$

Thus W is ρ_ξ invariant. Hence $W = \{0\}$ or $W = V$. We now have two cases:

(1) If $W = \{0\}$, we are done as then $\sum_{x \in G} \rho_\xi(x) = 0$, the zero representation, so $\widehat{\lambda}(\xi) = 0$.

(2) If $W = V$, we will have a contradiction with $\rho_\xi(x_0) \neq I_\xi$. Indeed, when $W = V$, this means that $\sum_{x \in G} \rho_\xi(x)$ is invertible. This is impossible since by the bijectivity of $x \mapsto x_0^{-1}x$ we have

$$\sum_{x \in G} \rho_\xi(x) = \sum_{x \in G} \rho_\xi(x_0)\rho_\xi(x_0^{-1}x) = \rho_\xi(x_0) \sum_{x \in G} \rho_\xi(x),$$

which, after taking inverses of $\sum_{x \in G} \rho_\xi(x)$ from the right gives

$$I_\xi = \rho_\xi(x_0)I_\xi,$$

so $\rho_\xi(x_0) = I_\xi$, a contradiction.

Example 6.23 (Relation to harmonic analysis in $\mathbb{Z}_p$)

Representation theory and Fourier transforms with respect to representations extend naturally the case of $\mathbb{Z}_p$.

(1) The dual group $\widehat{\mathbb{Z}_p}$ can be identified with $\mathbb{Z}_p$. Indeed, every unitary representation of $\mathbb{Z}_p$ is isomorphic to the unitary representation:

$$\rho_k(t) = e^{-2\pi ikt/p}, \quad t \in \mathbb{Z}_p,$$

for each $k \in \mathbb{Z}_p$. Then we can identify $e^{-2\pi ikt/p} \in \mathbb{C}$ as a 1×1 matrix in $\mathbb{C}$, that is,

$$e^{-2\pi ikt/p} = (e^{-2\pi ikt/p}) \quad \text{as a matrix on } \mathbb{C}.$$

Then the action on the elements $z \in \mathbb{C}$ in the vector space $\mathbb{C}$ are defined naturally by

$$(e^{-2\pi ikt/p})z = e^{-2\pi ikt/p}z \in \mathbb{C}$$

so $\rho_k(t)$ is a rotation in $\mathbb{C}$ with angle $\theta = -2\pi kt/p$ so each $\rho_k(t)$ is an invertible 1×1 matrix in $\mathbb{C}$. Thus $\rho_k(t) \in \mathrm{U}(\mathbb{C})$ with dimension of the representation $d_{\rho_k} = 1$ for all $k \in \mathbb{Z}_p$ as the dimension of the vector space $V = \mathbb{C}$ is 1 when thinking $\mathbb{C}$ as the scalars.

(2) We see that $\rho_k : \mathbb{Z}_p \to \mathrm{U}(\mathbb{C})$ is a homomorphism: if $t, s \in \mathbb{Z}_p$, then for all $z \in \mathbb{C}$ we have:

$$\rho_k(t \oplus s)z = e^{-2\pi ik(t\oplus s)/p} = e^{-2\pi ikt/p}e^{-2\pi iks/p}z = \rho_k(t)\rho_k(s)z.$$

(3) Finally we see that every ρ_k is irreducible as the only subspaces of $\mathbb{C}$ are the trivial ones $\{0\}$ and $\mathbb{C}$. Hence the definition of the Fourier transform on $\mathbb{Z}_p$, after identifying $\mathbb{C}$ scalars by 1×1 matrices in $\mathbb{C}$, that for any $k \in \mathbb{Z}_p$ we have

$$\widehat{f}(k) = \sum_{t\in\mathbb{Z}_p} f(t)\rho_k(t) = \sum_{t\in\mathbb{Z}_p} f(t)e^{-2\pi ikt/p}$$

Then Fourier transform on G obeys a convolution theorem. Here, recall, we use always the right-convolution $* = *_R$.

Theorem 6.24 (Convolution theorem)

For all $f : G \to \mathbb{C}$ and $\xi \in \widehat{G}$, we have

$$\widehat{f * g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi).$$

Proof

Firstly, we have the following invariance for summations: for every $h : G \to \mathbb{C}$ and $b \in G$ we

have

$$\sum_{x\in G} h(x) = \sum_{x\in G} h(xy^{-1}). \tag{6.1}$$

This is just a reparametrisation: the map $x \mapsto xy^{-1}$ is a bijection $G \to G$ so we will count each value in both sums in (6.1) exactly once.

Fix $\xi \in \widehat{G}$ and thus an irreducible representation $\rho_\xi : G \to U(V_\xi)$. Then after changing the order of summation and using $x = xy^{-1}x$, we have after abbreviating $\rho = \rho_\xi$:

$$\begin{aligned}
\widehat{f * g}(\xi) &= \sum_{x\in G} f * g(x)\rho(x) \\
&= \sum_{x\in G} \Big(\sum_{y\in G} f(xy^{-1})g(y) \Big)\rho(x) \\
&= \sum_{x\in G}\sum_{y\in G} f(xy^{-1})g(y)\rho(x) \\
&= \sum_{y\in G}\sum_{x\in G} f(xy^{-1})g(y)\rho(x) \\
&= \sum_{y\in G}\sum_{x\in G} f(xy^{-1})g(y)\rho(xy^{-1})\rho(y) \\
&= \sum_{y\in G} g(y)\Big(\sum_{x\in G} f(xy^{-1})\rho(xy^{-1}) \Big)\rho(y) \\
&= \sum_{y\in G} g(y)\Big(\sum_{x\in G} f(x)\rho(x) \Big)\rho(y) \\
&= \sum_{y\in G} g(y)\widehat{f}(\rho)\rho(y) \\
&= \widehat{f}(\xi) \sum_{y\in G} g(y)\rho(y) \\
&= \widehat{f}(\xi)\widehat{g}(\xi).
\end{aligned}$$

In the second last equality we applied (6.1) for $h(x) = f(x)\rho(x)$, $x \in G$. □

6.4 L^2 theory in G and the Upper Bound Lemma

Now the plan is to do L^2 theory in the group G using the Fourier transform in G, and then prove the Upper Bound Lemma. For this purpose, write $L^2(G)$ as the space of all functions $G \to \mathbb{C}$, that is,

$$L^2(G) := \{f : G \to \mathbb{C}\},$$

which is a vector space with the operation $(f+g)(x) = f(x) + g(x)$, $x \in G$, $f, g \in L^2(G)$. Then $L^2(G)$ can be equipped with an inner product as we did in $\mathbb{Z}_p$ using the definition

Definition 6.25 (Inner product in $L^2(G)$)

The **inner product** of $f, g \in L^2(G)$ is defined by

$$\langle f, g \rangle = \sum_{x \in G} f(x)\overline{g(x)}.$$

Then the L^2 norm of a single $f : G \to \mathbb{C}$ by

$$\|f\|_2 := \sqrt{\langle f, f \rangle}.$$

We now see that every $f \in L^2(G)$ can be written as the finite linear combination of Dirac masses $f = \sum_{x \in G} f(x)\delta_x$ and the finite set $\{\delta_x : x \in G\}$ form an orthonormal basis for $L^2(G)$:

$$\langle \delta_x, \delta_y \rangle = \begin{cases} 1, & x = y; \\ 0, & x \neq y. \end{cases}$$

Thus $L^2(G)$ is a finite dimensional complex vector space with the inner product given by $\langle \cdot, \cdot \rangle$ above with orthonormal basis $\{\delta_x : x \in G\}$.

Let $L(V_\xi)$ be the set of all linear maps $V_\xi \to V_\xi$. Denote $L^2(\widehat{G})$ as the space of all functions from the dual group $\widehat{G}$ to the union $\bigcup_{\xi \in \widehat{G}} L(V_\xi)$:

$$L^2(\widehat{G}) := \Big\{F : \widehat{G} \to \bigcup_{\xi \in \widehat{G}} L(V_\xi)\Big\}.$$

Thus in particular the Fourier transform $\widehat{f} \in L^2(\widehat{G})$. Then we can equip $L^2(\widehat{G})$ with the **Hilbert-Schmidt inner product** as follows. Recall from linear algebra that here for a linear map $A : V_\xi \to V_\xi$, the map $A^* : V_\xi \to V_\xi$ is the **adjoint** of A satisfying

$$\langle A^* v, w \rangle_{V_\xi} = \langle v, Aw \rangle_{V_\xi}, \quad v, w \in V_\xi$$

for the inner product $\langle \cdot, \cdot \rangle_{V_\xi}$ in V_ξ. Moreover, if $\{e_1, \ldots, e_{\dim(V_\xi)}\}$ if an orthonormal basis of V_ξ, then the **trace** of $A : V_\xi \to V_\xi$ is given by

$$\mathrm{Tr}_{V_\xi}(A) := \sum_{j=1}^{\dim(V_\xi)} \langle Ae_j, e_j \rangle_{V_\xi},$$

which is independent of the choice of the orthonormal basis.

Definition 6.26 (Hilbert-Schmidt inner product in $L^2(\widehat{G})$)

For $F, G \in L^2(\widehat{G})$, define

$$\langle F, G\rangle_{\mathrm{HS}} := \sum_{\xi\in\widehat{G}} \dim(V_\xi)\langle F(\xi), G(\xi)^*\rangle_{\mathrm{HS},\xi},$$

where $\langle\cdot,\cdot\rangle_{\mathrm{HS},\xi}$ is given by the trace:

$$\langle A, B\rangle_{\mathrm{HS},\xi} := \mathrm{Tr}_{V_\xi}(AB^*)$$

whenever $A, B \in L(V_\xi)$.

Then the **Hilbert-Schmidt norm** of a single $F \in L^2(G)$ by

$$\|F\|_{\mathrm{HS}} := \sqrt{\langle F, F\rangle_{\mathrm{HS}}}$$

and we also write for fixed $\xi \in \widehat{G}$ that

$$\|F(\xi)\|_{\mathrm{HS},\xi} := \sqrt{\langle F(\xi), F(\xi)\rangle_{\mathrm{HS},\xi}} = \sqrt{\mathrm{Tr}_{V_\xi}(F(\xi)F(\xi)^*)}$$

so

$$\|F\|^2_{\mathrm{HS}} = \sum_{\xi\in\widehat{G}} \dim(V_\xi)\|F(\xi)\|^2_{\mathrm{HS},\xi}$$

We define the inner product in this way as now we indeed have the Plancherel's theorem:

Theorem 6.27 (Plancherel's theorem)

Let $f, g : G \to \mathbb{C}$. Then

$$\langle f, g\rangle = \frac{1}{|G|}\langle \widehat{f}, \widehat{g}\rangle_{\mathrm{HS}},$$

In the case $f = g$, this gives

$$\|f\|_2 = \frac{1}{\sqrt{|G|}}\|\widehat{f}\|_{\mathrm{HS}}.$$

To eventually prove Plancherel's theorem, we need an analogue of the geometric summation formula

$$\sum_{k=0}^{p-1} e^{i\theta k} = \frac{1 - e^{i\theta p}}{1 - e^{i\theta}}$$

that we used extensively in the $\mathbb{Z}_p$ case. The analogue of this for general groups comes from representings a very large representation of G, called regular representation, and then we can see any irreducible representation of G will be isomorphic to a subrepresentation of the regular representation. Then taking traces of both side of this expression, gives us a formula involving

so called *characters* of the group G, which we will call the *trace lemma* below, that will be the analogue of the geometric summation formula.

To make this all precise, we will now define the regular representation and define character theory needed to make this precise. Recall that $L^2(G)$ is a complex finite dimensional vector space with an inner product $\langle \dot{;} \dot{\rangle}$ defined earlier. A natural representation $\rho_G : G \to U(L^2(G))$ is given by the regular representation:

Definition 6.28 (Regular representation)

The map $\rho_G : G \to U(L^2(G))$ defined by

$$\rho_G(x)f(y) = f(x^{-1}y), \quad y \in G, f \in L^2(G)$$

is called the **regular representation** of G.

Now, our aim is to formally decompose ρ_G into a direct sum over the irreducible representations, and then take traces of this, and for this purpose we will define the direct sum of representations first.

Definition 6.29 (Direct sum of representations)

If $\rho_1 : G \to U(V_{\rho_1})$ and $\rho_2 : G \to U(V_{\rho_2})$ are representations, then we can define their **direct sum** as $\rho_1 \oplus \rho_2 : G \to U(V_{\rho_1} \oplus V_{\rho_2})$ formally as

$$\rho_1 \oplus \rho_2(x)(v_1, v_2) = (\rho_1(x)v_1, \rho_2(x)v_2), \quad x \in G, (v_1, v_2) \in V_{\rho_1} \oplus V_{\rho_2},$$

where $V_{\rho_1} \oplus V_{\rho_2} = V_{\rho_1} \times V_{\rho_2}$ is the direct sum of the vector spaces equipped with the operation $(v_1, v_2) + (v_1', v_2') = (v_1 + v_1', v_2 + v_2')$ for $v_1, v_1' \in V_1$ and $v_2, v_2' \in V_2$. Then in particular the trace

$$\mathrm{Tr}_{V_{\rho_1} \oplus V_{\rho_2}}(\rho_1 \oplus \rho_2(x)) = \mathrm{Tr}_{V_{\rho_1}}(\rho_1(x)) + \mathrm{Tr}_{V_{\rho_2}}(\rho_2(x)). \tag{6.2}$$

If we now have a direct sum $\rho_1 \oplus \cdots \oplus \rho_d$ such that each $\rho_1, \ldots, \rho_k$ are isomorphic to each other, that is, they all belong to the same equivalence class $\xi \in \widehat{G}$, then we write $d\rho_\xi := \rho_1 \oplus \cdots \oplus \rho_d$. Now the reason we introduced this notation is that we will now justify why

$$\rho_G = \bigoplus_{\xi \in \widehat{G}} \dim(V_\xi)\rho_\xi, \quad x \in G. \tag{6.3}$$

In other words, $L^2(G)$ splits into a direct sum of the spaces V_ξ, $\xi \in V_\xi$, each counted $\dim(V_\xi)$ times. To justify why this is true, we need to introduce **character theory**.

Definition 6.30 (Characters)

Given an irreducible representation ρ_ξ, the corresponding **character** is the mapping $\chi_\xi : G \to \mathbb{C}$, defined by

$$\chi_\xi(g) := \mathrm{Tr}_{V_\xi}(\rho_\xi(g)), \quad g \in G.$$

Characters are examples of **class functions**, that is, constant along the so called **conjugacy classes**

$$C(g) = \{h^{-1}gh : h \in G\}$$

of the group G, which form a partition of G in terms of the equivalence relation $x \sim y$ if and only if $x = h^{-1}yh$ for some $h \in G$:

Lemma 6.31

Let $C(g)$ be a conjugacy class in G for some $g \in G$. Then $\chi_\xi(z) = \chi_\xi g_0$ for all $z \in C(g)$.

Proof
indeed, if $h \in G$, then as trace always satisfies $\mathrm{Tr}_{V_\xi}(AB) = \mathrm{Tr}_{V_\xi}(BA)$ for any two linear maps A, B, we have by the homomorphism property of ρ_ξ that

$$\chi_\xi(h^{-1}gh) = \mathrm{Tr}_{V_\xi}(\rho_\xi(h)^{-1}\rho_\xi(g)\rho_\xi(h)) = \mathrm{Tr}_{V_\xi}(\rho_\xi(g)\rho_\xi(h)^{-1}\rho_\xi(h)) = \chi_\xi(g).$$

□

Also, an important property of characters are that there are exactly the same number of them as the number of conjugacy classes:

Lemma 6.32

The cardinality of $\widehat{G}$ is the same as the number of conjugacy classes in G.

Proof
We leave this as an exercise, but it is done in e.g. [7, Proposition 2.30]. □

Lemmas 6.31 and 6.32 together give rise to the notion of **character tables**, two dimensional tables of the (constant) values $\chi_\xi(g)$ on each conjugacy class, where on rows we list each irreducible representation ρ_ξ, $\xi \in \widehat{G}$, and on the columns we have chosen a single conjugacy class representative $g \in G$.

Characters satisfy an important orthogonality relation that follows from Schur's lemma (Lemma 6.18):

Lemma 6.33 (Schur orthogonality relations)

We have

$$\langle \chi_\xi, \chi_\eta \rangle = \sum_{g \in G} \chi_\xi(g)\overline{\chi_\eta(g)} = \begin{cases} |G|, & \text{if } \xi = \eta; \\ 0, & \text{if } \xi \neq \eta. \end{cases}$$

Proof

First of all, if (e_i) and $(\tilde{e}_j)$ be orthonormal bases of V_ξ and V_η respectively, then by the definition of trace

$$\langle \chi_\xi, \chi_\eta \rangle = \sum_i \sum_j \sum_{g \in G} \langle \rho_\xi(g) e_i, e_i \rangle_{V_\xi} \overline{\langle \rho_\eta(g)\tilde{e}_j, \tilde{e}_j \rangle_{V_\eta}}.$$

Let us prove that this equals to 0 if $\xi \neq \eta$ and if $\xi = \eta$, it is equal to

$$\sum_{i=1}^{\dim(V_\xi)} \sum_{j=1}^{\dim(V_\xi)} \frac{|G|}{\dim(V_\xi)} \overline{\langle e_i, e_j \rangle_{V_\xi}} \langle e_i, e_j \rangle_{V_\xi} = |G|.$$

First of all, if $f : V_\xi \times V_\eta \to \mathbb{C}$ is *sesquilinear*, that is, linear on the first coordinate and conjugate linear in the second, then there exists a linear map $\phi : V_\xi \to V_\eta$ such that $f(v, w) = \langle \phi(v), w \rangle_{V_\eta}$. Indeed, by sesquilinearity $f(v, w) = f(\sum_i v_i e_i, \sum_j w_j \tilde{e}_j) = \sum_j (\sum_i f(e_i, \tilde{e}_j) v_i) \overline{w}_j = \langle \phi(v), w \rangle_{V_\eta}$ with the matrix $\phi = (f(e_i, \tilde{e}_j))_{i,j}$. Fix now $v_0 \in V_\xi$ and $w_0 \in V_\eta$ and define

$$f(v_0, w_0, v, w) := \sum_{g \in G} \langle \rho_\xi(g) v, v_0 \rangle_{V_\xi} \overline{\langle \rho_\eta(g) w, \tilde{w}_0 \rangle_{V_\eta}}.$$

Now, $(v, w) \mapsto f(v_0, w_0, v, w)$ is sesquilinear, so we can find a linear $\phi_{v_0, w_0} : V_\xi \to V_\eta$ such that $f(v_0, w_0, v, w) = \langle \phi_{v_0, w_0}(v), w \rangle_{V_\eta}$ for all v and w. Moreover, by the definition of $f(v_0, w_0, v, w)$ we see that ϕ is a morphism of representations: $\phi \rho_\xi(g) = \rho_\eta(g)\phi$ for all $g \in G$. Thus by Schur's lemma (Lemma 6.18) there exists $g(v_0, w_0) \in \mathbb{C}$ such that

$$\phi_{v_0, w_0} = \begin{cases} g(v_0, w_0) I_{V_\xi}, & \text{if } \xi = \eta; \\ 0, & \text{if } \xi \neq \eta. \end{cases}.$$

Thus this completes the proof if $\eta \neq \xi$ since we can use this for $v = v_0 = e_i$ and $w = w_0 = \tilde{w}_j$.

Now if $\eta = \xi$ and $v, w \in V_\xi$ are fixed, the map $(v_0, w_0) \mapsto f(v_0, w_0, v, w)$ from $V_\xi \times V_\xi \to \mathbb{C}$ also is also sesquilinear. This then implies that the map $g : V_\xi \times V_\xi \to \mathbb{C}$ is sesquilinear. Thus, by Schur's lemma, we can find a constant $\kappa \in \mathbb{C}$ such that

$$g(v_0, w_0) = \kappa \overline{\langle v_0, w_0 \rangle_{V_\xi}}.$$

Since $\rho_\xi(x)$ is unitary, we have

$$\sum_{i=1}^{\dim(V_\xi)} |\langle \rho_\xi(x) e_1, e_j \rangle_{V_\xi}|^2 = 1$$

so by the earlier formula applied with $v = e_1$, $v_0 = e_i$, $w = e_1$, $w_0 = e_i$ we have

$$|G| = \sum_{i=1}^{\dim(V_\xi)} \sum_{x \in G} \langle \rho_\xi(g) e_1, e_i \rangle_{V_\xi} \overline{\langle \rho_\xi(g) e_1, e_i \rangle_{V_\eta}} = \sum_{i=1}^{\dim(V_\xi)} \kappa \langle e_i, e_i \rangle_{V_\xi} \langle e_1, e_1 \rangle_{V_\xi} = \kappa \dim(V_\xi)$$

by the orthonormality of (e_i). Thus $\kappa = |G|/\dim(V_\xi)$. Now, as

$$\langle \chi_\xi, \chi_\xi \rangle = \sum_i \sum_j \sum_{g\in G} \langle \rho_\xi(g)e_i, e_i\rangle_{V_\xi} \overline{\langle \rho_\xi(g)e_j, e_j\rangle_{V_\xi}} = \sum_i \sum_j \kappa \overline{\langle e_i, e_j\rangle_{V_\xi}} \langle e_i, e_j\rangle_{V_\xi}$$

the proof is complete. □

Finally, we need the following decomposition lemma that allows us to split any representation (e.g. the regular one ρ_G) into direct sum of its subrepresentations. This will be crucial for proving (6.3) later.

Lemma 6.34

Let $\rho : G \to \mathrm{GL}(V)$ be a representation and $W < V$ a ρ-invariant subspace. Then There exists ρ-invariant $U < V$ with $W \cap U = \{0\}$ such that $V = W \oplus U$.

Proof
We can always choose a vector space $W' < V$ such that $V = W \oplus W'$. Let $\pi' : W \oplus W' \to W$ be the projection $\pi'(w \oplus w') = w$, for $w \in W$ and $w' \in W'$. For $v = w \oplus w' \in V$, define

$$\pi(v) = \frac{1}{|G|} \sum_{g\in G} \rho(g)\pi'(\rho(g^{-1})v).$$

Since W is ρ invariant and $\mathrm{Im}(\pi') \subset W$, we see that $\mathrm{Im}(w) \subset W$ and the restriction to W satisfies: $\pi|_W = I_W = \pi'|_W$. Thus $\pi : V \to W$ is a projection. Writing $U = \mathrm{Ker}(\pi) = \{v \in V : \pi(v) = 0\}$ we see that $V = W \oplus U$ but also that U is ρ invariant. Indeed, if $\pi(v) = 0$, then at any $h \in G$, as the map $g \mapsto hg$ is a bijection and ρ a homomorphism, we see that

$$0 = \rho(h)\pi(v) = \frac{1}{|G|} \sum_{g\in G} \rho(h)\rho(g)\pi'(\rho(gh^{-1}hv) = \frac{1}{|G|} \sum_{g\in G} \rho(hg)\pi'(\rho((hg)^{-1}hv) = \pi(\rho(h)v).$$

□

We can then use Schur's lemma to establish the following, which is the analogue of the geometric summation formula in $\mathbb{Z}_p$ which was the cornerstone for many of the proofs.

Lemma 6.35 (Trace lemma)

Let $x \in G$. Then

$$\sum_{\xi\in\hat{G}} \dim(V_\xi)\chi_\xi(x) = \begin{cases} \sum_{\xi\in\hat{G}} \dim(V_\xi)^2 = |G|, & \text{if } x = 1; \\ 0, & \text{if } x \neq 1. \end{cases}$$

Proof
By definition of the trace and as the $\rho_G(1)$ action on the orthonormal basis of $L^2(G)$ is identity: $\rho_G(1)\delta_y = \delta_y$ for all $y \in G$, we have

$$\mathrm{Tr}_{L^2(G)}\rho_G(1) = \sum_{y\in G} \langle \rho_G(1)\delta_y, \delta_y\rangle = \sum_{y\in G} \langle \delta_y, \delta_y\rangle = |G|.$$

Moreover, when $x \neq 1$, we have $\rho_G(x)\delta_y = \delta_{xy}$, so as $xy \neq y$, we have by the orthonormality $\langle \rho_G(x)\delta_y, \delta_y \rangle = 0$ proving the trace

$$\mathrm{Tr}_{L^2(G)}\rho_G(x) = 0.$$

Now, we can conclude the claim if we can verify for all $x \in G$ that

$$\mathrm{Tr}_{L^2(G)}\rho_G(x) = \sum_{\xi \in \widehat{G}} \dim(V_\xi)\mathrm{Tr}_{V_\xi}\rho_\xi(x). \tag{6.4}$$

Now this follows if we can argue that ρ_G is isomorphic to

$$\bigoplus_{\xi \in \widehat{G}} \dim(V_\xi)\rho_\xi$$

Notice that we can identify $L^2(G)$ as the vector space $\mathbb{C}^G$ and for any irreducible representation ρ_ξ of G, the space V_ξ as a $\dim(V_\xi)$ dimensional subspace of $\mathbb{C}^{|G|}$. Thus by iterating Lemma 6.34 until we reach irreducible representations V_ξ, we see that ρ_G is isomorphic to

$$\bigoplus_{\xi \in \widehat{G}} m_\xi \rho_\xi$$

for some $m_\xi \geqslant 0$. We now just need to verify $m_\xi = \dim(V_\xi)$. Indeed, by the definition of trace

$$\dim(V_\xi) = \overline{\dim(V_\xi)} = \overline{\chi_\xi(1)} = \frac{1}{|G|}\sum_{x \in G} \mathrm{Tr}_{L^2(G)}\rho_G(x)\overline{\chi_\xi(1)}$$

since

$$\mathrm{Tr}_{L^2(G)}\rho_G(x) = \begin{cases} |G|, x = 1 \\ 0, x \neq 1. \end{cases}$$

Now, as ρ_G is isomorphic to $\bigoplus_{\xi \in \widehat{G}} m_\xi \rho_\xi$, we have

$$\frac{1}{|G|}\sum_{x \in G} \mathrm{Tr}_{L^2(G)}\rho_G(x)\overline{\chi_\xi(1)} = \frac{1}{|G|}\sum_{x \in G}\sum_{\eta \in \widehat{G}} m_\eta \chi_\eta(x)\overline{\chi_\xi(1)} = \sum_\eta m_\eta \frac{1}{|G|}\langle \chi_\eta, \chi_\xi \rangle = m_\xi$$

by Lemma 6.33, so we are done. □

We can now first establish the **inverse Fourier transform** that we will also use in the proof of Plancherel's theorem:

Theorem 6.36 (Inverse Fourier transform)

Given $F \in L^2(\widehat{G})$, the **inverse Fourier transform** $\check{F}(x)$ of F at $x \in G$ is given by

$$\check{F}(x) := \frac{1}{|G|}\sum_{\xi \in \widehat{G}} \dim(V_\xi)\mathrm{Tr}_{V_\xi}(\rho_\xi(x^{-1})F(\xi)).$$

Then $\check{\check{f}} = f$ for all $f : G \to \mathbb{C}$.

Proof
Fix $x \in G$. We need to verify

$$f(x) = \frac{1}{|G|} \sum_{\xi \in \widehat{G}} \dim(V_\xi) \mathrm{Tr}_{V_\xi}(\rho_\xi(x^{-1}) \widehat{f}(\xi)).$$

Any function $f : G \to \mathbb{C}$ can be written as $f = \sum_{y \in G} f(y)\delta_y$, so by linearity it is enough to verify the above for just $f = \delta_y$ for some $y \in G$. We have $\widehat{\delta_y}(\xi) = \rho_\xi(y)$ so the right-hand side with $f = \delta_y$ equals as $\rho_\xi(x^{-1})\rho_\xi(y) = \rho_\xi(x^{-1}y)$ that

$$\frac{1}{|G|} \sum_{\xi \in \widehat{G}} \dim(V_\xi) \mathrm{Tr}_{V_\xi}(\rho_\xi(x^{-1})\rho_\xi(y)) = \frac{1}{|G|} \sum_{\xi \in \widehat{G}} \dim(V_\xi) \mathrm{Tr}_{V_\xi}(\rho_\xi(x^{-1}y))$$

which, by Lemma 6.35 equals to $\delta_y(x)$ as claimed. □

We can now prove Plancherel's theorem in G:

Proof
(Proof of Plancherel's Theorem) We want to prove

$$\sum_{x \in G} f(x)\overline{g(x)} = \frac{1}{|G|} \sum_{\xi \in \widehat{G}} \dim(V_\xi) \mathrm{Tr}_{V_\xi}(\widehat{f}(\xi)\widehat{g}(\xi)^*).$$

Define the involution map $g^* : G \to \mathbb{C}$ by

$$g^*(x) := \overline{g(x^{-1})}, \quad x \in G,$$

where $\overline{z}$ denotes the modulus of a complex number. Then by the definition of convolution of f and g^*, we have

$$f * g^*(1) = \sum_{x \in G} f(1x^{-1})g^*(x) = \sum_{x \in G} f(x^{-1})\overline{g(x^{-1})} = \sum_{x \in G} f(x)\overline{g(x)}$$

as the map $x \mapsto x^{-1}$ is a bijection (every element in G has a unique inverse).

On the other hand, using the inverse Fourier transform to the map $f * g^*$ at 1, we have

$$f * g^*(1) = \frac{1}{|G|} \sum_{\xi \in \widehat{G}} \dim(V_\xi) \mathrm{Tr}_{V_\xi}(\rho_\xi(1^{-1})\widehat{f * g^*}(\xi)) = \frac{1}{|G|} \sum_{\xi \in \widehat{G}} \dim(V_\xi) \mathrm{Tr}_{V_\xi}(\widehat{f * g^*}(\xi)).$$

Then note that for all $\xi \in \widehat{G}$ we have the following relation:

$$\widehat{g^*}(\xi) = \widehat{g}(\xi)^*.$$

Thus by the convolution theorem we have

$$\mathrm{Tr}_{V_\xi}(\widehat{f * g^*}(\xi)) = \mathrm{Tr}_{V_\xi}(\widehat{f}(\xi)\widehat{g^*}(\xi)) = \mathrm{Tr}_{V_\xi}(\widehat{f}(\xi)\widehat{g}(\xi)^*)$$

which gives the claim. □

Now we are ready to prove the Upper Bound Lemma by Diaconis and Shashahani for general finite groups G:

Theorem 6.37 (Upper Bound Lemma for G)

For any probability distribution $\mu : G \to [0,1]$ and $n \in \mathbb{N}$:

$$d(\mu^{*n}, \lambda) \leqslant \frac{1}{2}\sqrt{\sum_{\substack{\xi \in \widehat{G} \\ \xi \neq 1}} \dim(V_\xi) \left\| \widehat{\mu}(\xi)^n \right\|_{\mathrm{HS},\xi}^2}.$$

Proof

We follow the same general steps as the proof of the Upper Bound Lemma (Theorem 5.1)

(1) Use the L^1 identity for the total variation distance and Cauchy-Schwartz inequality to obtain the inequality:

$$4d(\mu^{*n}, \lambda)^2 \leqslant |G| \|\mu^{*n} - \lambda\|_2^2.$$

(2) Now using the Fourier transform formula, we obtain

$$\|\mu^{*n} - \lambda\|_2^2 = \frac{1}{|G|} \sum_{\xi \in \widehat{G}} \dim(V_\xi) \|\widehat{\mu^{*n}}(\xi) - \widehat{\lambda}(\xi)\|_{\mathrm{HS},\xi}^2.$$

(3) At the trivial representation ρ_1, as μ^{*n} is a probability distribution, we have

$$\widehat{\mu^{*n}}(1) = I_1,$$

where I_1 is the identity matrix of V_1. Recall that for the uniform distribution $\lambda(x) = 1/|G|$ we have for all $\xi \in \widehat{G}$ that

$$\widehat{\lambda}(\xi) = \begin{cases} I_1, & \xi = 1; \\ 0, & \xi \neq 1, \end{cases}$$

where 0 is the zero-representation: $0(x) = 0$ the zero matrix for all $x \in G$. Hence we have by the convolution theorem

$$\sum_{\xi \in \widehat{G}} \dim(V_\xi) \|\widehat{\mu^{*n}}(\xi) - \widehat{\lambda}(\xi)\|_{\mathrm{HS},\xi}^2 = \sum_{\substack{\xi \in \widehat{G} \\ \xi \neq 1}} \dim(V_\xi) \|\widehat{\mu}(\xi)^n\|_{\mathrm{HS},\xi}^2,$$

which gives the claim.

□

Finally to understand the **ergodicity** and **mixing** of a random walk (X_n) driven by μ in G, we need again need to talk about the notion of **spectral gap** for μ. In our context this means that we need to analyse what are the dimensions of the irreducible representations and how close the Hilbert-Schmidt norm $\|\widehat{\mu}(\xi)\|_{\mathrm{HS},\xi}$, $\xi \neq 1$, is to $1 = \|\widehat{\mu}(1)\|_{\mathrm{HS},1}$ that corresponds to the the trivial representation. This is where understanding the character table of the group in question becomes crucial, which will open the door on bounding the Fourier transform of μ, and thus $d(\mu^{*n}, \lambda)$. Let us demonstrate in the following sections we will see how this is done more precisely in the cases of dice rolling and random transposition shuffles mentioned in the beginning of the course.

6.5 Representation theory of symmetric groups.

As the next applications to dice rolling and card shuffling need us to find the character tables for symmetric groups (in dice rolling it is S_4 and for card shuffling S_{52}). Let us now give an introduction to the representation theory of symmetric groups S_n that allow us to then bound the characters in the upper bound lemma.

First, we consider two important irreducible one-dimensional representations of S_n:

(1) *trivial representation*

$$\rho_{\xi_1}(\sigma) = \mathrm{id}_{V_{\xi_1}}(\sigma), \sigma \in S_n,$$

where $\dim V_{x_1} = 1$.

(2) *sign representation:*

$$\rho_{\xi_2}(\sigma) = \mathrm{sgn}(\sigma)\mathrm{id}_{V_{\xi_1}}(\sigma), \sigma \in S_n,$$

where $\mathrm{sgn}(\sigma)$ is the *sign* of the permutation σ, that is, $+1$ if σ even and -1 if σ is odd. Recall that a permutation is even, if there are even number of inversions for σ, i.e. pairs $i < j$ such that $\sigma(j) > \sigma(i)$.

Up to isomorphisms, these turn out to be the *only* irreducible representations of S_n:

Lemma 6.38 (Classification of dimension 1 irreducible representations of S_n)

Let $n \geqslant 2$. Then all one dimensional irreducible representations of S_n are isomorphic to either the trivial representation or the sign representation.

Proof
If ρ is a one-dimensional representation of S_n, then the transposition $\tau = (1, 2)$ (i.e. $\tau(1) = 2$ and $\tau(2) = 1$) must satisfy $\rho(\tau) = 1$ or $\rho(\tau) = -1$ since $\tau^2 = e$. On the other hand, any other transposition $\tau' = (i, j)$ satisfies $(1, i)(2, j)(1, 2)(2, j)(1, i) = (i, j)$ so τ' can be conjugated to τ. Thus $\rho(\tau') = \rho(\tau)$, which means ρ maps every transposition to a fixed number, either 1 or -1. Now the trivial representation ρ_{ξ_1} maps all transpositions to 1 and sgn to -1 as a transposition always contains one inversion. The trivial and sign representations are not isomorphic as they are distinct and the only way to conjugate them is by 1×1 matrix that is trivial. □

Case S_3. Let us now write down the character table for S_3. First of all, we know that S_3 has three conjugacy classes given explicitly by the collection $\{\{e\}, \{(12), (13), (23)\}, \{(123), (132)\}\}$ so there are exactly three non-isomorphic irreducible representations and the dual group is $\widehat{S}_3 = \{\xi_1, \xi_2, \xi_3\}$. By Lemma 6.38, we have two different one dimensional irreducible representations given by the trivial one and

(1) *trivial irreducible representation:*

$$\rho_{\xi_1}(\sigma) = \mathrm{id}_{V_{\xi_1}}(\sigma), \sigma \in S_4,$$

where we have $\dim(V_{\xi_1}) = 1$ due to irreducibility. Thus

$$\chi_{\xi_1}(e) = 1, \chi_{\xi_1}((12)) = 1, \chi_{\xi_1}((123)) = 1$$

(2) *sign irreducible representation:*

$$\rho_{\xi_2}(\sigma) = \mathrm{sgn}(\sigma)\mathrm{id}_{V_{\xi_1}}(\sigma), \sigma \in S_4.$$

Thus, $V_{\xi_2} = V_{\xi_1}$, $\dim(V_{\xi_2}) = 1$, and

$$\chi_{\xi_2}(e) = 1, \chi_{\xi_2}((12)) = -1, \chi_{\xi_2}((123)) = 1$$

Finally, the third one is given by the following two dimensional representation:

(3) *standard representation:* $\rho_{\xi_3}(\sigma)(x_1, x_2, x_3) = (z_{\sigma^{-1}(1)}, z_{\sigma^{-1}(2)}, z_{\sigma^{-1}(3)})$ mapping $S_3 \to V_{\xi_3} := \{(z_1, z_2, z_3) \in \mathbb{C}^3 : z_1 + z_2 + z_3 = 0\}$. Then $\dim(V_{\xi_3}) = 2$ and ρ_{ξ_3} is irreducible and using the definition of ρ_{ξ_3} we can work out

$$\chi_{\xi_3}(e) = 2, \chi_{\xi_3}((12)) = -1, \chi_{\xi_3}((123)) = 0$$

Thus we have worked out the character table and dimensions of irreducible representations for S_3.

Case S_4. Let us now write down the character table for S_4 and the dimensions of the irreducible representations. This goes slightly beyond the scope of the course and is thus not examinable.

First of all, it turns out that there are exactly 5 disjoint conjugacy classes: determined whether the permutation is an identity, transposition, product of two disjoint transpositions, cycle of length 3 or a cycle of length 4 of S_4, and we can list the classes $C(\sigma_j)$ by the elements:

$$\sigma_1 = e, \sigma_2 = (12), \sigma_3 = (12)(34), \sigma_4 = (123), \sigma_5 = (1234),$$

and the cardinalities of each of these classes are $1, 6, 3, 8, 6$ respectively which add to 24, the cardinality of S_4. Recall that the number of irreducible representations will be the same as the number of conjugacy classes, so we can list the dual group $\widehat{S}_4 = \{\xi_1, \ldots, \xi_5\}$.

What are these 5 irreducible representations ρ_{ξ_j} corresponding to each $\xi_j \in \widehat{S}_4$, and their character table? From Lemma 6.38 we know the first two already:

(1) *trivial irreducible representation:*

$$\rho_{\xi_1}(\sigma) = \mathrm{id}_{V_{\xi_1}}(\sigma), \sigma \in S_4,$$

where we have $\dim(V_{\xi_1}) = 1$ due to irreducibility. Thus

$$\chi_{\xi_1}(e) = 1, \chi_{\xi_1}((12)) = 1, \chi_{\xi_1}((12)(34)) = 1, \chi_{\xi_1}((123)) = 1, \chi_{\xi_1}((1234)) = 1$$

(2) *sign irreducible representation:*

$$\rho_{\xi_2}(\sigma) = \mathrm{sgn}(\sigma)\mathrm{id}_{V_{\xi_1}}(\sigma), \sigma \in S_4.$$

Thus, $V_{\xi_2} = V_{\xi_1}$, $\dim(V_{\xi_2}) = 1$, and

$$\chi_{\xi_2}(e) = 1, \chi_{\xi_2}((12)) = -1, \chi_{\xi_2}((12)(34)) = 1, \chi_{\xi_2}((123)) = 1, \chi_{\xi_2}((1234)) = -1$$

Next we will work out a third one, similarly as in S_3, but a bit more elaborately now:

(3) *standard representation* ρ_{ξ_3} is defined as follows. Consider the representation $\tilde{\rho}$, which maps any $\sigma \in S_4$ to the corresponding permutation matrix P_{σ_4}, which permutes the columns of the 4×4 identity matrix I_4 according to σ. Then the subspace $W = \text{span}(e_1 + e_2 + e_3 + e_4)$, where e_j are the basis vectors of $\mathbb{R}^4$, is $\tilde{\rho}$-invariant. Then the *standard representation* ρ_{ξ_3} is the subrepresentation of $\tilde{\rho}$ to the $\tilde{\rho}$ invariant orthogonal complement $W^{\perp} = \text{span}(e_2 - e_1, e_3 - e_1, e_4 - e_1)$, so $\dim(V_{\xi_3}) = 3$. Using this basis one could compute the character χ_{ξ_3}:

$$\chi_{\xi_3}(e) = 3, \chi_{\xi_3}((12)) = 1, \chi_{\xi_3}((12)(34)) = -1, \chi_{\xi_3}((123)) = 0, \chi_{\xi_3}((1234)) = -1$$

Moreover, using the definition of our inner product:

$$\frac{1}{|S_4|}\langle \chi_{\xi_3}, \chi_{\xi_3} \rangle = 1,$$

we see that ρ_{ξ_3} is irreducible. This follows from the fact that ρ is a representation of S_4, then $g \mapsto \text{Tr}_{V_\rho}(\rho(g))$ is a linear combination of characters of S_5 with integer coefficients (Exercise).

This allows us to find a fourth one:

(4) *sign tensored with standard representation:* this is an irreducible representation formally defined as the tensor product $\rho_{\xi_4} = \rho_{\xi_2} \otimes \rho_{\xi_3}$, that is,

$$\rho_{\xi_4}(\sigma)(v_1 \otimes v_2) := (\rho_{\xi_2}(\sigma)v_1) \otimes (\rho_{\xi_3}(\sigma)v_2), v_1 \otimes v_2 \in V_{\xi_2} \otimes V_{\xi_3},$$

where $V_{\xi_2} \otimes V_{\xi_3}$ is the tensor product of the vector spaces. The *tensor product* of vector spaces can be defined using their basis vectors and then extended to all vectors. In particular, if B is the basis of V and B' is the basis of V', then $v \otimes v'$ for $v \in B, v' \in B'$ is defined as the mapping that maps (v, w) onto 1 and all other elements of $B \times B'$ to 0. In particular, this gives $V_{\xi_4} = V_{\xi_2} \otimes V_{\xi_3}$ giving $\dim(V_{\xi_4}) = 3$. This goes slightly beyond the scope of this course, but with the method of *inner tensor products* in the literature, using the already computed characters of ξ_2 and ξ_3, we can work out that

$$\chi_{\xi_4}(e) = 3, \chi_{\xi_4}((12)) = -1, \chi_{\xi_4}((12)(34)) = -1, \chi_{\xi_4}((123)) = 0, \chi_{\xi_4}((1234)) = 1$$

(5) Finally, the last representation ρ_{ξ_5} can be formally defined as follows:

$$\rho_{\xi_5} = \rho_{\xi_3} \otimes \rho_{\xi_3} - \rho_{\xi_1} - \rho_{\xi_3} - \rho_{\xi_4}$$

where for $W = V_{\xi_3} \otimes V_{\xi_3}$ we have

$$\text{Tr}_W \rho_{\xi_3} \otimes \rho_{\xi_3}(e) = 9, \text{Tr}_W \rho_{\xi_3} \otimes \rho_{\xi_3}((12)) = 1$$

$$\text{Tr}_W \rho_{\xi_3} \otimes \rho_{\xi_3}((12)(34)) = 1, \text{Tr}_W \rho_{\xi_3} \otimes \rho_{\xi_3}((123)) = 0, \text{Tr}_W \rho_{\xi_3} \otimes \rho_{\xi_3}((1234)) = 1,$$

from which we work out:

$$\chi_{\xi_5}(e) = 2, \chi_{\xi_5}((12)) = 0, \chi_{\xi_5}((12)(34)) = 2, \chi_{\xi_5}((123)) = -1, \chi_{\xi_5}((1234)) = 0$$

Moreover, using the definition of our inner product:

$$\frac{1}{|S_4|}\langle \chi_{\xi_5}, \chi_{\xi_5} \rangle = 1,$$

we see that ρ_{ξ_5} is irreducible. Finally $\dim(V_{\xi_5}) = 2$ since $\sum_{\xi \in \widehat{S_4}} \dim(V_\xi)^2 = |S_4|$ by Lemma 6.35.

Thus we have worked out the precise character table for S_4.

Case S_n for general n. Now to go to cases $n > 4$ like $n = 52$, the character tables can be come very complicated to work out. However, there is a very powerful method to study these using a method of identifying each irreducible representation of ρ of S_n with a *partition* $\lambda = (\lambda_1, \ldots, \lambda_r)$ of n, where $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \lambda_r > 0$ and $n = \lambda_1 + \cdots + \lambda_r$, we refer to [4] for more details on this. Using this identification, one has a very effective way to compute characters, given by the following *Frobenius' theorem*, that we will use later also in the card shuffling application instead of working out the whole character table:

Theorem 6.39 (Frobenius' theorem)

Let $\lambda = (\lambda_1, \ldots, \lambda_r)$ be a partition associated to an irreducible representation $\rho_\xi : S_n \to U(V_\xi)$. Then for any $\sigma \in S_n$ we have

$$\frac{\chi_\xi(\sigma)}{\dim(V_\rho)} = \frac{1}{n(n-1)} \sum_{j=1}^{r} \lambda_j^2 - (2j-1)\lambda_j.$$

See e.g. [4] for more details on this and references.

6.6 How many dice rolls are enough?

Recall the dice rolling group $\mathcal{D}$, which we can identify as S_4. In Questions 1.7, recall that we described the random dice rolling using the three rotations α, β and γ with probability 1/3 each: Thus to model this random dice rolling in our language, we can model this as a random

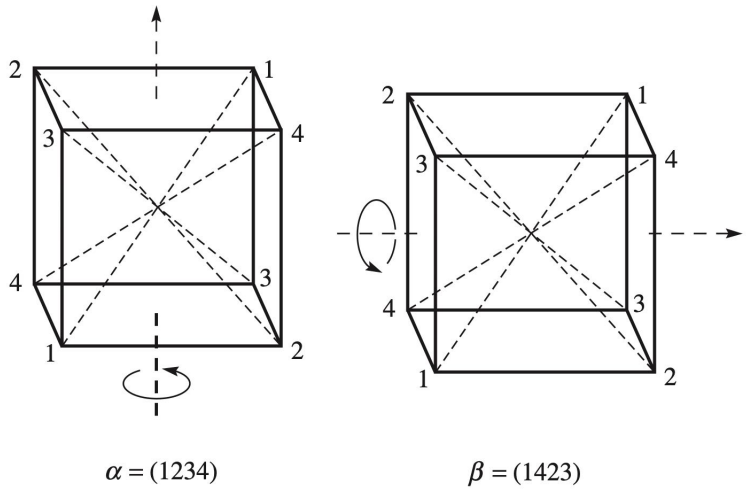

Figure 6.2: Three rotations $\alpha = (1234)$, $\beta = (1423)$ and $\gamma = (123)$ (cycles diagonal 1 to 2 and 2 to 3, and 3 to 1) that we use to describe the random dice rolling, image from [8, Figure 7.3].

walk on S_4 driven by the measure:

$$\mu = \frac{1}{3}\delta_\alpha + \frac{1}{3}\delta_\beta + \frac{1}{3}\delta_\gamma.$$

Thus, in order to calculate the probabilities in Questions 1.7, we need to control the total variation distance $d(\mu^{*n}, \lambda)$ for $n \in \mathbb{N}$, where λ is the uniform distribution on S_4.

By the upper bound lemma (Theorem 6.37 earlier), we have

$$d(\mu^{*n}, \lambda) \leqslant \frac{1}{2}\sqrt{\sum_{\substack{\xi \in \widehat{S_4} \\ \xi \neq 1}} \dim(V_\xi) \left\|\widehat{\mu}(\xi)^n\right\|^2_{\mathrm{HS},\xi}}.$$

Thus to continue, we need to use the character table of S_4 in order to bound the Fourier transform of μ corresponding to $\xi \in \widehat{S_4}$, $\xi \neq 1$. Using now the character table for S_4 we did earlier in Section 6.5, let us now proceed with bounding the right-hand side of the inequality

$$d(\mu^{*n}, \lambda) \leqslant \frac{1}{2}\sqrt{\sum_{\substack{\xi \in \widehat{S_4} \\ \xi \neq 1}} \dim(V_\xi) \left\|\widehat{\mu}(\xi)^n\right\|^2_{\mathrm{HS},\xi}}.$$

Fix $\xi \in \widehat{S_4}$ and an irreducible unitary representation $\rho_\xi : S_4 \to U(V_\xi)$ for some vector space V_ξ of dimension $\dim(V_\xi)$. Then we have

$$\widehat{\mu}(\xi) = \sum_{\sigma \in S_4} \mu(\sigma)\rho_\xi(\sigma).$$

.

Now, if we fix a permutation $\gamma \in S_4$, then for any $\sigma \in S_4$ we have

$$\mu(\sigma) = \mu(\gamma^{-1}\sigma\gamma)$$

Taking Fourier transform from both sides in the representation ρ_ξ gives us

$$\begin{aligned}
\widehat{\mu}(\xi) &= \sum_{\sigma\in S_4} \mu(\gamma^{-1}\sigma\gamma)\rho_\xi(\sigma) \\
&= \sum_{\sigma\in S_4} \mu(\gamma^{-1}\sigma\gamma)\rho_\xi(\gamma)\rho_\xi(\gamma^{-1}\sigma\gamma)\rho_\xi(\gamma^{-1}) \\
&= \rho_\xi(\gamma)\Big(\sum_{\sigma\in S_4} \mu(\gamma^{-1}\sigma\gamma)\rho_\xi(\gamma^{-1}\sigma\gamma)\Big)\rho_\xi(\gamma^{-1}) \\
&= \rho_\xi(\gamma)\widehat{\mu}(\xi)\rho_\xi(\gamma^{-1}),
\end{aligned}$$

where we used $\rho_\xi(\gamma) = \rho_\xi(\gamma)\rho_\xi(\gamma^{-1}\sigma\gamma)\rho_\xi(\gamma^{-1})$ as ρ_ξ is a homomorphism and that μ has scalar valued (in $[0,1]$ in fact). Hence the map $\widehat{\mu}(\xi)$ is a morphism so as ρ_ξ is irreducible, Schur's lemma (Lemma 6.18) implies

$$\widehat{\mu}(\xi) = \kappa_\xi I_\xi$$

for some constant $\kappa_\xi \in \mathbb{C}$ and $I_\xi : V_\xi \to V_\xi$ is the identity matrix. Now, if we take a trace from both sides, we we have $\mathrm{Tr}_{V_\xi}\widehat{\mu}(\xi) = \dim(V_\xi)\kappa_\xi$, so we have arrived to the formula:

$$\widehat{\mu}(\xi) = \frac{\mathrm{Tr}_{V_\xi}\,\widehat{\mu}(\xi)}{\dim(V_\xi)} I_\xi.$$

On the other, hand, by the definition of μ, if we directly plug-in to the definition of $\widehat{\mu}$ and take a trace, using linearity and the definition of characters, we obtain:

$$\mathrm{Tr}_{V_\xi}\,\widehat{\mu}(\xi) = \frac{1}{3}(\chi_\xi(\alpha) + \chi_\xi(\beta) + \chi_\xi(\gamma)) =: r(\xi).$$

This in particular gives us

$$\widehat{\mu}(\xi)^n = \Big(\frac{r(\xi)}{\dim(V_\xi)}\Big)^n I_\xi.$$

Notice that as a diagonal matrix $\widehat{\mu}(\xi)^n$ is in particular now self-adjoint: $(\widehat{\mu}(\xi)^n)^* = \widehat{\mu}(\xi)^n$ so $\widehat{\mu}(\xi)^n(\widehat{\mu}(\xi)^n)^* = \kappa_\xi^{2n} I_\xi$. Thus the Hilbert-Schmidt norm

$$\|\widehat{\mu}(\xi)^n\|^2_{\mathrm{HS},\xi} = \mathrm{Tr}_{V_\xi}\,(\widehat{\mu}(\xi)^n(\widehat{\mu}(\xi)^n)^*) = \dim(V_\xi)\kappa_\xi^{2n} = \dim(V_\xi)\Big(\frac{r(\xi)}{\dim(V_\xi)}\Big)^{2n}$$

Thus

$$\sum_{\substack{\xi\in\widehat{S_4}\\ \xi\neq 0}} \dim(V_\xi)\|\widehat{\mu}(\xi)^n\|^2_{\mathrm{HS},\xi} = \sum_{\substack{\xi\in\widehat{S_4}\\ \xi\neq 0}} \dim(V_\xi)^2\Big(\frac{r(\xi)}{\dim(V_\xi)}\Big)^{2n}.$$

Now, provided that we can prove that $|r(\xi)| < \dim(V_\xi)$ whenever $\xi \neq 1$, that is,

$$|r(\xi_j)| < \dim(V_{\xi_j})$$

for $j = 2,3,4,5$, we can ensure decay for the total variation distance and exponential mixing for the dice rolling. Here we can finally use the character table and dimensions, which which we need to have a look at the chosen permutations $\alpha = (1234)$, $\beta = (1423)$ and $\gamma = (123)$. Note that $\beta = (1423) \in C((1234))$, so the characters have same value here as on α. We can thus compute:

$j = 2$: We have $\dim(V_{\xi_2}) = 1$, $\chi_{\xi_2}(\alpha) = -1$, $\chi_{\xi_2}(\beta) = -1$, $\chi_{\xi_2}(\gamma) = -1$, so

$$|r(\xi_2)| = \frac{|-1-1+1|}{3} = \frac{1}{3} < 1 = \dim(V_{\xi_2})$$

$j = 3$: In this case $\dim(V_{\xi_3}) = 3$, $\chi_{\xi_3}(\alpha) = -1$, $\chi_{\xi_3}(\beta) = -1$, $\chi_{\xi_3}(\gamma) = 1$, so

$$|r(\xi_3)| = \frac{|-1-1+0|}{3} = \frac{2}{3} < 3 = \dim(V_{\xi_3})$$

$j = 4$: As before $\dim(V_{\xi_4}) = 3$, $\chi_{\xi_4}(\alpha) = 1$, $\chi_{\xi_4}(\beta) = 1$, $\chi_{\xi_4}(\gamma) = -1$, so

$$|r(\xi_4)| = \frac{|1+1+0|}{3} = \frac{2}{3} < 3 = \dim(V_{\xi_4})$$

$j = 5$: We have $\dim(V_{\xi_5}) = 2$, $\chi_{\xi_5}(\alpha) = 0$, $\chi_{\xi_5}(\beta) = 0$, $\chi_{\xi_5}(\gamma) = 0$, so

$$|r(\xi_5)| = \frac{|0+1-1|}{3} = 0 < 2 = \dim(V_{\xi_5})$$

We have proved exponential mixing of μ:

$$d(\mu^{*n}, \lambda) \to 0$$

exponentially as $n \to \infty$ by the previous estimate on $d(\mu^{*n}, \lambda)$, which now allows us to compute the answers to Questions 1.7 by estimating the mixing time, which follows by taking e.g. the worst estimate above of the characters, which we can use as the spectral gap and rate.

6.7 How many shuffles is enough?

Let us now concentrate on the symmetric group S_n, with $n \geqslant 2$. In particular, we are interested of the case $n = 52$, which corresponds to the case of card shuffling. We will concentrate on the random transposition shuffle measure μ Example 6.6, defined as $\mu : S_{52} \to [0, 1]$ with the formula

$$\mu(\sigma) = \begin{cases} \frac{1}{52}, & \text{if } \sigma = e; \\ \frac{2}{52^2}, & \text{if } \sigma \text{ is a transposition} \\ 0, & \text{otherwise .} \end{cases}$$

The Fourier transform of $\widehat{\mu}$ is easiest to understand in our context. The case of μ defining riffle shuffle is similar, but the Fourier transform attains a bit more complicated form. The main result of Diaconis and Shahshahani for the random transposition shuffle is the following, which says 270 random transposition shuffles is enough to make the deck sufficiently random:

Theorem 6.40 (Diaconis-Shahshahani)

For the random transposition probability distribution μ defined in Example 6.6, we have for any $c > 0$ that

$$d(\mu^{*n}, \lambda) \leqslant 6e^{-c}$$

for $n \geqslant 103 + 26c$. Hence if $n \geqslant 270$, we have

$$d(\mu^{*n}, \lambda) \leqslant \frac{1}{100}$$

so 270 shuffles is enough to make the deck random enough under random transpositions.

Let us now outline how to approach Theorem 6.40. Fix $\xi \in \widehat{S_{52}}$ and an irreducible unitary representation $\rho_\xi : S_{52} \to U(V_\rho)$ for some vector space V_ξ of dimension $\dim(V_\xi)$. Then we have

$$\widehat{\mu}(\xi) = \sum_{\sigma \in S_{52}} \mu(\sigma)\rho_\xi(\sigma).$$

. Now, similarly as with the dice rolling, Schur's lemma (Lemma 6.18) implies

$$\widehat{\mu}(\xi) = \kappa_\xi I_\xi$$

for some constant $\kappa_\xi \in \mathbb{C}$ and $I_\xi : V_\xi \to V_\xi$ is the identity matrix. Now, if we take a trace from both sides, we get that

$$\mathrm{Tr}_{V_\xi}\, \widehat{\mu}(\xi) = \dim(V_\xi)\kappa_\xi$$

Since the support of the random transposition measure μ is by definition

$$\mathrm{spt}\, \mu = \{e\} \cup \{\tau \in S_{52} : \tau \text{ is a transposition}\}$$

we have, as we defined characters $\chi_\xi(\sigma) = \mathrm{Tr}_{V_\xi}\rho_\xi(\sigma)$ that

$$\mathrm{Tr}_{V_\xi}\, \widehat{\mu}(\xi) = \sum_{\sigma \in S_{52}} \mu(\sigma)\chi_\xi(\xi) = \mu(e)\chi_\xi(e) + \sum_{\tau \text{ is a transposition}} \mu(\tau)\chi_\xi(\tau).$$

On the other hand

$$\mu(e) = \frac{1}{52}$$

and

$$\rho_\xi(e) = I \quad \Longrightarrow \quad \chi_\xi(e) = \dim(V_\xi)$$

Thus

$$\mu(e)\chi_\xi(e) = \frac{1}{52} \cdot \dim(V_\xi).$$

If τ and τ' are transpositions, then the measures

$$\mu(\tau) = \mu(\tau') = \frac{2}{52^2}$$

and the characters

$$\chi_\xi(\tau) = \chi_\xi(\tau')$$

as transpositions are in the same conjugacy class of S_{52}. Let N be the number of transpositions. Then

$$N = \binom{52}{2} = \frac{52!}{2!(52-2)!} = \frac{52!}{2 \cdot 50!} = \frac{51 \cdot 52}{2}.$$

Thus by fixing some (in fact any one is fine) transposition $\tau_\xi \in S_{52}$, then

$$\sum_{\tau \text{ is a transposition}} \mu(\tau)\chi_\xi(\tau) = N\mu(\tau_\xi)\chi_\xi(\tau_\xi) = \frac{51 \cdot 52}{2} \cdot \frac{2}{52^2} \cdot \chi_\xi(\tau_\xi) = \frac{51}{52} \cdot \chi_\xi(\tau_\xi)$$

Therefore we have proved

$$\mathrm{Tr}_{V_\xi}\, \widehat{\mu}(\xi) = \frac{1}{52} \cdot \dim(V_\xi) + \frac{51}{52} \cdot \chi_\xi(\tau_\xi).$$

On the other hand, we earlier saw that Schur's lemma implied $\mathrm{Tr}_{V_\xi}\, \widehat{\mu}(\xi) = \dim(V_\xi)\kappa_\xi$ so after dividing by $\dim(V_\xi)$ gives

$$\kappa_\xi = \frac{1}{52} + \frac{51}{52} \cdot \frac{\chi_\xi(\tau_\xi)}{\dim(V_\xi)}$$

and so the Fourier transform takes the form

$$\widehat{\mu}(\xi) = \Big(\frac{1}{52} + \frac{51}{52} \cdot \frac{chi_\xi(\tau_\xi)}{\dim(V_\xi)}\Big) I_\xi,$$

where I_ξ is the identity matrix of V_ξ so

$$\widehat{\mu}(\xi)^n = \Big(\frac{1}{52} + \frac{51}{52} \cdot \frac{\chi_\xi(\tau_\xi)}{\dim(V_\xi)}\Big)^n I_\xi.$$

Notice that as a diagonal matrix $\widehat{\mu}(\xi)^n$ is in particular now self-adjoint: $(\widehat{\mu}(\xi)^n)^* = \widehat{\mu}(\xi)^n$ so $\widehat{\mu}(\xi)^n(\widehat{\mu}(\xi)^n)^* = \kappa_\xi^{2n} I_\xi$. Thus the Hilbert-Schmidt norm

$$\|\widehat{\mu}(\xi)^n\|^2_{\mathrm{HS},\xi} = \mathrm{Tr}_{V_\xi}\, (\widehat{\mu}(\xi)^n(\widehat{\mu}(\xi)^n)^*) = \dim(V_\xi)\kappa_\xi^{2n} = \dim(V_\xi)\Big(\frac{1}{52} + \frac{51}{52} \cdot \frac{\chi_\xi(\tau_\xi)}{\dim(V_\xi)}\Big)^{2n}$$

Thus by the Upper Bound Lemma (Theorem 6.37), we obtain

$$4d(\mu^{*n},\lambda)^2 \leqslant \sum_{\substack{\xi\in\widehat{S_{52}}\\ \xi\neq 0}} \dim(V_\xi)\big\|\widehat{\mu}(\xi)^n\big\|_{\mathrm{HS},\xi}^2 = \sum_{\substack{\xi\in\widehat{S_{52}}\\ \xi\neq 0}} \dim(V_\xi)^2\Big(\frac{1}{52}+\frac{51}{52}\cdot\frac{\chi_\xi(\tau_\xi)}{\dim(V_\xi)}\Big)^{2n}.$$

Since $\widehat{G}$ is finite, there exists the maximum

$$r := \max\Big\{\frac{\chi_\xi(\tau_\xi)}{\dim(V_\xi)} : \xi\neq 1\Big\}. \tag{6.5}$$

By establishing $r<1$ (i.e. spectral gap for μ), as

$$\frac{1}{52}+\frac{51}{52}\cdot\frac{\chi_\xi(\tau_\xi)}{\dim(V_\xi)} \leqslant \frac{1}{52}+\frac{51}{52}r,$$

we have proved exponential mixing of μ:

$$d(\mu^{*n},\lambda)\to 0$$

exponentially as $n\to\infty$ by the previous estimate on $4d(\mu^{*n},\lambda)^2$.

In order to prove $r<1$ and give quantitative estimates to it and thus the mixing time of the random walk driven by μ, we just need to again use the bounds for the character table as we did with the case of S_4. Now, instead of working out the whole large character table for S_{52}, we can rely on the Frobenius theorem (Theorem 6.39) mentioned earlier in Section 6.5. Write $2\in\widehat{S_{52}}$ corresponding to the class, which realises the maximum r in (6.5). It turns out the representation $\rho_2 : S_{52}\to U(V_2)$ corresponds to the partition $\lambda=(51,1)$ of 52 for which the dimension $\dim(V_2)=51$, see [4] for more details on this. Frobenius theorem (Theorem 6.39) applied to the partition $\lambda=(51,1)$ then gives

$$r=\frac{\chi_2(\tau_2)}{\dim(V_2)}=\frac{1}{52\cdot 51}((1^2-1)+(51^2-3\cdot 51))=\frac{48}{52}$$

so we have $r<1$. But we can use this exact form for r to get the desired quantitative estimate as follows.

In particular as $\dim(V_2)=51$, we have

$$\dim(V_2)^2\Big(\frac{1}{52}+\frac{51}{52}\cdot r\Big)^{2n} \leqslant 51^2\Big(1-\frac{2}{52}\Big)^{2n}.$$

Since $1-x\leqslant e^{-x}$ this is bounded from above by e^{-2c} when $n\geqslant 103+26c$ and $c>0$. Moreover, again using Frobenius theorem for the other irreducible representations, we can then bound

$$\sum_{\substack{\xi\in\widehat{S_{52}}\\ \xi\neq 0}} \dim(V_\xi)^2\Big(\frac{1}{52}+\frac{51}{52}\cdot\frac{\chi_\xi(\tau_\xi)}{\dim(V_\xi)}\Big)^{2n} \leqslant 144e^{-2c}.$$

Dividing by 4 and taking square root gives the claim. In particular, we see with $n:=270$ and $c:=-\frac{1}{2}\log(1/14400)\approx 4.78749...>0$ so that $n\geqslant 103+26c$ and thus

$$d(\mu^{*n},\lambda)\leqslant 144e^{-2c}=\frac{1}{100}.$$

How about riffle shuffles? In this case we need to again understand the irreducible representation ρ_2 "closest" to the trivial one ρ_1 and what values the Fourier transform of the Gilbert-Shannon-Reeds probability distribution μ corresponding to the riffle shuffle gives at $\widehat{\mu}(2)$. This is done in the works of Bayer and Diaconis "*Trailing the dovetail shuffle to its lair*" from 1992, see [1], which we refer to the interested reader.

6.8 Random walks on the circle

Let $\alpha \in (0,1)$. Consider the following i.i.d. random walk $X_1, X_2, \dots$ on the circle $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ where with probability $1/2$ we add α and with probability $1/2$ we substract α modulo 1. Now, if α is a rational number $\alpha = p/q$, then we can see that $X_1 + X_2 + \dots$ gets trapped into a periodic orbit and an arithmetic progression depending on the integers p and q.

However, if α is irrational, then things get more interesting. It can be proven, but we do not do it in this course, that if α is irrational, then the random walk $X_1 + X_2 + \dots$ spreads around evenly in the whole circle $\mathbb{S}^1$ (equidistribution), which is an analogue of the ergodicity of the random walk in $\mathbb{Z}_p$. Formally this can be written as follows: for any interval $I \subset [0,1)$

$$\lim_{n\to\infty} \mathbb{P}(X_1 + \cdots + X_n \in I) = |I|,$$

where $|I|$ is the length of I.

Then if we want a rate for the equidistribution of the random walk, that is, rate of mixing, it highly depends on how 'well approximated' by rationals α is, or more quantitatively, how close $n\alpha$ gets to a rational number when n grows. Thus we find a connection to Diophantine approximation.

Definition 6.41 (Badly approximable numbers)

We say that a real number $\alpha \in (0,1)$ is **badly approximable** (with rational numbers) if there exists a constant $c > 0$ such that for any integer $n \in \mathbb{N}$ we have

$$\|n\alpha\| \geqslant \frac{c}{n},$$

where $\|x\| = \min\{|x - p| : p \in \mathbb{Z}\}$. In order words, for some $c > 0$ we have for all rationals $p/q \in \mathbb{Q}$

$$\left|\alpha - \frac{p}{q}\right| \geqslant \frac{c}{q^2}.$$

If α is badly approximable by rational numbers, the random walk $X_1 + \cdots + X_n$ on the group $\mathbb{S}^1$ behaves quite chaotically:

Theorem 6.42

Prove that if $\alpha \in (0,1)$ is badly approximable, then there exists a constant $C > 0$ such that for any interval $I \subset [0,1)$ and $n \in \mathbb{N}$ we have

$$|\mathbb{P}(X_1 + \cdots + X_n \in I) - |I|| \leqslant \frac{C}{\sqrt{n}}.$$

In order to prove Theorem 6.42, we can use a similar idea what we have done with $\mathbb{Z}_p$. However, we need to introduce analogues of harmonic analysis to this context. Here, thankfully,

the random walk associated to $\pm\alpha \mod 1$ can be described similarly as a discrete probability distribution as the pass the broccoli random walk in $\mathbb{Z}_p$ as follows. Each X_n is identically distributed according to the probability distribution

$$\frac{1}{2}\delta_\alpha + \frac{1}{2}\delta_{-\alpha},$$

where $-\alpha = 1 - \alpha \mod 1$ in $\mathbb{S}^1$, where δ_y, at $y \in \mathbb{S}^1$, is called a *Dirac delta mass*, which we here define formally just as a function with the property $\delta_y(x) = 1$ if $x = y$ and $\delta_y(x) = 0$ otherwise.

For distributions of the form above, we can form some basic Fourier theory as we did for $\mathbb{Z}_p$. Let X be a discrete random variable on the group $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ with the probability distribution

$$\mu = \sum_{j=1}^{N} p_j \delta_{x_j}$$

where $x_j \in \mathbb{S}^1$ and $p_1 + \cdots + p_N = 1$ with $0 \leqslant p_j \leqslant 1$. We can then define the **Fourier transform of** μ by

$$\widehat{\mu}(\xi) := \sum_{j=1}^{N} \lambda(x_j) e^{-2\pi i \xi x_j}$$

at $\xi \in \mathbb{R}$. This notion of Fourier transform satisfies the convolution theorem in the same form $\widehat{\mu^{*n}} = \widehat{\mu}^n$. Moreover, we have the following analogue of the Upper Bound Lemma:

Theorem 6.43 (**Erdös-Turán inequality**)

For any interval $I \subset [0,1)$ and integer $M \in \mathbb{N}$ we have

$$|\mathbb{P}(X \in I) - |I|| \leqslant \frac{4}{M+1} + \frac{4}{\pi} \sum_{m=1}^{M} \frac{1}{m} |\widehat{\mu}(m)|.$$

Towards Theorem 6.42, using Erdös-Turán inequality, we obtain:

Lemma 6.44

For any $0 < \alpha < 1$, and for the measure

$$\mu = \frac{1}{2}\delta_\alpha + \frac{1}{2}\delta_{-\alpha}$$

we have for the random walk $X_1, X_2, \ldots, X_n$ driven by μ that for any interval $I \subset [0,1)$, $k \in \mathbb{N}$ and $M \in \mathbb{N}$ we have

$$|\mathbb{P}(X_1 + \cdots + X_n \in I) - |I|| \leqslant \frac{4}{M+1} + \frac{4}{\pi} \sum_{m=1}^{M} \frac{1}{m} e^{-4n\|2m\alpha\|^2}.$$

Proof
First of all, we have

$$\cos(2\pi x) \leqslant 1 - 4\|2x\|^2$$

for all $x \in \mathbb{R}$.

Moreover, the Fourier transform is by the cosine identity

$$\widehat{\mu}(m) = \frac{1}{2}e^{2\pi i m\alpha} + \frac{1}{2}e^{-2\pi i m\alpha} = \cos(2\pi m\alpha).$$

Hence we have

$$|\widehat{\mu}(m)| \leqslant 1 - 4\|2m\alpha\|^2$$

Moreover, we can bound using the exponential as follows:

$$1 - 4\|2m\alpha\|^2 \leqslant e^{-4\|2m\alpha\|^2}$$

using the Taylor series of exponential function for example.

Hence by the convolution theorem

$$\widehat{\mu^{*n}}(m) = \widehat{\mu}(m)^n$$

and the Erdös-Turán inequality we have that

$$\begin{aligned} |\mathbb{P}(X_1 + \cdots + X_n \in I) - |I|| &\leqslant \frac{4}{M+1} + \frac{4}{\pi}\sum_{m=1}^{M}\frac{1}{m}|\widehat{\mu^{*k}}(m)| \\ &\leqslant \frac{4}{M+1} + \frac{4}{\pi}\sum_{m=1}^{M}\frac{1}{m}e^{-4k\|2m\alpha\|^2} \end{aligned}$$

as claimed. □

We are now ready to prove Theorem 6.42:

Proof
(Proof of Theorem 6.42) Write

$$S = \sum_{m=1}^{M}\frac{1}{m}e^{-4n\|2m\alpha\|^2}.$$

By Lemma 6.44, it is enough for us to find a constant $c_0 > 0$ such that

$$S \leqslant \frac{c_0}{M+1}.$$

Choose M such that

$$M \leqslant \frac{1}{2}c\sqrt{n} < M + 1,$$

where c is the constant from the definition of badly approximability of α. Choose an integer J such that

$$2^{J-1} \leqslant M \leqslant 2^J - 1.$$

Group the sum S into J cohorts of integers $m \in [2^{j-1}, 2^j - 1]$ for $j = 1, \ldots, J$ and apply the badly approximability of α in each cohort in the way

$$\|2m\alpha\| \geqslant \frac{c}{2m} \geqslant \frac{c}{2^{j+1}} =: s_j$$

for each $m \in [2^{j-1}, 2^j - 1]$. Moreover, we have that $m_1, m_2 \in [2^{j-1}, 2^j - 1]$ distinct that

$$\|2(m_1 - m_2)\alpha\| \geqslant s_j.$$

Thus any subinterval of $[0, 1]$ of length s_j can contain at most one of the points $\|2m\alpha\|$, $m \in [2^{j-1}, 2^j - 1]$.

With this in mind, divide now $[0, 1]$ into disjoint intervals of side length s_j starting from 0 until 1, with the last interval being of length at most s_j. As any interval of length s_j can contain at most one $\|2m\alpha\|$, the distance $\|2m\alpha\| \geqslant \ell s_j$ for some integer ℓ. In the worst case they are in all of the intervals nearest to 0 or 1, except the ones containing 0 or 1, and in these case the integer ℓ is the smallest possible.

Hence we have the following crude upper bound

$$\sum_{m=2^{j-1}}^{2^j-1} e^{-4k\|2m\alpha\|^2} \leqslant \sum_{\ell=1}^{M} e^{-4k(s_j\ell)^2}$$

Thus

$$\begin{aligned} S &\leqslant \sum_{j=1}^{J} \sum_{m=2^{j-1}}^{2^j-1} \frac{1}{m} e^{-4k\|2k\alpha\|^2} \\ &\leqslant \sum_{j=1}^{J} \sum_{m=2^{j-1}}^{2^j-1} \frac{1}{2^{j-1}} e^{-4k\|2k\alpha\|^2} \\ &\leqslant \sum_{j=1}^{J} \frac{1}{2^{j-1}} \sum_{\ell=1}^{M} e^{-4k(s_j\ell)^2} \end{aligned}$$

Now as $M \geqslant 2^{J-1}$ and $M \leqslant \frac{1}{2}c\sqrt{n}$ we have $k \geqslant 2^{2J}/c^2$ and $s_j = \frac{c}{2^{j+1}}$ so we have from above

$$S \leqslant \sum_{j=1}^{J} \frac{1}{2^{j-1}} \sum_{\ell=1}^{\infty} e^{-\ell^2 \cdot 4^{J-j}}.$$

The sum over ℓ is decreasing and a geometric series so the the inner sum is bounded by the first term $\ell = 1$ times the constant $1/(1 - e^{-1})$. Thus there exists a constant $c_1 > 0$ such that

$$S \leqslant c_1 \sum_{j=1}^{J} 2^{-j+1} e^{-4^{J-j}}.$$

Again, the terms in the sum over j are decreasing in j and the sum decreases geometrically with ratio at least $1/2$, so the sum is therefore bounded by twice the final term at $j = J$, which gives

$$S \leqslant 2c_1 2^{-J+1} e^{-1}.$$

We have $M \leqslant 2^J - 1$ so there does indeed exist a constant $c_0 > 0$ such that

$$S \leqslant \frac{c_0}{M+1}.$$

□

Going beyond $\mathbb{S}^1$ and other random walks, we would need to introduce the notion of Lebesgue measure and Lebesgue integration and Haar measures. This would allow us to talk about random walks on matrix groups and other more general Lie groups, which is currently a very active field of research. We refer to the book by Benoist and Quint [2] for an overview of the field and future topics.

Bibliography

[1] **D. Bayer, P. Diaconis**: Trailing the dovetail shuffle to its lair. *Ann. Probab.*, 2, 294-313, 1992.

[2] **Y. Benoist, J-F. Quint**: Random Walks on Reductive Groups, Springer, 2016

[3] **Berestycki, N., Durrett, R.** A phase transition in the random transposition random walk. Probab. Theory Relat. Fields 136, 203–233 (2006)

[4] **P. Diaconis**: *Group Representations in Probability and Statistics*, IMS Lecture Series volume 11, Institute of Mathematical Statistics, Hayward, California, 1988

[5] **F. Ceccherini-Silberstein, T. Scarabotti, F. Tolli**: *Harmonic Analysis on Finite Groups*. Cambridge University Press, 2008.

[6] **Durrett, R.** Journal of Theoretical Probability (2003) 16: 725.

[7] **Fulton, W.** Representation Theory, A First Course, Springer.

[8] **Gallian, J.A.** Contemporary Abstract Algebra, Brooks/Cole, 7th edition, 2010.

[9] **Hannehalli, S. and Pevzner, P.A.** (1995) Transforming cabbage into turnip (polynomial algorithm for sorting signed permutations by reversals). Proceedings of the 27th Annual Symposium on the Theory of Computing, 178–189.

[10] **R. Lyons, Y. Peres**: *Probability on Trees and Networks*, Cambridge Series in Statistical and Probabilistic Mathematics, 2017

[11] **Ranz, J. M., Casals, F., and Ruiz, A.** (2001). How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus Drosophila. Genome Research 11, 230–239.

[12] **E. M. Stein, R. Shakarchi**: *Fourier Analysis: An Introduction*, Princeton Lectures in Analysis, 2011

[13] **T. Tao**: An uncertainty principle for cyclic groups of prime order. *Mathematical Research Letters* 12(1), 2003.

[14] **T. Tao, V. Vu**: *Additive Combinatorics*, Cambridge university Press, 2006

[15] **P. Walters**: *An Introduction to Ergodic Theory*, Springer, 1982