The Cut-Off Phenomenon in Random Walks on Finite Groups

A thesis submitted to the National University of Ireland, Cork for the degree of
Master of Science

Supervisor: Dr. Stephen Wills

Head of Department: Prof. Martin Stynes

Department of Mathematics

College of Science, Engineering and Food Science

National University of Ireland, Cork

September 2010

Contents

Abstract

How many shuffles are needed to mix up a deck of cards? This question may be answered in the language of a random walk on the symmetric group, S_{52}. This generalises neatly to the study of *random walks on finite groups* — themselves a special class of Markov chains. *Ergodic* random walks exhibit nice limiting behaviour, and both the quantitative and qualitative aspects of the convergence to this limiting behaviour is examined. A particular qualitative behaviour — the *cut-off phenomenon* — occurs in many examples. For random walks exhibiting this behaviour, after a period of time, convergence to the limiting behaviour is abrupt.

The aim of this thesis is to present the general theory of random walks on finite groups, with a particular emphasis on the cut-off phenomenon. It is an open problem to determine which random walks exhibit the cut-off phenomenon. There are various formulations of the cut-off phenomenon; the original — that of *variation distance* cut-off — is considered here. At present, progress is made on this problem in a case-by-case basis. There are general techniques for attacking a particular case — and many of these are presented here — but there are no truly universal results.

Throughout the thesis, examples are used to demonstrate the theory. The last chapter presents some new heuristics developed by the author in the course of his studies.

Acknowledgements

In the first instance, I would like to sincerely thank my supervisor Stephen Wills for proposing such an interesting area to work in. I am grateful for the fact that he provided assistance whenever I needed it — be it in terms of mathematics, day-to-day life in the School of Mathematical Science, or practical advice for the writing of this thesis. I am very much looking forward to working with him on my Ph.D work.

I am sincerely grateful to Teresa Buckley and her team in the School office; never once were my problems left unresolved. I would also like to thank Martin Stynes and his engineering students (also students of Stephen Wills). As their tutor, I had to go over some old material: this undoubtedly saved me from some embarrassing mistakes in my thesis!

I would like to mention my friends in the Mathematics Research lab for enhancing my life in the School. I would like to share my appreciation of my grandmother Roses' cooking: it was a regular crutch over the last year! Also my housemates Ballsie, Swarley and Aliss — it was always easy to unwind in their company after a long day at the sums.

Chapter 1

Introduction

The question, how many shuffles are required to mix up a deck of cards , does not appear to have an obvious mathematical answer. Before any kind of analysis can be done, the terms *deck of cards*, *shuffle* and *mixed up* need a precise mathematical realisation.

Consider a fresh deck of cards; in the order, $K\heartsuit, Q\heartsuit, \ldots, A\heartsuit, K\spadesuit, \ldots, A\clubsuit$. In this order, each card can be labeled $1, \ldots, 52$, and given any arrangement of the deck, a permutation $\sigma : \{1, \ldots, 52\} \to \{1, \ldots, 52\}$ can encode the arrangement:

$$\begin{pmatrix} 1 & 2 & \cdots & 52 \\ \sigma(1) & \sigma(2) & \cdots & \sigma(52) \end{pmatrix} \tag{1.1}$$

In the language of group theory, the deck of cards may be modelled by S_{52}.

A shuffle, meanwhile, takes the deck, and, *independently[1] of the arrangement of the deck*, permutes the cards. For example, a perfect cut shuffle takes off the top half of the cards and places it under the bottom half of the deck is a shuffle. It is not hard to see that a shuffle is a function $S : S_{52} \to S_{52}$, whose action is by multiplication by some $\sigma_S \in S_{52}$; i.e. $S(\sigma) = \sigma_S \sigma$. Indeed the perfect-cut shuffle is realised by multiplication by $(1, 27)(2, 28) \cdots (26, 52)$.

Now the question of when is a deck mixed-up needs to addressed. In the first instance, it is always assumed that the deck started in some known order; e.g. the one given above. Secondly, when is a deck totally random?

---

[1] *in general (!), one doesn't shuffle while looking at the labels on the cards. To be technical, not all functions $S : S_{52} \to S_{52}$ are considered shuffles. For example, the 'shuffle' swapping the positions of $A\heartsuit$ and $A\spadesuit$ is not a shuffle.*

If one is handed a deck of cards, face down, and if each possible order of the cards is equally possible then the deck is considered random. It should be clear from group theory, that if any perfect shuffle is repeated, then the deck will never get random in this sense. If the deck is always shuffled by σ_S, then after k shuffles the deck will be in the order σ_S^k. Hence, to get random, there has to be some randomisation in how the deck is shuffled. As an example of a suitable randomisation, pick two distinct cards at random[2], and let the shuffle swap the positions of these two cards. Assuming now that after a number of shuffles, every arrangement of the deck is approximately equally likely, various notions of 'how close' the deck is to random may be formulated, and a clear definition of mixed-up may be given.

Consider the *riffle shuffle*: at each step the deck is cut into two packs which are then *riffled* together. A model for such shuffles on n rather than just 52 cards, due to Gilbert, Shannon and (independently) Reeds, was completely analysed in a remarkable paper by Bayer & Diaconis [6]. In this paper, a phenomenon called the *cut-off phenomenon* was proven to occur for the riffle shuffle. Namely, for n large, the deck is far from random in a certain sense after less than $t_n = (3\log_2 n)/2$ shuffles, but close to random after more than t_n shuffles: the transition from order to random takes place at about t_n steps and it makes sense to say it takes t_n steps to mix-up the cards. For the case $n = 52$, seven shuffles are necessary and sufficient to mix up the cards.

Random walks on finite groups generalise card shuffling by replacing the symmetric group by any finite group. This thesis aims to present the general theory of random walks on finite groups, with an emphasis on the cut-off phenomenon. In particular, care has been shown to take no liberties with assumptions, and all the 'obvious' elements of the theory are revisited and questioned. For example, Theorem 1.3.2 is standard in the field but almost all references do not carry the non-trivial proof. The questioning of 'obvious' facets of the theory allowed some new perspectives.

---

[2]to be careful maybe two distinct card *positions*, e.g. top card, second card, etc.

In making the thesis modest, some interesting and often powerful aspects of the theory have been omitted. The aforementioned riffle shuffle was not studied — neither was the familiar over-hand shuffle. In fact, in terms of the development of the subject, the riffle shuffle is a pathological example. Despite its apparent complexity, the shuffle has been more or less completely understood and analysed by Bayer & Diaconis, albeit through some deeper mathematics than the subject usually requires.

The Diaconis-Fourier theory is an attractive machinery in the field that is presented here. However it is only applied in two Abelian examples: neither of which needed require the full theory anyway. Its greatest success has been in the analysis of the random transposition shuffle, a random walk on the symmetric group, however the representation theory of the symmetric group is not covered here. Diaconis [12] is an excellent reference. A great survey of techniques, including those not mentioned here is [27].

There are a number of interesting generalisations of random walks on groups, such as to homogenous spaces and Gelfand pairs. These are not covered here: Ceccherini-Silberstein *et al* [7] is an excellent book and pursues these areas.

Despite these restrictions, a great variety of mathematical techniques are used. Probability, measure theory, representation theory, functional analysis, geometry and, naturally, group theory is used throughout the thesis. The cut-off phenomenon is not just a theory for random walks on groups, it occurs for some more general Markov chains also. A breakthrough in the theory of random walks on groups will surely have an impact for the Markov chain community. In his introduction, Chen [8] discusses a few examples where the existence of a cut-off has a significant impact for applications.

This first chapter introduces the general discrete time Markov chain theory on a finite set. Random walks on groups are introduced as a special class of Markov chains and necessary and sufficient conditions for a random walk to 'get random' are developed.

Chapter 2 discusses what it means for a random walk to be 'close to random'. A number of measures of closeness to random are introduced. A distinguished distance, namely the *variation distance*, is identified as the conventional measure

of closeness to random in this study. An interpretation of variation distance by Switzer is shown to be correct here. Much of the spectral analysis of the stochastic operator is done in this chapter and this yields upper bounds on the distance to random — many related to the eigenvalues of the associated *stochastic operator*. Next techniques for finding lower bounds on the distance to random are discussed. Finally, methods of procuring bounds for these eigenvalues via the geometry of the group are presented.

Chapter 3 develops the representation theory of finite groups. In conjunction with Fourier analysis for finite groups, this machinery, so well pioneered by Diaconis, is a powerful technique for generating bounds on the distance to random. Here the full, general, theory is developed. Two Abelian examples, the simple walk on the circle and the simple walk with loops on the n-Cube, are analysed.

Chapter 4 introduces the cut-off phenomenon and its formulation. In particular, it is seen that the phenomenon is defined with respect to the limiting behavior of a family of random walks on groups, $\{G_n : n \in \mathbb{N}\}$, as the size of the group increases to infinity ($n \to \infty$). There is a discussion of the present understanding of the cut-off phenomenon, and reasons for its existence are mentioned.

Chapter 5 presents some probabilistic methods for bounding the distance to random. These powerful methods — strong uniform times and coupling — are occasionally very transparent and help explain *why* cut-offs occur.

Finally in Chapter 6 some new viewpoints and generalisations are presented. Although the motion of a particle in a random walk *is* random (in general, after k steps the position of the particle is unknown), its distribution after k steps is deterministic. Thus the random walk has the structure of a dynamical system. Here an attempt is made to develop this further. Also the question of whether or not the invertibility of the stochastic operator has implications for a random walk is addressed. A study of invertible stochastic operators is, as far as this author knows, non-existent in the literature. A few basic properties and questions are explored. Finally, a conjecture of the author, namely that if the stochastic operator is invertible, then the cut-off phenomenon will not be exhibited, is explored and disproved.

1.1 Markov Chain Theory

Essentially, a Markov Chain is a construction of a mathematical model for a certain type of discrete motion of a particle in a space. The particle begins at some initial point and at certain times $t_1, t_2, \ldots$ moves to another point in the space chosen 'at random'. The probability that the particle moves to a certain point y at a time t is dependent only upon its position x at the previous time. This is the Markov property.

To formulate, let X be a finite set. Denote by $M_p(X)$ the probability measures on X. Let δ^x be the element of $M_p(X)$ which puts a measure of 1 on x (and zero elsewhere). These *Dirac measures*, $\{\delta^x : x \in X\}$, are the canonical basis for $\mathbb{R}^{|G|} \supseteq M_p(X)$. A probability measure $\nu \in M_p(X)$ is *strict* if $\nu(x) > 0$, for all $x \in X$. Denote by $F(X)$ the complex functions on X and $L(V)$ the linear operators on a vector space V. The similarly defined *Dirac functions*, $\{\delta_x : x \in X\}$, are the canonical basis for $F(X)$. With respect to this basis $P \in L(F(X))$ has a matrix representation $[p(x,y)]_{xy}$. $P \in L(F(X))$ is a *stochastic operator* if:

(i) $p(x,y) \geq 0, \ \forall x, y$

(ii) $\sum_y p(x,y) = 1, \forall x$ (row sum is unity)

Given $\nu \in M_p(X)$, a stochastic operators P acts on ν as $\nu P(x) = \sum_y \nu(y) p(y, x)$. Stochastic operators are readily characterised without using matrix elements as being $M_p(X)$-stable in the sense that $M_p(X)P \subset M_p(X)$ if and only if P is a stochastic operator. It is an immediate consequence that if P and Q are stochastic, then so is PQ.

1.1.1 Definition

Let X be a finite set and $\nu \in M_p(X)$, P a stochastic operator on X, and $(Y, \mathcal{A}, \mu)$ a probability space. A sequence $\{\xi_k\}_{k=0}^n$ of random variables $\xi_k : Y \to X$ are a *Markov Chain with initial distribution ν and stochastic operator P*, if

  (i) $\mu(\xi_0 = x_0) = \nu(x_0)$.

  (ii) $\mu(\xi_{k+1} = x_{k+1} \,|\, \xi_0 = x_0, \ldots, \xi_k = x_k) = p(x_k, x_{k+1})$,
      assuming $\mu(\xi_0 = x_0, \ldots, \xi_k = x_k) > 0$.

If $\nu = \delta^x$ in (i) the Markov chain is said to *start deterministically* at x. Condition (ii) is the *Markov property*. Subsequent references to a Markov Chain ξ refer to a Markov Chain $(\{\xi_k\}_{k=0}^n, P, \nu)$.

In terms of existence, given ν and P, let

$$Y := X^{n+1} = \underbrace{X \times X \times \cdots \times X}_{n+1 \text{ copies}}$$

Define $\xi_k : Y \to X$ by $\xi_k(x_0, \ldots, x_n) = x_k$ and

$$\mu(x_0, \ldots, x_n) = \nu(x_0) p(x_0, x_1) \cdots p(x_{n-1}, x_m).$$

Then $\mu \in M_p(Y)$, and ξ is a Markov Chain for ν and P.

1.1.2 Example: Two State Markov Chain

Consider the set $X = \{1, 2\}$ and $\nu \in M_p(X)$. Suppose the probability of going from 1 to 2 is p and the probability of going from 2 to 1 is q. Then the two state Markov chain has stochastic operator

$$P = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix}$$
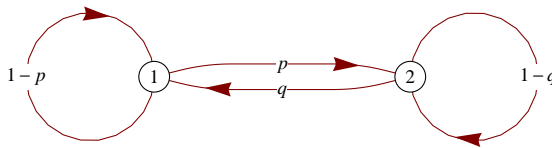
for $p, q \in [0, 1]$.

Figure 1.1: A graphical representation of the two state Markov chain.

1.2   Ergodic Theory

Ergodic theory is concerned with the longtime behaviour of a Markov chain. A central question is for a given chain whether or not the ξ_k display limiting behaviour as $k \to \infty$? If 'ξ_∞' exists, what is its distribution?

One possible debarring of the existence of a limit is periodicity. Consider a Markov chain ξ on a set $X = X_0 \cup X_1$ with $X_0 \cap X_1 = \emptyset$ and neither of the $X_i = \emptyset$ for $i = 1, 2$. Suppose ξ has the property that $\xi_{2k+i} \in X_i$, for $k \in \mathbb{N}_0$, $i = 0, 1$. Then 'ξ_∞' cannot exist in the obvious way. In a certain sense ξ must be *aperiodic* for limiting behaviour to exist.

Suppose ξ is a Markov chain and the limit $\nu P^n \to \theta$ exists. Loosely speaking, after a long time N, ξ_N has distribution $\mu(\xi_N) \sim \theta$:

$$\nu P^N \sim \theta$$
$$\Rightarrow \nu P^N P \sim \theta P$$
$$\Rightarrow \nu P^{N+1} \sim \theta P$$

But $\nu P^{N+1} \sim \theta$ also and hence $\theta P \sim \theta$. So if 'ξ_∞' exists then its distribution θ may have the property $\theta P = \theta$. Such a distribution is said to be a *stationary distribution for P*. Relaxing the supposition on 'ξ_∞' existing, do stationary distributions exist? Clearly they are left eigenvectors of eigenvalue 1 that have positive entries summing to 1.

If $k(x) \in F(X)$ is any constant function then $Pk = k$ so k is a right eigenfunction of eigenvalue 1. Let u be a left eigenvector of eigenvalue 1. By the triangle inequality, $|u(x)| = |\sum_y u(y)p(y,x)| \leq \sum_y |u(y)|p(y,x)$. Now

$$\sum_{z \in X} |u(z)| \leq \sum_{z \in X} \left(\sum_{y \in X} |u(y)|p(y,z) \right) = \sum_{y \in X} |u(y)| \underbrace{\left(\sum_{z \in X} p(y,z) \right)}_{=1} = \sum_{y \in X} |u(y)|$$

Hence the inequality is an equality so $\sum_z \left(\sum_y |u(y)|p(y,z) - |u(z)| \right) = 0$ is a sum of non-negative terms. Hence $|u|P = |u|$, and by a scaling, $\pi(x) := |u(x)|/\sum_y |u(y)|$, is a stationary distribution.

How many stationary distributions exist? Consider Markov Chains ξ and ζ on disjoint finite sets X and Y, with stochastic operators P and Q. The block matrix

$$R = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \tag{1.2}$$

is a stochastic operator on $X \cup Y$. If π and θ are stationary distributions for P and Q then

$$\phi_c = (c\pi, (1-c)\theta), \quad c \in [0,1]$$

is an infinite family of stationary distributions for R. The dynamics of this walk are that if the particle is in X it stays in X, and vice versa for Y (the graph of R has two disconnected components). This example shows that, in general, the stationary distribution need not be unique. Rosenthal [26] shows that a sufficient condition for uniqueness is that the Markov chain ξ has the property that every point is accessible from any other point; i.e. for all $x, y \in X$, there exists $r(x,y) \in \mathbb{N}$ such that $p^{(r(x,y))}(x,y) > 0$. A Markov chain satisfying this property is said to be *irreducible*.

So for the existence of a unique, stationary distribution it may be sufficient that the Markov chain is both aperiodic and irreducible. Call a stochastic operator P *ergodic* if there exists $n_0 \in \mathbb{N}$ such that

$$p^{(n_0)}(x,y) > 0, \quad \forall x, y \in X$$

In fact, ergodicity is equivalent to aperiodic and irreducible (see [26][3] Lemma 8.3.9), and the following theorem asserts that it is both a necessary and sufficient condition for the existence of a strict distribution for 'ξ_∞'. These precluding remarks suggest the distribution of 'ξ_∞' is in fact stationary and unique, and indeed this will be seen to be the case. A nice, non-standard proof of this well-known theorem is to be found in [7].

1.2.1 Markov Ergodic Theorem

A stochastic operator P is ergodic if and only if there exists a strict $\pi \in M_p(X)$ such that

$$\lim_{n \to \infty} p^{(n)}(x, y) = \pi(y) \,, \ \forall x, \, y \in X \tag{1.3}$$

In this case π is the unique stationary distribution for P $\bullet$

In the special class of ergodic Markov chains, (1.3) indicates that statistically speaking, the system that evolves for a long time 'forgets' its initial state. Another special class of Markov chains are *reversible* Markov chains. A stochastic operator P is *reversible* if there exists a strict $\pi \in M_p(X)$ such that

$$\pi(x)p(x, y) = p(y, x)\pi(y) \,, \ \forall \, x, y \in X \tag{1.4}$$

This is equivalent to $D_\pi P = P^T D_\pi$ where D_π is the diagonal matrix with (x, x)-component $\pi(x)$. Suppose further that P is ergodic and (1.4) holds for some strict $\pi \in M_p(G)$. A quick calculation shows that then π is the unique, strict, stationary distribution. The definition of a reversible chain appears at odds with our interpretation of what reversible means. However, it may be shown (see [7]) that the condition is equivalent to

(i) $p(x, y) > 0 \Rightarrow p(y, x) > 0$

(ii) for all $n \in \mathbb{N}$, $x_0, x_1, \ldots, x_n \in X$,

$$p(x_0, x_1)p(x_1, x_2) \cdots p(x_{n-1}, x_n)p(x_n, x_0) = p(x_0, x_n)p(x_n, x_{n-1}) \cdots p(x_1, x_0)$$
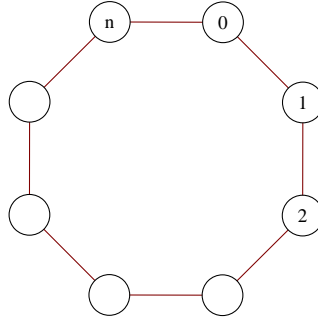
---

[3]although aperiodic hasn't been defined here

Figure 1.2: For a reversible Markov Chain, the probability of going in a cycle from $0 \to 0$ is equal for clockwise and anti-clockwise orientations.

1.3 Random Walks on Finite Groups

1.3.1 Introduction

A particularly nice class of Markov chain is that of a *random walk on a group*. The particle moves from group element to group element by choosing an element h of the group 'at random' and moving to the product of h and the present position g, i.e. the particle moves from g to hg. To avoid trivialities, the random walk on the trivial group is not considered. Naturally the group structure of the walk induces strong symmetry conditions: this allows the generation of much stronger results than that of general Markov chain theory.

To formulate, let G be a finite group of order $|G|$ and identity e. Let $\nu \in M_p(G)$ and (Y, μ) be a probability space. Let $\{\zeta_k\}_{k=0}^n : (Y, \mu) \to G$ be a sequence of i.i.d. random variables with distributions $\mu(\zeta_0 = g_0) = \delta^e(g_0)$ and $\mu(\zeta_k = g) = \nu(g)$. The sequence of random variables $\{\xi_k\}_{k=0}^n : (Y, \mu) \to G$

$$\xi_k = \zeta_k \zeta_{k-1} \cdots \zeta_1 \zeta_0 \tag{1.5}$$

is a *right-invariant random walk on G*.

This construction makes ξ into a Markov Chain on G with initial distribution δ^e and stochastic operator $P = p(s, t)$ is induced by the *driving probability*, ν: $p(s, t) = \nu(ts^{-1})$. The random walk is called *right invariant* because $p(s, t) = p(sh, th)$. This is obvious as

$$p(sh, th) = \nu(th(sh)^{-1}) = \nu(ts^{-1}) = p(s, t)$$

Example: Card Shuffling

Card shuffling provides the motivation for the study of random walks on groups and remains the canonical example. Everyday shuffles such as the overhand shuffle or the riffle shuffle, as well as simpler but more tractable examples such as top-to-random or random transpositions all have the structure of a random walk on S_{52}. Each shuffle may be realised as sampling from a probability distribution $\nu \in M_p(S_{52})$. For example, consider the case of repeated random transpositions. A *random transposition* consists chooses two cards at random (with replacement) from the deck and swapping the positions of these two cards. Suppose without loss of generality that the first card chosen is the ace of spades. The probability of choosing the ace of spaces again is $1/52$. Swapping the ace the spades with itself leaves the deck unchanged. The choice of the first card is independent hence the probability that the shuffle leaves the deck unchanged is $1/52$. What is the probability of transposing two given (distinct) cards? Consider, again without loss of generality, the probability of transposing the ace of spades and the ace of hearts. There are two ways this may be achieved: choose $A\spadesuit$-$A\heartsuit$ or choose $A\heartsuit$-$A\spadesuit$. Both of these have probability of $1/52^2$. Any other given shuffle (not leaving the deck unchanged or transposing two cards) is impossible. Hence the shuffle may be modelled as sampling by

$$\nu(s) := \begin{cases} 1/52 & \text{if } s = e \\ 2/52^2 & \text{if } s \text{ is a transposition} \\ 0 & \text{otherwise} \end{cases}$$

It is a straightforward calculation to show that the stochastic operator of a random walk on a group is doubly stochastic — column sums are also 1. As a corollary, the uniform distribution, $\pi(g) = 1/|G|$, is a strict, stationary distribution. To keep terminology to a minimum, the uniform distribution shall be referred to as the *random distribution* and conversely π will refer to this random distribution.

If $\Sigma = \text{supp}\,(\nu)$, then, in general, $\xi_k \in \Sigma^k$ however if $\langle\Sigma\rangle = G$ and $e \in \Sigma$ then certainly $\Sigma^k \subset \Sigma^l$, for any $k \leq l$. Indeed:

$$\{e\} = \Sigma^0 \subset \Sigma \subset \Sigma^2 \subset \cdots \subset \Sigma^T = G$$

where T is called the *cover time* of the walk. In this case P is ergodic with $n_0 = T$. From Section 1.2, it is known that 'ξ_∞' exists in a nice way if the stochastic operator P is ergodic. Conveniently, this condition may be translated into a condition on the driving probability on the group, ν. The below theorem falls under the category of a 'folklore theorem' in that almost all references refer to the proof in older hard-to-source references — if at all. A proof outline is given by Fountoulakis [19] in his lecture notes but here a full proof is given.

1.3.2 Ergodic Theorem for Random Walks on Groups

Let G be a group and $\nu \in M_p(G)$ with support Σ. A right-invariant random walk on G is ergodic if and only if $\Sigma \not\subset K$ for any proper subgroup K of G and $\Sigma \not\subset Hx$ for any coset of any proper normal subgroup $H \lhd G$.

In this case, π is the unique, strict stationary distribution for P.

Proof. Assume $\Sigma \subset K$ a proper subgroup of G. $\langle\Sigma\rangle \subset K$ by closure in K; hence $\xi_k \in K$, for all $k \in \mathbb{N}$. Let, $s \in K$, $t \notin K$. Now for all $n \in \mathbb{N}$, $p^{(n)}(s,t) = 0$. Hence P is *not* ergodic.

Assume $\Sigma \subset Hx$ for some coset of a proper normal subgroup $H \lhd G$. Now $\xi_0 \in He$ and $\xi_1 \in HxHe = Hx$, so by induction $\xi_n \in (Hx)^n = Hx^n$, for all $n \in \mathbb{N}$. Let $n \in \mathbb{N}$. Let $s \in G\backslash Hx^n$: $p^{(n)}(e,s) = 0$. Hence P is *not* ergodic.

Assume now $\Sigma \not\subset K$ a proper subgroup of G and $\Sigma \not\subset Hx$ for any coset of any proper normal subgroup $H \lhd G$.

Clearly the inclusions $\Sigma \subset \langle \Sigma \rangle \subset G$ hold with $\langle \Sigma \rangle$ a subgroup of G. By assumption Σ does not lie in a proper subgroup hence $\langle \Sigma \rangle = G$. Hence for all $s, t \in G$, there exists $n(s,t) \in \mathbb{N}$ such that $p^{(n(s,t))}(s,t) > 0$.

Let $L_\Sigma(e) := \{(\sigma_{i_1}, \ldots, \sigma_{i_N}) : e = \sigma_{i_1} \cdots \sigma_{i_N} \,; \sigma_{i_m} \ldots \sigma_{i_n} \neq e \,, n - m < N - 1 \,; \sigma_{i_j} \in \Sigma\}$ be the set of all distinct minimal Σ-presentations of e.

Claim 1: If $|L_\Sigma(e)| = 1$, $G = \mathbb{Z}_{|G|}$ and Σ is in a coset of a proper normal subgroup.

Proof. If $|L_\Sigma(e)| = 1$ there is only one minimal Σ-presentations of e. But $\sigma_1^{o(\sigma_1)}$ and $\sigma_2^{o(\sigma_2)}$ are two distinct minimal Σ-presentations of e. Hence $\sigma_1 = \sigma_2$. Hence $\Sigma = \{\sigma\}$. But $\langle \Sigma \rangle = G$, hence G is cyclic and in particular $\Sigma \subset \{e\}\sigma$ the coset of the proper normal subgroup $\{e\}$ $\bullet$

Claim 2: Assume $|L_\Sigma(e)| > 1$. If Σ is not contained in a coset of a proper normal subgroup of G, then, where L is the set of word lengths of the elements of $L_\Sigma(e)$, $\gcd L = 1$.

Proof. Suppose $\gcd L = k > 1$. Then every Σ-presentation of e has length 0 mod k. Let $N_k \subset G$ be the subgroup generated by all elements of G with a length 0 mod k Σ-presentation. Clearly $e \in N_k$. Let $t \in G$. Suppose t has a length p mod k Σ-presentation. Then t^{-1} has a length $-p$ mod k Σ-presentation since $t^{-1}t = e$ has length 0 mod k Σ-presentation. Let $n \in N_k$. By definition, n has a length 0 mod k Σ-presentation and so $t^{-1}nt$ has a length 0 mod k Σ-presentation. So N_k is normal.

Let $\sigma \in \Sigma$ and suppose $\sigma \in N_k$. Then

$$\sigma\sigma^{-1} = e = (\sigma_{i_1} \cdots \sigma_{i_{qk}})(\sigma_{j_1} \cdots \sigma_{j_{lk-1}})$$

that is e would have a length -1 mod k Σ-presentation, which is not allowed. Hence $\sigma \notin N_k$, so N_k is a proper normal subgroup of G.

Let $\sigma_1 \in \Sigma$. Then $\sigma\sigma_1^{-1} \in N_k$ for all $\sigma \in \Sigma$ as Σ-presentations of any σ^{-1} have length -1 mod k. Hence $\Sigma \subset N_k\sigma_1$ and this contradicts the assumption on Σ. Hence $\gcd L_\Sigma(e) = 1$ $\bullet$

Let S be the set of lengths of all[4] distinct Σ-presentations of e. As $L \subset S$, $\gcd S = 1$. Hence there exist $l_1, \ldots, l_m \in S$, $k_i \in \mathbb{Z}$ such that [22]:

$$k_1 l_1 + \cdots + k_m l_m = 1 \tag{1.6}$$

Let $l \in S$ and $n(e, s)$ as above.

Let

$$M = l_1 |k_1| + \cdots + l_m |k_m| \tag{1.7}$$

and

$$n_0(e, s) = lM + n(e, s) \tag{1.8}$$

If $n \geq n_0(e, s)$, and letting

$$r = \left\lfloor \frac{n - n(e, s)}{l} \right\rfloor , \text{ and}$$

$$n = n(e, s) + rl + a$$

where $0 \leq a < l$ and $r \geq M$. Now as

$$\sum_{i=1}^{m} k_i l_i = 1 , \text{ and } \sum_{i=1}^{m} l_i |k_i| = M,$$

n may be written

$$n = n(e, s) + rl \underbrace{- lM + l \left(\sum_{i=1}^{m} l_i |k_i| \right)}_{=0} + a \left(\sum_{i=1}^{m} k_i l_i \right)$$

$$= (r - M)l + \sum_{i=1}^{m} (l|k_i| + ak_i) l_i + n(e, s)$$

where the $(l|k_i| + ak_i) \geq 0$. Let $x, y, \lambda \in \mathbb{N}$. Note that the probability of going from s to t in $x + \lambda y$ steps is certainly greater than going from s to t in x steps and returning to t every y steps λ times:

$$p^{(x+\lambda y)}(s, t) \geq p^{(x)}(s, t) \left(p^{(y)}(t, t) \right)^{\lambda} \tag{1.9}$$

Hence as $l, l_i \in S$ (so that $p^{(l)}(e, e) > 0$) and $p^{(n(e,s))}(e, s) > 0$;

$$p^{(n)}(e, s) \geq \left(p^{(l)}(e, e) \right)^{r-M} \left[\prod_{i=1}^{m} \left(p^{(l_i)}(e, e) \right)^{l|k_i|+ak_i} \right] p^{(n(e,s))}(e, s) > 0$$

---

[4]not just minimal presentations

Now let n_0 be the maximum of $n_0(e, s)$ as s runs over G. Let $s, t \in G$. By right invariance

$$p^{(n)}(s, t) = p^{(n)}(e, ts^{-1}) > 0 \text{ , for } n > n_0$$

Hence P is ergodic $\bullet$

Chapter 2

Distance to Random

2.1 Introduction

The previous chapter demonstrates that under mild conditions a random walk on a group converges to the random distribution. Therefore, initially the walk is 'far' from random and eventually the walk is 'close' to random. An appropriate question therefore, is given a control $\varepsilon > 0$, how large should k be so that the walk is ε-close to random after k steps? The first problem here is to have a measure of 'close to random'. This chapter introduces a few measures of 'closeness to random', discusses the relationship between them and presents some bounds. In the rest of the work, all walks are assumed ergodic unless stated otherwise.

Let ν and $\mu \in M_p(G)$. The *convolution* of ν and μ is the probability

$$\nu \star \mu(s) := \sum_{t \in G} \nu(st^{-1})\mu(t). \tag{2.1}$$

In particular denote $\nu^{\star n+1} := \nu \star \nu^{\star n}$. The distribution of a random walk after one step is given by ν. If $s \in G$, then the walk can go to s in two steps by going to *some* $t \in G$ after one step and going from there to s in the next. The probability of going from t to s is given by the probability of choosing st^{-1}, i.e. $\nu(st^{-1})$. By summing over all intermediate steps $t \in G$, and noting that $\nu \star \delta^e = \nu$, it is seen that if $\{\xi_k\}_{k=0}^n$ is a random walk on G driven by ν, then $\nu^{\star k}$ is the probability distribution of ξ_k. In terms of the stochastic operator induced by $\nu \in M_p(G)$, P, given any $\mu \in M_p(G)$, $\mu P = \nu \star \mu$.

2.2   Measures of Randomness

The preceding remarks indicate that $\nu^{\star k} \to \pi$ thus a measure of closeness to random can be defined by defining a metric on $M_p(G)$ or putting a norm on $\mathbb{R}^{|G|} \supseteq M_p(G)$. Then a precise mathematical question may be asked: given $\varepsilon > 0$, how large should k be so that $\|\nu^{\star k} - \pi\| < \varepsilon$ or $d(\nu^{\star k}, \pi) < \varepsilon$? Straightaway it is clear that any of the p-norms may be used. Also multiples of p-norms may be used, for example, Diaconis & Saloff-Coste [15] introduce the distance $d_p(k) := |G|^{1-1/p}\|\nu^{\star k} - \pi\|_p$.

Another notion of closeness to random, although not a metric, is that of *separation distance*:

$$s(k) := |G| \max_{t \in G} \left\{ \frac{1}{|G|} - \nu^{\star k}(t) \right\} \tag{2.2}$$

Clearly $s(k) \in [0,1]$ with $s(k) = 1$ if and only if $\nu^{\star k}(g) = 0$ for some g; and $s(k) = 0$ if and only if $\nu^{\star k} = \pi$. The separation distance is submultiplicative in the sense that $s(k+l) \le s(k)s(l)$, for $k, l \in \mathbb{N}$ [4]. This immediately implies that $s(nk) \le [s(k)]^n$. Suppose however that $\nu^{\star k}(g) = 0$ for some $g \in G$. Then $s(k) = 1$ and $s(nk) \le 1$ which is useless. However because the walk is ergodic there exists a time n_0 when $\nu^{\star k}$ is supported on the entire group. Let $L := \min\{\nu^{\star n_o}(s) : s \in G\}$. Then $s(n_0) = (1 - |G|L)$, thence $s(kn_0) \le (1 - |G|L)^k$. An example where this bound is easily applied is the simple walk on $\mathbb{Z}_n$, n odd, where $\nu(\pm 1) = 1/2$. Then $n_0 = n - 1$, $L = 2^{1-n}$ and thence $s(k(n-1)) \le (1 - n.2^{1-n})^k$.

A further measure of randomness is that of the average Shannon Entropy of the distribution; $H(\mu) = \sum_t \mu(t) \log\left(1/\mu(t)\right)$. A quick calculation shows that $H(\delta^e) = 0$, $H(\pi) = \log|G|$; and also that $H(\nu^{\star k})$ increases to $\log|G|$ monotonically [11]. Therefore $\sigma(k) := \log|G| - H(\nu^{\star k})$ is a measure of closeness to random. A lower bound, adapted from [2], is $\sigma(k) \ge (1 - k)\log|G| + k\sigma(1)$.

The default measure of closeness to random in this work, however, is *variation distance*. If $\mu, \nu \in M_p(G)$, their *variation distance* is

$$\|\mu - \nu\| := \max_{A \subset G} |\mu(A) - \nu(A)| \tag{2.3}$$

Diaconis [12] notes an interpretation of variation distance of Paul Switzer. Consider $\mu, \nu \in M_p(G)$. Given a single observation of G, sampled from μ or ν with probability $1/2$, guess whether the observation, o, was sampled from μ or ν. The classical strategy presented here gives the probability of being correct as $1/2(1 + \|\mu - \nu\|)$:

1. Evaluate $\mu(o)$ and $\nu(o)$.

2. If $\mu(o) \geq \nu(o)$, choose μ.

3. If $\nu(o) > \mu(o)$, choose ν.

To see this is true, let $\{\mu > \nu\}$ be the set $\{t \in G : \mu(t) > \nu(t)\}$. Suppose o is sampled from μ. Then the strategy is correct if $o \in \{\mu = \nu\}$ or $o \in \{\mu > \nu\}$:

$$\mathbb{P}[\text{guessing correctly} \,|\, \mu] = \mathbb{P}[o \in \{\mu = \nu\} \,|\, \mu] + \mathbb{P}[o \in \{\mu > \nu\} \,|\, \mu]$$

with a similar expression for $\mathbb{P}[\text{guessing correctly} \,|\, \nu]$. Note that $\mathbb{P}[o \in \{\mu = \nu\}] = \mu(\{\mu = \nu\}) = \nu(\{\mu = \nu\})$ and also $\mathbb{P}[o \in \{\mu > \nu\} \,|\, \mu] = \mu(\{\mu > \nu\})$ (and similar for $o \in \{\mu < \nu\}$). Thus

$$\mathbb{P}[\text{guessing correctly}] = \frac{1}{2}\mathbb{P}[\text{guessing correctly} \,|\, \mu] + \frac{1}{2}\mathbb{P}[\text{guessing correctly} \,|\, \nu]$$

$$= \frac{1}{2}\left(\nu(\{\mu = \nu\}) + \mu(\{\mu > \nu\})\right) + \frac{1}{2}\left(\nu(\{\mu < \nu\})\right)$$

It is easily shown that

$$\|\mu - \nu\| = \mu(\{\mu > \nu\}) - \nu(\{\mu > \nu\}).$$

Hence

$$\mathbb{P}[\text{guessing correctly}] = \frac{1}{2}\left(\underbrace{\nu(\{\mu = \nu\}) + \nu(\{\mu > \nu\}) + \nu(\{\mu < \nu\})}_{=1} + \|\mu - \nu\|\right).$$

Also the separation distance controls the variation distance as

$$\|\nu^{\star k} - \pi\| = \sum_{t \in \{\nu^{\star k} < \pi\}} \left(\frac{1}{|G|} - \nu^{\star k}(t)\right) \leq s(k).$$

It is a straightforward exercise, however, to show that $\|\mu - \nu\|$ is simply half of the usual l^1-distance $\|\mu - \nu\|_1$. Hence, with P doubly stochastic ($\|P\|_{l^1 \to l^1} = 1$ as column sums are 1), the quick calculation

$$\|\nu^{\star k+1} - \pi\|_1 = \|(\nu^{\star k} - \pi)P\|_1 \leq \|\nu^{\star k} - \pi\|_1 \|P\|_{l^1 \to l^1} = \|\nu^{\star k} - \pi\|_1$$

shows that $\|\nu^{\star k} - \pi\|$ is decreasing in k.

At this juncture Aldous [2] denotes by $\tau(\varepsilon)$ the time to get ε-close to random: $\min\{k : \|\nu^{\star k} - \pi\| < \varepsilon\}$. Call $\tau := \tau(1/2e)$ the *mixing time*. The reason the random walk driven by $\nu \in M_p(G)$ is defined to start deterministically at e is because due to right-invariance a random walk driven by the same measure starting deterministically at $g \neq e$ will converge to random at the same rate. Also, if ξ_0 is distributed as $\theta = \sum_t a_t \delta^t$, then the walk looks like $\bigoplus_t a_t \xi^t$ where ξ^t is the walk which begins deterministically at t. All these constituent walks converge at the same rate, however, as might be expected:

$$\|\theta P^k - \pi\| = \frac{1}{2}\sum_{s \in G}\left|\left(\sum_{t \in G} a_t \delta^t P^k(s)\right) - \pi(s)\right| = \frac{1}{2}\sum_{s \in G}\left|\sum_{t \in G} a_t\left(\delta^t P^k(s) - \pi(s)\right)\right|$$

$$\leq \frac{1}{2}\sum_{s \in G}\sum_{t \in G} a_t|\delta^t P^k(s) - \pi(s)| = \sum_{t \in G} a_t\left(\frac{1}{2}\sum_{s \in G}|\delta^t P^k(s) - \pi(s)|\right)$$

$$\leq \|\nu^{\star k} - \pi\| \tag{2.4}$$

Certainly there is equality if θ is a Dirac measure or the random distribution, π.

2.3   Spectral Analysis

In the case of reversible random walks, where π is the random distribution, $\pi(g)p(g,h) = p(h,g)\pi(h)$. Hence the driving probability is symmetric:

$$p(g,h) = p(h,g) \Leftrightarrow \nu(hg^{-1}) = \nu(gh^{-1}) \Leftrightarrow \nu(s) = \nu(s^{-1})\,,\ \forall\, s \in G$$

Also in the $\{\delta_t : t \in G\}$ basis the matrix representation of the stochastic operator is symmetric: $p(x,y) = p(y,x)$. Let $(\,|\,)$ be the inner product on $F(G)$:

$$(\phi|\psi) := \frac{1}{|G|}\sum_{s \in G}\phi(s)\psi(s)^{\star}$$

When the walk is reversible:

$$(P\phi|\psi) = \frac{1}{|G|} \sum_{s \in G} \left(\sum_{t \in G} p(s,t)\phi(t) \right) \psi(t)^{\star}$$

$$= \frac{1}{|G|} \sum_{t \in G} \phi(t) \left(\sum_{s \in G} p(t,s)\psi(s) \right)^{\star} = (\phi|P\psi),$$

and so the stochastic operator is self-adjoint. By the spectral theorem for self-adjoint maps P has an (left) eigenbasis $\mathcal{B} = \{u_1, \ldots, u_{|G|}\}$. Suppose further that $\mathcal{B}$ is normalised such that $\delta^e = \sum a_t u_t$ with $u_1 = \pi$ and $a_1 = 1$ (in fact for any $\theta \in M_p(G)$ this normalisation is unique. Let $v \in \mathbb{R}^n$. Call the sum of the entries of v its *weight*. The eigenvectors u_t, $t \neq 1$, are orthogonal to π. Thence these eigenvectors have weight 0 so in order for the linear combination to be a probability distribution the weight needs to be 1, hence a_1 must be 1.). If P is ergodic, then the eigenvalue 1 has multiplicity 1. A quick calculation shows that if $\lambda_1 = 1$, then also $|\lambda_t| \leq 1$, for all $t \neq 1$. Using an elegant graph-theoretic argument, Ceccherini-Silberstein *et al* [7] show that if P is ergodic then -1 is not an eigenvalue. Therefore in the case of reversible walks (real eigenvalues), $|\lambda_t| < 1$, for all $t \neq 1$ (this is also a consequence of the Perron-Frobenius Theorem), and then

$$\nu^{\star k} = \delta^e P^k = \pi + \sum_{t \neq 1} a_t \lambda_t^k u_t \tag{2.5}$$

Therefore, letting $\lambda_\star := \max\{|\lambda_t| : t \neq 1\}$;

$$\|\nu^{\star k} - \pi\| = \frac{1}{2} \left| \sum_{t \neq 1} a_t \lambda_t^k u_t \right| = \frac{1}{2} \sum_{s \in G} \left| \sum_{t \neq 1} a_t \lambda_t^k u_t(s) \right|$$

$$\leq \frac{1}{2} \sum_{s \in G} \sum_{t \neq 1} |a_t| |\lambda_t|^k |u_t(s)|$$

$$\leq \lambda_\star^k \underbrace{\frac{1}{2} \sum_{s \in G} \sum_{t \neq 1} |a_t| |u_t(s)|}_{=C} = C\lambda_\star^k$$

Hence the rate of convergence is controlled by the second highest eigenvalue in magnitude. In Corollary 2.3.3 an explicit C is given. The importance of the second largest eigenvalue is a mantra in Markov chain theory, however it is only in the reversible case that the importance is so obvious.

Suppose now that P is a not-necessarily-reversible stochastic operator. Following [29], put P in Jordan normal form:

$$P = \begin{pmatrix} 1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_m \end{pmatrix}$$

where the Jordan blocks J_i have form:

$$J_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ 0 & \lambda_i & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_i \end{pmatrix}$$

and have size equal to the algebraic multiplicity of λ_i. Note the first entry of P will be just 1 as 1 is an eigenvalue of multiplicity 1. The Jordan block J_i is the sum of the diagonal matrix $\lambda_i I$ and the superdiagonal, and thus nilpotent, matrix N_i. With $P^n = \text{diag}(1, J_1^n, \ldots, J_m^n)$, and noting $N_i^{d_i} = 0$ where d_i is the multiplicity of λ_i;

$$J_i^k = (\lambda_i I + N_i)^k = \sum_{j=0}^{k} \binom{k}{j} \lambda_i^{k-j} N_i^j = \sum_{j=0}^{d_i-1} \binom{k}{j} \lambda_i^{k-j} N_i^j.$$

Now, for $j < d_i$, N_i^j is the matrix with ones on the jth diagonal above the main diagonal. Hence J_i^k is a matrix whose lower diagonal entries are zero and have equal entries along this 'jth diagonal', namely

$$(J_i^k)_j = \binom{k}{j} \lambda_i^{k-j}$$

Hence the magnitude of the entries along the jth diagonal is bounded by (as $|\lambda_i| < 1$):

$$\left| (J_i^k)_j \right| \leq |\lambda_i|^k \binom{k}{j}$$

The remaining manipulations are dependent on the relation of k to d_i. Assuming $k > 2d_i$ for example:

$$\left| (J_i^k)_j \right| \leq |\lambda_i|^k \binom{k}{d_i}$$

In Jordan normal form, P converges to the matrix with 1 in the $(1,1)$ entry and zero elsewhere. Clearly it is the block corresponding to the second largest eigenvalue in magnitude which is the slowest to converge and hence this eigenvalue controls convergence.

Taking the approach of [7], more explicit bounds for the reversible case may be found. If the walk is reversible then P has an (right) orthonormal basis $\mathcal{B} = \{v_t : t \in G\}$ with corresponding eigenvalues $\{\lambda_t : t \in G\}$. Let v_1 be the constant function with value 1 (so that $\lambda_1 = 1$). Put $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{|G|})$. Now

$$Pv_s(g) = \sum_t p(g,t)v_s(t) = v_s(g)\lambda_s \Leftrightarrow PU = U\Lambda$$

where $U = [v_1|\cdots|v_{|G|}]$. From orthonormality

$$(v_s|v_h) = \frac{1}{|G|}\sum_t v_s(t)v_h(t) = \delta_s(h) \Leftrightarrow U^T U = |G|I$$

As a matrix of eigenvectors, U is invertible with $U^{-1} = U^T/|G|$. Hence $P = U\Lambda U^T/|G|$, and so:

$$P^k = \frac{1}{|G|^k} U\Lambda \underbrace{U^T U}_{=|G|} \underbrace{\Lambda U^T \cdots \Lambda U^T}_{k \text{ copies}} = U\Lambda^k U^T/|G|$$

Or, in terms of coordinates,

$$p^{(k)}(g,h) = \frac{1}{|G|}\sum_{t\in G} v_t(g)\lambda_t^k v_t(h)$$

2.3.1 Proposition

Suppose ν is symmetric. Then in the notation above

$$\|\nu^{\star k} - \pi\|_2^2 = \frac{1}{|G|}\sum_{t\neq 1}\lambda_t^{2k}v_t(e)^2 \tag{2.6}$$

Proof. By definition

$$\|\nu^{\star k} - \pi\|_2^2 = \sum_{s\in G}(\nu^{\star k}(s) - \pi(s))^2$$

$$= \sum_{s\in G}\left(\sum_{t\neq 1} v_t(e)\lambda_t^k v_t(s)/|G|\right)^2$$

$$= \sum_{t_1,t_2\neq 1} v_{t_1}(e)v_{t_2}(e)\lambda_{t_1}^k\lambda_{t_2}^k \sum_{s\in G} v_{t_1}(s)v_{t_2}(s)/|G|^2$$

27

But $U^T U / |G| = I$; equivalently

$$\sum_{s \in G} v_{t_1}(s) v_{t_2}(s) / |G| = \delta_{t_1}(t_2)$$

and so

$$\|\nu^{\star k} - \pi\|_2^2 = \frac{1}{|G|} \sum_{t \neq 1} v_t(e)^2 \lambda_t^{2k} \quad \bullet$$

2.3.2 Corollary: Upper Bound Lemma

Using the same notation, where $\| \cdot \|$ is the variation distance:

$$\|\nu^{\star k} - \pi\|^2 \leq \frac{1}{4} \sum_{t \neq 1} v_t(e)^2 \lambda_t^{2k} \qquad (2.7)$$

Proof. The proof is a rudimentary application of the Cauchy-Schwarz Inequality:

$$
\begin{aligned}
\|\nu^{\star k} - \pi\|^2 &= \frac{1}{4} \|\nu^{\star k} - \pi\|_1^2 \\
&= \frac{1}{4} \left(\sum_{t \in G} \left\{ |\nu^{\star k}(t) - \pi(t)| \sqrt{|G|} \right\} \cdot \frac{1}{\sqrt{|G|}} \right)^2 \\
&\leq \frac{1}{4} \underbrace{\left(\sum_{t \in G} |\nu^{\star k}(t) - \pi(t)|^2 |G| \right)}_{= |G| \|\nu^{\star k} - \pi\|_2^2} \left(\sum_{t \in G} \frac{1}{|G|} \right) \quad \bullet
\end{aligned}
$$

2.3.3 Corollary

In the same notation:

$$\|\nu^{\star k} - \pi\|^2 \leq \frac{|G| - 1}{4} (\lambda_\star)^{2k} \qquad (2.8)$$

Proof. Since $|\lambda_t| \leq \lambda_\star$ for all $t \neq 1$,

$$\|\nu^{\star k} - \pi\|^2 \leq \frac{1}{4} \sum_{t \neq 1} (\lambda_t)^{2k} v_t(e)^2 \leq \frac{(\lambda_\star)^{2k}}{4} \sum_{t \neq 1} v_t(e)^2$$

Note that the eigenvectors of symmetric matrices can be chosen to be real-valued [1], so that $v_t \overline{v_t} = v_t^2$. Also $UU^T = |G|I$ and hence $U^T U = |G|I$ thus

$$\sum_{t \in G} v_t(e)^2 = |G|$$

$$v_1(e)^2 + \sum_{t \neq 1} v_t(e)^2 = 1 + \sum_{t \neq 1} v_t(e)^2 = |G| \; \bullet$$

When ν is symmetric, the associated stochastic operator, P, is symmetric and hence has real eigenvalues which can be ordered $1 = \lambda_1 > \lambda_1 \geq \cdots \geq \lambda_{|G|} > -1$. So now $\lambda_\star = |\lambda_2|$ or $|\lambda_{|G|}|$. Of course, if the spectrum of P can be calculated then these bounds are immediately applicable, however more often one must do with estimates. Diaconis and Saloff-Coste [15] has many examples. Lemma 1 in that paper is a standard result in the field and is proved by consideration of the probability $\nu' = (\nu - \nu(e)\delta^e)/(1 - \nu(e))$ however a quick application of Gershgorin's circle theorem [24] shows the $\lambda_{|G|} \geq -1 + 2\nu(e)$ result also. As the Gershgorin result is mentioned in the sequel, and not typically used by the random walk community, it is presented here:

2.3.4   Gershgorin's Circle Theorem

Let A be a complex $n \times n$ matrix with entries a_{ij}. Let $R_i = \sum_{j \neq i} |a_{ij}|$ be the sum of the absolute values of the entries in the ith row, excluding the diagonal element. If $B[a_{ii}, R_i]$ is the closed disc centered at a_{ii} with radius R_i, then each of the eigenvalues of A is contained in at least one of the $B[a_{ii}, R_i]$.

Proof. Let λ be an eigenvalue of A with eigenvector v. Let $|v(k)| = \max_j |v(j)|$. Now

$$Av(k) = \sum_{j=1}^n a_{kj} v(j) = \lambda v(k).$$

That is

$$\sum_{j \neq k} a_{kj} v(j) = \lambda v(k) - a_{kk} v(k).$$

Divide both sides by $v(k)$;

$$\lambda - a_{kk} = \frac{\sum_{j \neq k} a_{kj} v(j)}{v(k)}$$

Now as $|v(j)| \leq |v(k)|$,

$$\left| \frac{\sum_{j \neq k} a_{kj} v(j)}{v(k)} \right| \leq \sum_{j \neq k} |a_{kj}| \left| \frac{v(j)}{v(k)} \right| \leq \sum_{j \neq k} |a_{kj}| = R_k.$$

In other words $|\lambda - a_{kk}| \leq R_k$ ●

Note that the diagonal entries of a stochastic operator driven by $\nu \in M_p(G)$ are all $\nu(e)$. Hence the radii, R_t, are all equal to $1 - \nu(e)$. The theorem says, for any eigenvalue of the stochastic operator, λ, $|\lambda - \nu(e)| \leq 1 - \nu(e)$. Note Tr $P = |G|\nu(e)$. If P is put in Jordan form, since the trace is basis independent, it is found that Tr $P = \sum_t \lambda_t$. Hence the average of the eigenvalues[1] is equal to $\nu(e)$. Therefore, as 1 is an eigenvalue, there are eigenvalues less than $\nu(e)$, i.e. $\lambda \leq \nu(e)$. In the symmetric case, therefore, $\lambda_{|G|} - \nu(e) \leq 0$ so that $-\lambda_{|G|} + \nu(e) \leq 1 - \nu(e)$, thus $\lambda_{|G|} \geq -1 + 2\nu(e)$. In the general, not-necessarily-symmetric case, the eigenvalues are not necessarily real. However, with $|\lambda - \nu(e)| \leq 1 - \nu(e)$, if $\nu(e) > 1/2$, then the eigenvalues are bounded away from zero so that the stochastic operator, P, is invertible.

2.4   Comparison Techniques

Whilst some random walks yield easily to analysis, others do not. There are a number of techniques, due to Diaconis & Saloff-Coste [15], however, that allow comparison with a simpler walk. Often the continuous analogue of a discrete random walk yields readily to analysis. Diaconis & Saloff-Coste [15] present, in the symmetric case, the most general relationship between the discrete and continuous time version of a given random walk. This paper also uses Dirichlet forms and the Courant minimax principle to estimate eigenvalues on a complicated walk from a simpler version.

---

[1]it would be interesting to try to apply this to obtain a bound for $\|\nu^{\star k} - \pi\|$

2.5 Lower Bounds

The definition of variation distance immediately gives a technique for generating lower bounds. Given a test set $B \subset G$, immediately $\|\nu^{\star k} - \pi\| \geq |\nu^{\star k}(B) - \pi(B)|$. A very simple application uses the fact that $|\mathrm{supp}(\nu^{\star k})| \leq |\Sigma|^k$. Let $A_k \subset G$ be the set where $\nu^{\star k}$ vanishes. Clearly

$$|\nu^{\star k}(A_k) - \pi(A_k)| = \pi(A_k) \geq \frac{1}{|G|}(|G| - |\Sigma|^k) = 1 - \frac{|\Sigma|^k}{|G|}.$$

Another elementary method for generating a lower bound using a test function is apparent via

$$\|\mu - \nu\| = \frac{1}{2} \max_{\|\phi\| \leq 1} \left| \sum_{t \in G} (\mu(t) - \nu(t))\phi(t) \right|. \tag{2.9}$$

The discussion in Section 6.2 implies that if the (right) eigenvector v_s is normalised to have $\|v_s\|_\infty = 1$, then v_s will have expectation zero under the random distribution, $\sum_t \pi(t)v_s(t) = 0$.

2.5.1 Proposition

Let u_λ be a real left eigenvector with eigenvalue $\lambda \neq 1$ and normalised such that $\pi + u_\lambda \in M_p(G)$. Then

$$\|\nu^{\star k} - \pi\| \geq \frac{1}{2}\|u_\lambda\|_1 |\lambda|^k$$

Proof. Let $\theta = \pi + u_\lambda$. Using the fact that a non-Dirac initial distribution θ converges faster than any Dirac measure (see (2.4)), it is clear that $\|\theta P^k - \pi\|$ is a lower bound for $\|\nu^{\star k} - \pi\|$;

$$\|\theta P^k - \pi\| = \|\pi + \lambda^k u_\lambda - \pi\| = \frac{1}{2}\|u_\lambda\|_1 |\lambda|^k \quad \bullet$$

2.6 Volume & Diameter Bounds

By elegant analysis of the properties of the geometry of the random walk, bounds may be put on the eigenvalues of P and applications of the bounds of this Chapter give bounds on the variation distance. The geometry of the random walk is determined by its *Cayley graph*. Suppose that ξ is a random walk on G with driving probability supported on a generating set Σ. The Cayley graph of the random walk is a directed graph with vertex set identified with G. For any $g \in G$, $\sigma \in \Sigma$, the vertices corresponding to the elements g and σg are joined by a directed edge. Thus the edge set consists of pairs of the form $(g, \sigma g)$. The *growth function* of the random walk is $V(k) := |\Sigma^k|$ and the *diameter* of ξ, Δ, is the minimum k such $V(k) = |G|$. Say a random walk has (A, d) *moderate growth* if

$$\frac{V(k)}{V(\Delta)} \geq \frac{1}{A}\left(\frac{k}{\Delta}\right)^d, \quad 1 \leq k \leq \Delta. \tag{2.10}$$

The following theorem appears in Diaconis & Saloff-Coste [16]. The proof — via the heavy machinery of path analysis, flows, two particular quadratic forms and some functional analysis — is omitted. More details are to be found in [15]. First an attractive lemma:

2.6.1 Lemma

Let ξ be a symmetric random walk with diameter Δ. Let $L := \min\{\nu(s) : s \in \Sigma\}$. Then, where λ_2 is the second largest eigenvalue[2]

$$\lambda_2 \leq 1 - \frac{L}{\Delta^2} \tag{2.11}$$

---

[2] *i.e. not necessarily $\lambda_\star$*

2.6.2 Theorem

Let ξ be a symmetric random walk with (A, d) moderate growth. Then for $k = (1 + c)\Delta^2/L$, with $c > 0$:

$$\|\nu^{\star k} - \pi\| \leq Be^{-c} \tag{2.12}$$

where $B = 2^{d(d+3)/4}\sqrt{A}$.
Conversely, for $k = c\Delta^2/(2^{4d+2}A^2)$:

$$\|\nu^{\star k} - \pi\| \geq \frac{1}{2}e^{-c} \tag{2.13}$$

Example: The Heisenberg Group

Consider the set of matrices:

$$H_3(n) = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \tag{2.14}$$

where a, b, $c \in \mathbb{Z}_n$. With matrix multiplication modulo n, $H_3(n)$ forms a group of order n^3. The random walk driven by the measure $\nu_n \in M_p(H_3(n))$ constant on the matrices $(a, b, c) = (\pm 1, 0, 0)$, $(0, 0, \pm 1)$, $(0, 0, 0)$ is ergodic. Diaconis & Saloff-Coste [16] have shown that the random walk has diameter $n - 1 \leq \Delta \leq n + 1$ and volume growth function $V(k) \geq k^3/6$ ($1 \leq k \leq n+1$). With order[3] $|H_3(n)| \leq 8\Delta^3$,

$$\frac{V(k)}{V(\Delta)} \geq \frac{k^3/6}{8\Delta^3} = \frac{1}{48}\left(\frac{k}{\Delta}\right)^3 \text{, for } 1 \leq k \leq \Delta,$$

the random walk has $(48, 3)$ moderate growth.
Precise application of Theorem 2.6.2 yields for constants A, A', B, B':

$$A'e^{-B'k/n^2} \leq \|\nu_n^{\star k} - \pi\| \leq Ae^{-Bk/n^2} \tag{2.15}$$

Hence order n^2 steps are necessary for convergence to random.

---

[3] $n^3 \leq 8(n-1)^3 \leq 8\Delta^3$ for $n \geq 4$.

Chapter 3

Diaconis-Fourier Theory

Much of the precluding analysis passes neatly into the case of the classical Markov theory for a random walk on a finite set X. It has been seen that this analysis culminates in the result that the rate of convergence of to a stationary state is related heavily to the second largest eigenvalue of the stochastic operator. As a rule the calculation of the second highest eigenvalue is too cumbersome for larger groups and further the bound is not particularly sharp due to the information loss in disregarding the rest of the spectrum of the stochastic operator.

In his seminal monograph [12], Diaconis utilises the group structure to produce bounds for rates of convergence. He uses Fourier methods and representation theory to produce bounds that are invariably sharper as the entire spectrum is utilised. This chapter follows his approach.

3.1   Basics of Representations and Characters

A *representation* ρ of a finite group G is a group homomorphism from G into $GL(V)$ for some vector space V. The dimension of the vector space[1] is called the *dimension of* ρ and is denoted by d_ρ. If W is a subspace of V invariant under $\rho(G)$, then $\rho_{|W}$ is called a *subrepresentation*.

---

[1]at this point the underlying vector space may be infinite dimensional but later it will be seen that the only representations of any interest are of finite dimension. Also the underlying field is unspecified at this point but later it will be seen that the only representations of any interest will be over complex vector spaces.

If $\langle .,. \rangle$ is an inner product on V, $\langle u, v \rangle_\rho = \sum_t \langle \rho(t)u, \rho(t)v \rangle$ defines another, and further the orthogonal complement of W with respect to $\langle .,. \rangle_\rho$, $W^\perp$, is also invariant under ρ. Hence, every representation splits into a direct sum of subrepresentations. Both $\{\mathbf{0}\}$ and V itself yield trivial subrepresentations. A representation ρ that admits no non-trivial subrepresentations is called *irreducible*. An example of an irreducible representation is the *trivial representation*, τ, which maps G to 1: $\rho(s)z = z$, $z \in \mathbb{C}$. Inductively, therefore, every representation is a direct sum of irreducible representations. A quick calculation shows $\langle \rho(s)u, \rho(s)v \rangle_\rho = \langle u, v \rangle_\rho$, hence $\|u\|_\rho = \|\rho(s)v\|_\rho$ so the operators $\rho(s)$ are isometries and are thus unitary. Two representations, ρ acting on V and ϱ acting on W; are *equivalent as representations*, $\rho \equiv \varrho$, if there is a bijective linear map $f \in L(V, W)$ such that $\varrho \circ f = f \circ \rho$. In this context f is said to *intertwine* ϱ and ρ.

Example: A Two Dimensional Representation of the Dihedral Group

The dihedral group D_4, the group of symmetries of the square, admits a natural representation ρ. The elements of D_4 are the rotations r_0, $r_{\pi/2}$, r_π, $r_{3\pi/2}$ and reflections (12), (13), (14), (23). If the vertices of the square are inscribed in a unit circle at the poles[2] then $\rho(r_\theta)$ are the rotation matrices:

$$\rho(r_\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Similarly the reflections have action as reflection in $y = x$, $y = -x$, $y = 0$ and $x = 0$ which have matrix representations:

$$\rho((12)) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \rho((13)) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\rho((14)) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \rho((23)) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

---

[2] i.e. the coordinates $(\pm 1, 0)$, $(0, \pm 1)$.

3.1.1 Schur's Lemma

Let $\rho_1 : G \to GL(V_1)$ and $\rho_2 : G \to GL(V_2)$ be two irreducible representations of G, and let $f \in L(V_1, V_2)$ be an intertwiner. Then

1. *If ρ_1 and ρ_2 are not equivalent $f \equiv 0$.*

2. *If $V_1 =: V := V_2$ is complex, and $\rho_1 := \rho =: \rho_2$, $f = \lambda I$, for some, $\lambda \in \mathbb{C}$.*

Proof. The straightforward calculations $f(\rho_1(G)\ker f) = \rho_2(G)f(\ker f) = 0$ and $\rho_2(G)\text{Im } f = f(\rho_1(G)V_1)$ show that $\ker f$ and $\text{Im } f$ are invariant subspaces. By irreducibility both the kernel and image of f are trivial or the whole space.

1. Suppose $f \not\equiv 0$. Hence $\ker f = \{0\}$ and $\text{Im } f = V_2$ so f is an isomorphism as it is linear. However this would imply that ρ_1 and ρ_2 are equivalent as representations, a contradiction. Thence $f \equiv 0$.

2. If $f \equiv 0$ then $f = 0.I$. Suppose again $f \not\equiv 0$. Then f has a non-zero eigenvalue $\lambda \in \mathbb{C}$ with associated non-zero eigenvector $v_\lambda \neq 0$. Let $f_\lambda = f - \lambda I$. A quick calculation shows that $\rho(G)f_\lambda(V) = f_\lambda(\rho(G)V)$, hence f_λ is an intertwiner. Note that $\ker f_\lambda \neq \{0\}$ as $v_\lambda \in \ker f_\lambda$. Thence $\ker f_\lambda = V$, that is $f_\lambda \equiv 0$, which implies $f = \lambda I$ $\bullet$

Let $\rho_1 : G \to GL(V_1)$ and $\rho_2 : G \to GL(V_2)$ be two irreducible representations of G and $h_0 \in L(V_1, V_2)$. Let

$$h = \frac{1}{|G|} \sum_{t \in G} \rho_2^{-1}(t) h_0 \rho_1(t) \tag{3.1}$$

A quick verification shows that h is an intertwiner of ρ_1 and ρ_2, and by recourse to Schur's Lemma $h \equiv 0$ in the case where $\rho_1 \not\equiv \rho_2$, and $h = \lambda I$ when $\rho_1 \equiv \rho_2$. In the case $\rho_1 \equiv \rho_2$, taking traces gives $\lambda = \text{Tr } h/d_\rho$ and a further calculation shows $\text{Tr } h = \text{Tr } h_0$. Suppose ρ_1 and ρ_2 are given in matrix form as $\rho_1(s) = (r^1_{ij}(s))$ and $\rho_2(s) = (r^2_{ij}(s))$. The linear maps h and h_0 are defined by matrices x_{ij} and x^0_{ij}. In particular,

$$x_{ij} = \frac{1}{|G|} \sum_{\substack{t \in G \\ \lambda, \mu}} r^2_{i\lambda}(t^{-1}) x^0_{\lambda\mu} r^1_{\mu j}(t) \tag{3.2}$$

Suppose $\rho_1 \not\equiv \rho_2$ so that $h \equiv 0$ when defined by $h_0 = \delta_{kl}$. In this case $x_{ij} = 0$ and (3.2) collapses to

$$\frac{1}{|G|} \sum_{t \in G} r_{ik}^2(t^{-1}) r_{lj}^1(t) = 0 \ , \ \forall \, i, k, l, j. \tag{3.3}$$

In the case where $\rho_1 \equiv \rho_2$, $h = \lambda I$, where, in matrix elements, $\lambda = \sum_m x_{mm}^0 / d_\rho$. When h is defined by $h_0 = \delta_{kl}$, (3.2) collapses to

$$\frac{1}{|G|} \sum_{t \in G} r_{ik}^2(t^{-1}) r_{lj}^1(t) = \frac{\delta_{ij} \delta_{kl}}{d_\rho}. \tag{3.4}$$

Note again that $\rho(s)$ is a unitary operator so that $\rho(s)^\star = \rho^{-1}(s)$, thence $\overline{r_{ji}(s)} = r_{ij}(s^{-1})$. A simple rearrangement of (3.3) and (3.4) using this fact show that the matrix elements of the irreducible representations are orthogonal in $F(G)$.

If ρ is a representation, the *character* of ρ, $\chi_\rho(s) := \mathrm{Tr} \ \rho(s)$. Using the preceding remarks, it can be shown that the characters of the irreducible representations are orthonormal in $F(G)$. If ρ_1 and ρ_2 are representations with characters χ_1 and χ_2, by choosing a basis so that the matrix of $\rho_1 \oplus \rho_2$ is a block 2×2 matrix with ρ_1 in the $(1,1)$ position and ρ_2 in the $(2,2)$ position, taking traces shows that the character of $\rho_1 \oplus \rho_2$ is $\chi_1 + \chi_2$. Suppose now ρ is a representation with character ϕ that decomposes into a direct sum of irreducible representations $\rho = \rho_1 \oplus \cdots \oplus \rho_k$. If each of the ρ_i have character χ_i, then $\phi = \chi_1 + \cdots + \chi_k$. If ρ' is an irreducible representation with character χ, then $(\phi|\chi) = \sum_i (\chi_i|\chi)$. By orthonormality, $(\chi_i|\chi) = 0$ or 1 as χ_i is, or is not, equivalent to χ. Thence, the number of ρ_i equivalent to ρ' equals $(\phi|\chi)$.

A canonical representation is the *regular representation*; defined with respect to a complex vector space with basis $\{e_s\}$ indexed by $s \in G$ via $r(s)(e_t) := e_{st}$. Observe that the underlying vector space is isomorphic to $F(G)$. It is a simple exercise to show that $\chi_r(e) = |G|$ and zero elsewhere. This implies that for an irreducible representation ρ_i, $(\chi_r|\chi_i) = \chi_i(e)^\star = \overline{\mathrm{Tr} \ I_{d_i}} = d_i$ so that $\chi_r(s) = \sum_i d_i \chi_i(s)$, where the sum is over all irreducible representations. Letting $s = e$ here yields $\sum_i d_i^2 = |G|$. Now it can be seen that the matrix entries of the irreducible representations form an orthogonal basis for $F(G)$ because they are orthogonal and there are $\sum_i d_i^2 = |G|$ of them: $\dim F(G) = |G|$.

3.2 Fourier Theory

Let $f \in F(G)$ and ρ a representation of G. The *Fourier Transform of f at the representation ρ* is the operator $\widehat{f}(\rho) = \sum_s f(s)\rho(s)$. This Fourier transform satisfies an inversion theorem, a Plancherel Formula; and, of course, a Convolution Theorem $\widehat{f \star h}(\rho) = \widehat{f}(\rho)\widehat{h}(\rho)$ whose proof is rudimentary.

3.2.1 Fourier Inversion Theorem

Let $f \in F(G)$, then, where the sum is over irreducible representations,

$$f(s) = \frac{1}{|G|} \sum_i d_i \operatorname{Tr} \left(\rho_i(s^{-1})\widehat{f}(\rho_i)\right) \tag{3.5}$$

Proof. Both sides are linear in f so it is sufficient to check the formula for $f = \delta_t$. Then $\widehat{f}(\rho_i) = \rho_i(t)$, and the right hand side equals

$$\frac{1}{|G|} \sum_i d_i \operatorname{Tr} \left(\rho_i(s^{-1})\rho_i(t)\right) = \frac{1}{|G|} \sum_i d_i \chi_i(s^{-1}t) = \frac{1}{|G|} \chi_r(s^{-1}t)$$

When $s = t$ this equals 1; otherwise it is 0; i.e. it equals δ_t $\bullet$

3.2.2 Plancherel Formula

Let f, $h \in F(G)$, then

$$\sum_{s \in G} f(s^{-1})h(s) = \frac{1}{|G|} \sum_i d_i \operatorname{Tr} \left(\widehat{f}(\rho_i)\widehat{h}(\rho_i)\right) \tag{3.6}$$

Proof. Both sides are linear in f; so consider $f = \delta_t$. Using the Fourier Inversion Theorem

$$h(t^{-1}) = \sum_{s \in G} \delta_t(s^{-1})h(s) = \frac{1}{|G|} \sum_i d_i \operatorname{Tr} \left(\rho_i(t)\widehat{h}(\rho_i)\right)$$

However, $\rho_i(t)$ is nothing but $\widehat{\delta_t}(\rho_i)$ so the formula is verified $\bullet$

In the sequel, mostly elements of $M_p(G)$ viewed as elements of $F(G)$ are considered. Let $\mu \in M_p(G)$ and let $\check{\mu}(s) := \mu(s^{-1})$. After a reindex, $t \mapsto t^{-1}$, $\widehat{\check{\mu}}(\rho) = \sum_t \mu(t)\rho(t^{-1})$. With the unitary nature of the representation, and the fact that $\mu = \overline{\mu}$ as $\mu \in \mathbb{R}$, in fact $\widehat{\check{\mu}}(\rho) = \widehat{\mu}(\rho)^\star$. Hence, for $\nu \in M_p(G)$:

$$\sum_{t \in G} \mu(t)\nu(t) = \frac{1}{|G|} \sum d_i \operatorname{Tr} \left[\widehat{\nu}(\rho_i)\widehat{\mu}(\rho_i)^\star\right] \tag{3.7}$$

With the aid of two quick facts the celebrated Upper Bound Lemma of Diaconis and Shahshahani [12, 17] may be proven. The first of these is the straightforward calculation that for all $\nu \in M_p(G)$, at the trivial representation τ, $\widehat{\nu}(\tau) = \sum_t \nu(t) = 1$. The second comprises a lemma.

3.2.3 Lemma

At a non-trivial irreducible representation, ρ, the Fourier transform of the random distribution, π, vanishes: $\widehat{\pi}(\rho) = \mathbf{0}$.

Proof. First note that $h = \sum_{t \in G} \rho(t)$ is a linear map, invariant under any $\rho(s)$: $\rho(s)h = h = h\rho(s)$. As a consequence both $\ker h$ and $\operatorname{Im} h$ are invariant subspaces. By irreducibility, both the kernel and the image of h are trivial or the whole space. Suppose $\ker h = \{\mathbf{0}\}$ and $\operatorname{Im} h = V$. For any $v \in V$, $\rho(s)hv = hv$. Hit both sides with h^{-1}: $h^{-1}\rho(s)hv = v$. Now use the fact that $\rho(s)$ and h commute to show $\rho(s)v = v$. Hence ρ is trivial. Therefore $\ker h = V$, $\operatorname{Im} h = \{\mathbf{0}\}$, i.e. $h = \mathbf{0}$. Now $\widehat{\pi}(\rho) = \sum_t \pi(t)\rho(t) = h/|G| = \mathbf{0}$ •

3.2.4   Upper Bound Lemma

Let ν be a probability on a finite group G. Then

$$\|\nu - \pi\|^2 \le \frac{1}{4} \sum_i d_i \operatorname{Tr} \left(\widehat{\nu}(\rho_i)\widehat{\nu}(\rho_i)^\star\right), \qquad (3.8)$$

where the sum is over all non-trivial irreducible representations.

Proof. Using the Cauchy-Schwarz Inequality

$$4\|\nu - \pi\|^2 = \left\{\sum_{t \in G} |\nu(t) - \pi(t)|\right\}^2$$

$$\le |G| \sum_{t \in G} |\nu(t) - \pi(t)|^2 = |G| \sum_{t \in G} (\nu - \pi)(t)(\nu - \pi)(t),$$

where of course $\nu - \pi$ is a real function. Thus, by (3.7)

$$4\|\nu - \pi\|^2 \le \cancel{|G|}\frac{1}{\cancel{|G|}} \sum_i d_i \operatorname{Tr} \left[\widehat{(\nu - \pi)}(\rho_i)\widehat{(\nu - \pi)}(\rho_i)^\star\right]$$

Now $\widehat{(\nu - \pi)}(\rho) = \sum_t (\nu(t) - \pi(t))\rho(t) = \widehat{\nu}(\rho) - \widehat{\pi}(\rho)$. With the preceding facts:

$$\widehat{(\nu - \pi)}(\rho) = \begin{cases} \mathbf{0} & \text{if } \rho \text{ is trivial} \\ \widehat{\nu}(\rho) & \text{if } \rho \text{ is non-trivial and irreducible} \end{cases}$$

So therefore

$$4\|\nu - \pi\|^2 \le \sum_i d_i \operatorname{Tr} \left(\widehat{\nu}(\rho_i)\widehat{\nu}(\rho_i)^*\right),$$

where the sum is over all non-trivial representations ●

This bounds are applicable to $\|\nu^{\star k} - u\|$ via the Convolution Theorem: $\widehat{\nu^{\star k}}(\rho) = \widehat{\nu}(\rho)^k$.

3.3 Number of Irreducible Representations

Let G be a group and g, h elements of G. An element $g \in G$ is *conjugate* to h, $g \sim h$, if there exists $t \in G$ such that $h = tgt^{-1}$. Conjugacy is an equivalence relation on a group [22], and hence forms a partition of G into disjoint *conjugacy classes* $G = [s_1]_\sim \cup [s_2]_\sim \cup \cdots \cup [s_r]_\sim$, where

$$[s]_\sim = \{g \in G : \exists t \in G, g = tst^{-1}\} = \{tst^{-1} : t \in G\}. \qquad (3.9)$$

A complex function $f \in F(G)$ is a *class function* if for all conjugacy classes $[s_i]_\sim \subset G$, $f_{|[s_i]_\sim} = \lambda$, for some $\lambda \in \mathbb{C}$. Let $\mathfrak{Cl}(G)$ be the subspace of $F(G)$ consisting of all class functions. The characters of a representation are class functions. Let $f \in \mathfrak{Cl}(G)$ and ρ be an irreducible representation. Note that $\rho(s)\widehat{f}(\rho)\rho(s^{-1}) = \sum_t f(t)\rho(sts^{-1})$, and with a reindexing $t \mapsto s^{-1}ts$, it is clear that f is an intertwiner for ρ. Thus, by Schur's Lemma, $\widehat{f}(\rho) = \lambda I$. Taking traces gives $\lambda = \text{Tr}\ (\widehat{f}(\rho))/d_\rho = \sum_t f(t)\chi(t)/d_\rho = |G|(f|\chi^\star)/d_\rho$.

3.3.1 Theorem

The characters of the irreducible representations $\chi_1, \chi_2, \ldots, \chi_l$ form an orthonormal basis for $\mathfrak{Cl}(G)$.

Proof. Characters are orthonormal class functions. As $\mathfrak{Cl}(G)$ together with $(.|.)$ forms an inner product space, and $\Omega = \text{span}\{\chi_i\}$ is a subspace: $\mathfrak{Cl} = \Omega \oplus \Omega^\perp$. Let $f \in \mathfrak{Cl}$ have the decomposition $f = g + h$, with $g \in \Omega$, $h \in \Omega^\perp$. Therefore for all irreducible representations χ_i: $(h|\chi_i^\star) = 0$. The preceding remarks indicate that $\widehat{h}(\rho) = |G|(h|\chi_\rho^\star)I/d_\rho = \mathbf{0}$. The Fourier Inversion Theorem yields:

$$h(s) = \frac{1}{|G|} \sum_i d_i \text{Tr}\ \left(\rho_i(s^{-1})\widehat{h}(\rho_i)\right) \equiv 0.$$

Hence therefore $\Omega^\perp = \{\mathbf{0}\}$ and the characters of the irreducible representations span $\mathfrak{Cl}(G)$. $\bullet$

3.3.2 Theorem

The number of irreducible representations equals the number of conjugacy classes.

Proof. Theorem 3.3.1 gives the number of irreducible representations, l:

$$l = \dim(\mathfrak{Cl}(G))$$

A class function can be defined to have an arbitrary value on each conjugacy class, so $\dim(\mathfrak{Cl}(G))$ is the number of conjugacy classes ●

As an immediate corollary, all the irreducible representations of an Abelian group G have degree 1. To see this note if G is Abelian, there are $|G|$ conjugacy classes, so $|G|$ terms in the sum $\sum_i d_i^2 = |G|$, each of which must be 1. Hence if G has l conjugacy classes and l representations are found, if the l representations are inequivalent and irreducible, all the irreducible representations have been found.

3.3.3 Theorem

Two irreducible representations with the same character are equivalent.

Proof. Suppose χ_1, χ_2 are identical characters of non-equivalent irreducible representations ρ_1 and ρ_2,

$$(\chi_1|\chi_2) = \frac{1}{|G|} \sum_{t \in G} \chi_1(t)\overline{\chi_1(t)} \underset{\chi_1(e)\neq 0}{>} 0.$$

However the characters of irreducible representations are orthonormal. This is a contradiction; hence $\rho_1 \equiv \rho_2$ ●

3.3.4 Theorem

Let χ be the character of a representation ρ, then ρ is an irreducible representation if and only if $(\chi|\chi) = 1$.

Proof. Clearly if ρ is irreducible $(\chi|\chi) = 1$. Suppose for the converse that $(\chi|\chi) = 1$. Any representation ρ is the direct sum of irreducible representations $\{\rho_i\}$ with character $\chi = \chi_1 + \chi_2 + \cdots + \chi_m$. Therefore if $(\chi|\chi)$ must equal 1, then there exists a unique ρ_k such that $\rho \equiv \rho_k$ ●

Example: The Quaternion Group, $\mathcal{Q}$

Consider the quaternion group $\mathcal{Q} = \{\pm 1, \pm i, \pm j, \pm k\}$ where 1 is the identity. Multiplication in $\mathcal{Q}$ is defined by $(-1)^2 = 1$ and $i^2 = j^2 = k^2 = ijk = -1$, where -1 commutes with everything. The quaternion group has five conjugacy classes $\{1\}$, $\{-1\}$, $\{\pm i\}$, $\{\pm j\}$ and $\{\pm k\}$ and thus five irreducible representations. As $\sum_i d_i^2 = |G|$, the there must be one irreducible representation of degree 2 and four of degree 1. Consider the linear map $\rho : \mathcal{Q} \to GL(\mathbb{C}^2)$ given by:

$$\rho(i) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \qquad \rho(k) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\rho(j) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \qquad \rho(1) = I, \ \rho(-s) = -\rho(s) \tag{3.10}$$

Straightforward calculations show that ρ is a representation. Also $(\chi|\chi) = 1$, and in light of Theorem 3.3.4, ρ is the two dimensional irreducible representation. Let $\tau : \mathcal{Q} \to GL(\mathbb{C})$ be the trivial representation; it is the second irreducible representation. Let $\rho_i : \mathcal{Q} \to GL(\mathbb{C})$ (respectively ρ_j, ρ_k) be defined by:

$$\rho_i(s) := \begin{cases} 1 & \text{if } s \in \langle i \rangle \\ -1 & \text{if } s \notin \langle i \rangle \end{cases} \tag{3.11}$$

This is a one-dimensional representation so is irreducible. It is an easy calculation to show that $\{\tau, \chi_i, \chi_j, \chi_k\}$ is an orthogonal set so comprise four inequivalent representations. Hence the set of irreducible representations of $\mathcal{Q}$ are given by $\{\rho, \tau, \rho_i, \rho_j, \rho_k\}$.

3.4 Simple Walk on the Circle

Consider the walk on $\{\mathbb{Z}_n, \oplus\}$ driven by

$$\nu_n(s) := \begin{cases} \frac{1}{2} & \text{if } s = \pm 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

$\mathbb{Z}_n$ is an Abelian group, so all irreducible representations have degree 1. Any ρ is determined by the image of 1: $\rho(s) = \rho(1^s) = \rho(1)^s$. Also $1^n = 0$, hence $\rho(1)^n = \rho(1^n) = \rho(0) = 1$ so $\rho(1)$ must be a n-th root of unity.

There are n such: $e^{2\pi it/n}$, $t = 0, 1, 2, \ldots, n-1$. Each gives a representation $\rho_t(s) = e^{2\pi its/n}$. Now some results used in the Lower Bound; see Appendix A for proof.

3.4.1 Lemma

The following (in)equalities hold.

1. *For any odd n and $k \in \mathbb{N}$,*

$$\sum_{t=1}^{n-1} \cos^{2k}(2\pi t/n) = 2 \sum_{t=1}^{(n-1)/2} \cos^{2k}(\pi t/n) \tag{3.13}$$

2. *For $x \in [0, \pi/2]$,*

$$\cos x \leq e^{-x^2/2} \tag{3.14}$$

3. *For any $x > 0$*

$$\sum_{j=1}^{\infty} e^{-(j^2-1)x} \leq \sum_{j=0}^{\infty} e^{-3jx} \tag{3.15}$$

4. *For $x \in [0, \pi/6]$,*

$$\cos x \geq e^{-x^2/2-x^4/2} \tag{3.16}$$

3.4.2 Upper and Lower Bounds

For $k \geq n^2/40$, with n odd,

$$\|\nu_n^{\star k} - \pi_n\| \leq e^{-\pi^2 k/2n^2} \tag{3.17}$$

Conversely, for $n \geq 7$, and any k

$$\|\nu_n^{\star k} - \pi_n\| \geq \frac{1}{2} e^{-\pi^2 k/2n^2 - \pi^4 k/2n^4}. \tag{3.18}$$

Proof. The Fourier transform of ν_n at ρ_s is:

$$\widehat{\nu_n}(\rho_s) = \sum_{t=0}^{n-1} \nu_n(t) e^{2\pi ist/n} = \frac{1}{2} e^{2\pi is/n} + \frac{1}{2} e^{-2\pi is/n} = \cos\left(\frac{2\pi s}{n}\right).$$

The Upper Bound Lemma and (3.13) yield

$$\|\nu_n^{\star k} - \pi_n\|^2 \leq \frac{1}{4} \sum_{t=1}^{n-1} \cos^{2k}\left(\frac{2\pi t}{n}\right) = \frac{1}{2} \sum_{t=1}^{(n-1)/2} \cos^{2k}\left(\frac{\pi t}{n}\right).$$

Applying (3.14) yields

$$\|\nu_n^{\star k} - \pi_n\|^2 \leq \frac{1}{2} \sum_{t=1}^{(n-1)/2} e^{-\pi^2 t^2 k/n^2} \leq \frac{1}{2} e^{-\pi^2 k/n^2} \sum_{t=1}^{\infty} e^{-\pi^2(t^2-1)k/n^2},$$

and so with (3.15)

$$\|\nu_n^{\star k} - \pi_n\|^2 \leq \frac{1}{2} e^{-\pi^2 k/n} \sum_{t=0}^{\infty} e^{-3\pi^2 tk/n^2} = \frac{1}{2} \frac{e^{-\pi^2 k/n^2}}{1 - e^{-3\pi^2 k/n^2}}.$$

Now since $k \geq n^2/40$, $2\left(1 - e^{-3\pi^2 k/n^2}\right) > 1$, and it follows that

$$\|\nu_n^{\star k} - \pi_n\| \leq e^{-\pi^2 k/2n^2}$$

For the lower bound, consider the norm 1 function $\phi(s) = \rho_{\bar{s}}(s) = \cos(2\pi s\bar{s}/n)$ where $\bar{s} = (n-1)/2$. By Lemma 3.2.3, $\phi(s)$ has zero expectation under the random distribution. Now an application of (2.9) gives

$$\|\nu_n^{\star k} - \pi_n\| \geq \frac{1}{2} \left|\sum_{t \in G} \nu_n^{\star k}(t)\phi(t)\right| = \frac{1}{2} \left|\widehat{\nu_n^{\star k}}\right| = \frac{1}{2} |\widehat{\nu_n}(\rho_{\bar{s}})|^k$$

Now $\widehat{\nu_n}(\rho_{\bar{s}}) = \cos(2\pi\bar{s}/n) = -\cos(\pi/n)$ by a quick calculation. By (3.16), for $\pi/n \leq \pi/6$:

$$\|\nu_n^{\star k} - \pi_n\| \geq \frac{1}{2} \left|\cos\frac{\pi}{n}\right|^k \geq \frac{1}{2} e^{-\pi^2 k/2n^2 - \pi^4 k/2n^4} \quad \bullet$$

Remark

If n is even then $\{1, -1\}$ lies in the coset of odd numbers of the normal subgroup $\{0, 2, \ldots, n-2\} =: H \lhd \mathbb{Z}_n$, and so the walk is not ergodic by Theorem 1.3.2.
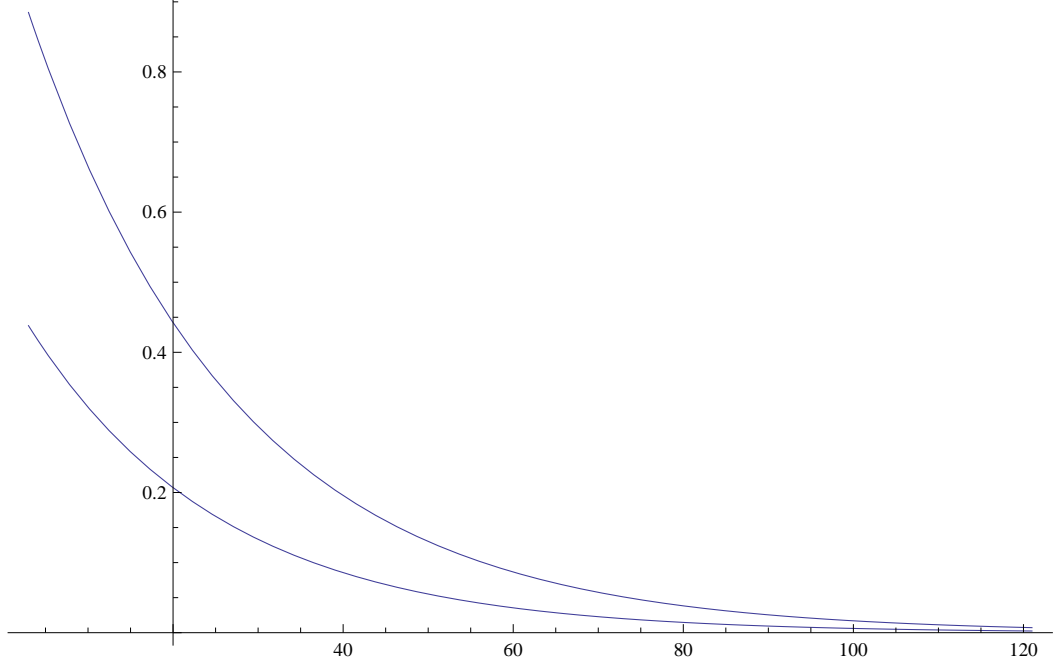

Figure 3.1: A plot of the upper and lower bound for $n = 11$.

3.5 Nearest Neighbour Walk on the n-Cube

Consider the walk on $\{\mathbb{Z}_2^n, \oplus_2^n\}$, $n > 1$, driven by

$$\nu_n(s) := \begin{cases} \frac{1}{n+1} & \text{if } w(s) = 0 \text{ or } 1 \\ \\ 0 & \text{otherwise} \end{cases} \tag{3.19}$$

where $w(s)$, the weight of $s = (s_1, s_2, \ldots, s_n)$, is given by the sum in $\mathbb{N}$:

$$w(s) = \sum_{i=1}^{n} s_i \tag{3.20}$$

$\mathbb{Z}_2^n$ is an Abelian group, so all irreducible representations have degree 1. It is a simple verification to show that each are given by $\rho_t(s) = (-1)^{t \cdot s}$.

46

Now some results used in the Upper Bound; see Appendix A for proof.

3.5.1 Lemma

The following inequalities hold.

1. *If $l \leq n/2$,*

$$\binom{n}{l}\left(1 - \frac{2l}{n+1}\right)^{2k} \geq \binom{n}{n+1-l}\left(1 - \frac{2(n+1-l)}{n+1}\right)^{2k} \tag{3.21}$$

2. *When $a \leq b$,*

$$\binom{a}{b} \leq \frac{a^b}{b!} \tag{3.22}$$

3. *Let $n \in \mathbb{N}$, $c > 0$. If $k = (n+1)(\log n + c)/4$*

$$\left(1 - \frac{2j}{n+1}\right)^{2k} \leq e^{-j\log n - jc} \tag{3.23}$$

3.5.2 Upper Bound

For $k = (n+1)(\log n + c)/4$, $c > 0$:

$$\|\nu_n^{\star k} - \pi_n\|^2 \leq \frac{1}{2}\left(e^{e^{-c}} - 1\right) \tag{3.24}$$

Proof. Let $\{e_i\}$ denote the standard basis[3] of $\mathbb{Z}_2^n$:

$$\widehat{\nu_n}(\rho_s) = \sum_{t \in \mathbb{Z}_2^n}(-1)^{s \cdot t}\nu_n(t) = \frac{1}{n+1}\left[1 + \sum_{i=1}^{n}(-1)^{s \cdot e_i}\right]$$

Now $s \cdot e_i = s_i$ so

$$\begin{aligned}
\widehat{\nu_n}(\rho_s) &= \frac{1}{n+1}\left[1 + \sum_{i=1}^{n}(-1)^{s_i}\right] \\
&= \frac{1}{n+1}\left[1 + \sum_{s_i=1}(-1) + \sum_{s_i=0}(1)\right] \\
&= \frac{1}{n+1}\left[1 + w(s)(-1) + (n - w(s))(1)\right] \\
&= \frac{n+1-2w(s)}{n+1} = 1 - \frac{2w(s)}{n+1}
\end{aligned}$$

---

[3]of the finite vector space $\mathbb{Z}_2^n$ with underlying field $\mathbb{Z}_2$.

Thus Upper Bound Lemma gives (summing over weights on the right equality):

$$\|\nu_n^{\star k} - \pi_n\|^2 \leq \frac{1}{4} \sum_{t \neq 0} \widehat{\nu_n}(\rho_t)^{2k} = \frac{1}{4} \sum_{j=1}^{n} \binom{n}{j} \left(1 - \frac{2j}{n+1}\right)^{2k}. \qquad (3.25)$$

Let $n/2 \leq j \leq n$ such that $j = n + 1 - l$ (i.e. $l \in \{1, 2, \ldots, \lfloor n/2 \rfloor\}$) and consider the $(n + 1 - l)$th (i.e. jth) term in this sum. By (3.21), the lth term dominates this term, and for $l \in \{1, 2, \ldots, \lfloor n/2 \rfloor\}$,

$$\binom{n}{l} \left(1 - \frac{2l}{n+1}\right)^{2k} + \binom{n}{n+1-l} \left(1 - \frac{2(n+1-l)}{n+1}\right)^{2k} \leq 2 \binom{n}{l} \left(1 - \frac{2l}{n+1}\right)^{2k}$$
$$(3.26)$$

Noting that the 'middle' term (i.e. n odd) is unaffected, (3.25) is thus dominated by a sum of $\lceil n/2 \rceil$ terms. Therefore, with (3.22)

$$\|\nu_n^{\star k} - \pi_n\|^2 \leq \frac{1}{2} \sum_{j=1}^{\lceil n/2 \rceil} \frac{n^j}{j!} \left(1 - \frac{2j}{n+1}\right)^{2k}.$$

Applying (3.23) and noting $n^j = e^{j \log n}$,

$$\|\nu_n^{\star k} - \pi_n\|^2 \leq \frac{1}{2} \sum_{j=1}^{\lceil n/2 \rceil} \frac{e^{j \log n}}{j!} e^{-j \log n - jc} = \frac{1}{2} \sum_{j=1}^{\lceil n/2 \rceil} \frac{e^{-jc}}{j!}$$

$$\leq \frac{1}{2} \sum_{j=1}^{\infty} \frac{(e^{-c})^j}{j!} = \frac{1}{2} \left(\sum_{j=0}^{\infty} \frac{(e^{-c})^j}{j!} - 1\right) = \frac{1}{2} \left(e^{e^{-c}} - 1\right) \quad \bullet$$

Chapter 4

The Cut-Off Phenomena

4.1 Introduction

Given an ergodic random walk ξ, a number of techniques for bounding $\|\nu^{\star k} - \pi\|$ have been developed. Recall the mixing time, τ, as the minimum k such that $\|\nu^{\star k} - \pi\| \leq 1/2e$. In particular, as $\|\nu^{\star k} - \pi\|$ is decreasing in k, if $\|\nu^{\star k} - \pi\| \leq 1/2e$, then $\tau \leq k$. In many random walks, behaviour called the *cut-off phenomenon* occurs and it makes sense to talk about the mixing time, τ, as the time when ξ is random.
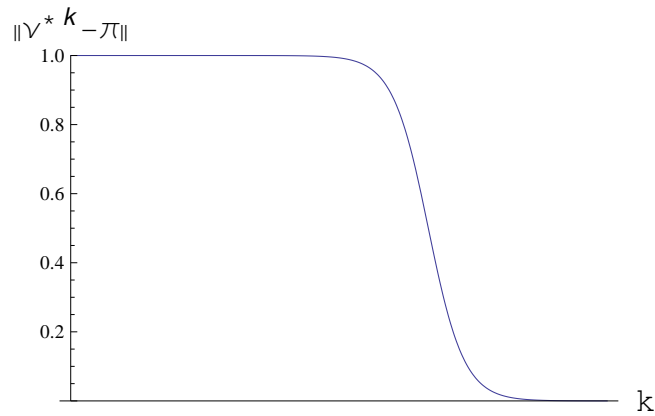

Figure 4.1: In the cut-off phenomenon, variation distance remains close to 1 initially until the mixing time τ when it rapidly converges to 0.

49

In the cut-off phenomenon, the random walk remains far from random until a certain time when there is a phase transition and the random walk rapidly becomes close to random.

4.1.1   Example: Random Transpositions

As described in Section 1.3.1, repeated random transpositions of n cards can be modelled as repeatedly convolving the measure:

$$\nu_n(s) := \begin{cases} 1/n & \text{if } s = e \\ 2/n^2 & \text{for } s \text{ a transposition} \\ 0 & \text{otherwise} \end{cases}$$

Careful analysis of the representation theory of the symmetric group and an application of the Upper Bound Lemma yields [12], for $k = (n \log n)/2 + cn$, for $c > 0$:

$$\|\nu_n^{\star k} - \pi_n\| \leq ae^{-2c} \tag{4.1}$$

for some constant a. For a lower bound, Diaconis considers the set $A \subset S_n$ of permutations with one or more fixed points. Two classical results of Feller[1] give sharp approximations of $\nu_n^{\star k}(A)$ and $\pi_n(A)$ and hence a lower bound for the variation distance may be given. For $k = (n \log n)/2 - cn$, $c > 0$, as $n \to \infty$:

$$\|\nu_n^{\star k} - \pi_n\| \geq \left(\frac{1}{e} - e^{-e^{-2c}}\right) + o(1) \tag{4.2}$$

Hence for n large, the random walk experiences a phase transition from order to random at $t_n = n \log n/2$. Indeed, this was the first problem where a cut-off was detected ([17]).

---

[1]namely the *matching problem* and the computation of the probability that when $2k$ balls are dropped into n boxes, that one or more of the boxes will be empty [18]

4.2 Formulation

There are a number of roughly equivalent formulations of the cut-off phenomenon. The subject developed from the question *how many times must a deck be shuffled until it is close to random?* Card shuffling is modelled by a random walk on S_n where the shuffle is defined by the driving probability $\nu \in M_p(G)$. In most cases, the driving probability ν is related to n so it makes sense to talk about a natural family of random walks (S_n, ν_n). When a good asymptote of the mixing times of these walks was accessible, it was found that in a number of examples that the cut-off behaviour becomes sharper as $n \to \infty$. As a corollary of this development, the cut-off phenomenon is defined with respect to the limiting behaviour of a *natural family* (G_n, ν_n).

In general, a formulation will be referenced to a particular distance of closeness to random. Surprisingly, given different norms on $M_p(G)$, a random walk exhibiting the cut-off phenomenon in the first need not exhibit the cut-off phenomenon in the second. There are a number of roughly equivalent formulations (see Chen's thesis [8]) that introduce a window size w_n. This means that the variation distance goes from 1 to 0 in w_n steps rather than 1 however these formulations still require that $\tau_n \gg w_n$ such that $w_n/\tau_n \to 0$ hence there is still abrupt convergence. The original formulation of Aldous & Diaconis [4] appeals to an arbitrary sharpness of convergence of variation distance to a step function:

4.2.1 Definition

A family of random walks (G_n, ν_n) exhibits the *cut-off phenomenon* if there exists a sequence of real numbers $\{t_n\}_{n=1}^{\infty}$ such that given $0 < \varepsilon < 1$, in the limit as $n \to \infty$, the following hold:

(a) $\|\nu_n^{\star \lfloor (1+\varepsilon)t_n \rfloor} - \pi_n\| \to 0$

(b) $\|\nu_n^{\star \lfloor (1-\varepsilon)t_n \rfloor} - \pi_n\| \to 1$

(c) $t_n \to \infty$

If τ_n is the mixing time of (G_n, ν_n) presenting cut-off, then the above formulation implies that $\tau_n \sim t_n$ so it makes sense to say that t_n is the time taken to reach random.

Example: Walk on the n-Cube

Recall the walk on the n-Cube from the last chapter. Along with the upper bound extracted from the Diaconis-Fourier theory, tedious but elementary calculations bound the variation distance away from 0 for $k = (n+1)(\log n - c)/4$ for n large and $c > 0$ ([7] — Th. 2.4.3). This is done via the test function $\phi(s) = n - 2w(s)$ whose expectation and variance under π are easy to calculate (namely 0 and n). The set $A_\beta \subset \mathbb{Z}_2^n$ is essentially defined as the elements whose weight is sufficiently close to $n/2$ for some β:

$$A_\beta := \{s \in \mathbb{Z}_2^d : |\phi(s)| < \beta\sqrt{n}\}$$

Use of the Markov inequality bounds $\pi_n(A_\beta)$ above $1 - 1/\beta^2$. More intricate calculations yield $\nu_n^{\star k}(A_\beta) \leq 4/\beta^2$ and thence

$$\|\nu_n^{\star k} - \pi_n\| \geq 1 - \frac{5}{\beta^2} \tag{4.3}$$

A more precise definition of β in terms of c makes this lower bound useful[2]. Hence it follows that the random walk has a cut-off at time $t_n = n \log n/4$.

---

[2] if $\beta = e^{c/2}/2$ then the lower bound is $1 - 20/e^c$, which clearly tends to 1 as c increases

Example: Simple Walk on the Circle

The simple walk on the circle does not exhibit cut-off. Considering the bounds developed in Section 3.4.2, note that at $k = n^2/2$, $\|\nu_n^{\star k} - \pi_n\| \leq e^{-\pi^2/4}$, and due to the decreasing nature of $\|\nu_n^{\star k} - \pi_n\|$ this is an upper bound for all $k \geq n^2/2$. Similarly at $k = 3n^2/2$:

$$\|\nu_n^{\star k} - \pi_n\| \geq \frac{1}{2} e^{-3\pi^2/4 - 3\pi^4/4n^2} \underset{n\to\infty}{\to} \frac{1}{2} e^{-3\pi^2/4}$$
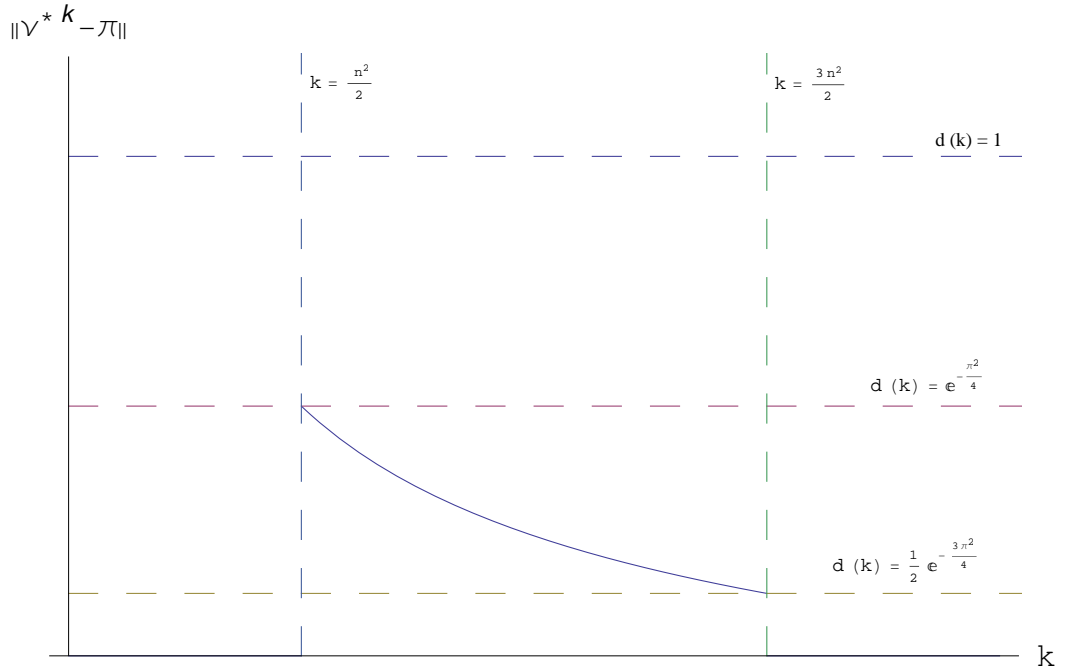
and this lower bound holds for all $k \leq 3n^2/2$.

Figure 4.2: In the limit as $n \to \infty$ the simple walk on the circle does not experience an abrupt transition from far from to close to random. Note that $d(k) := \|\nu_n^{\star k} - \pi_n\|$ and the graph is not to scale.

It is an open problem to determine for which families of random walks (G_n, ν_n) does cut-off occur. Unfortunately there does not appear to be a nice condition for an isolated random walk ξ to exhibit cut-off. In contrast, given G and $\nu \in M_p(G)$, the ergodic theorem 1.3.2 determines whether or not (G, ν) is ergodic.

An initial attempt at reformulation would be to have as fundamental a period of 'far from random' and a period of sharp transition to 'close to random'. Rather than being arbitrarily far from random and arbitrarily close to random (in the limit), this finitary formulation would have to define controls $a, b > 0$ for far and close to random:

4.2.2 Definition

A random walk on G driven by $\nu \in M_p(G)$ has (a, b, q) *finitary cut-off* if $A :=$ $\{k : \|\nu^{\star k} - \pi\| \geq 1 - a\}$, $B := \{k : b \leq \|\nu^{\star k} - \pi\| \leq 1 - a\}$ and $q = |A|/|B|$.

Therefore if (G_n, ν_n) presents cut-off, each member also has (a_n, b_n, q_n) finitary cut-off, where $a_n, b_n \to 0$, $|A_n| \to \infty$, and $q_n \to \infty$. However, consider the natural family $(\mathbb{Z}_n, \nu)$ where ν is uniform on $\{0, \pm 1\}$. This family has $(1/2, 1/4, \mathcal{O}(1))$ finitary cut-off but does not present the cut-off phenomenon. For a family, therefore, presenting cut-off is strictly stronger than presenting finitary cut-off. It is pretty clear that *all* random walks have some level of finitary cut-off. Is there an appropriate level of quality of cut-off?
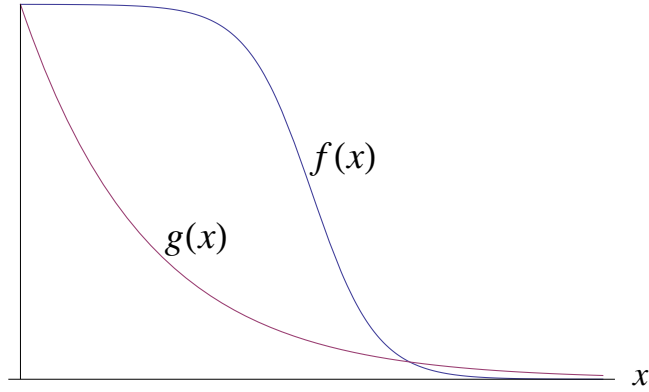

Figure 4.3: In a natural definition of cut-off, the exponential function g should not have cut-off. The other function, f, certainly exhibits some level of cut-off.

A continuous version of (a, b, q) finitary cut-off can be considered. Let $f :$ $\mathbb{R}_+ \to [0, 1]$ be a non-increasing continuous function with $f(0) = 1$ and $f(x) \to 0$. f exhibits (a, b, q) finitary cut-off where $A = \inf\{x : f(x) = 1 - a\}$, $B = \inf\{x :$ $f(x) = b\}$ and $q = A/(A - B)$.

In Figure 4.3, f has $(1/2e, 1/2e, 2.52)$ finitary cut-off while g has $(1/2e, 1/2e, 0.14)$ finitary cut-off. In a number of examples of established cut-off, e.g. the top-to-random shuffle [14], it has been shown that $\|\nu_n^{\star\lfloor(1-\varepsilon)t_n\rfloor} - \pi_n\| \to 1$ doubly exponentially as $\varepsilon \to 1$. Hence consider $(1/e^{2e}, 1/2e, 1)$ finitary cut-off as an appropriate level for cut-off. Indeed f has $(1/e^{2e}, 1/2e, 0.52)$ finitary cut-off while g has $(1/e^{2e}, 1/2e, 0.0026)$ finitary cut-off. However this too runs into problems. Consider the family of functions $f_d(x) = (1 - \tanh(d(x - 1/2)))/2$. This family has $(1/e^{2e}, 1/2e, 1)$ finitary cut-off for $d \gtrsim 12.4$.

Diaconis remarks [12] that Aldous & Diaconis have shown that for most probability measures on a finite group G, $\|\nu^{\star 2} - \pi\| \le 1/|G|$, so for large groups, most random walks are random after two steps.

Therefore, without an alternative formulation of the cut-off phenomenon, it seems likely there will never be a theorem of the form: *A random walk on G with driving probability $\nu \in M_p(G)$ presents 'the' cut-off phenomenon at time k if and only if property P is satisfied.*

4.3 What Makes it Cut-Off?

To demonstrate the intransigence of the problem note that the asymptotics of a reversible random walk $\|\nu^{\star k} - \pi\| \sim C\lambda_\star^k$ cannot detect cut-off. A critical idea for understanding of the cut-off phenomena is that variation distance is sensitive. Suppose a deck of cards is shuffled (by $\nu \in M_p(S_{52})$) but the shuffle leaves the ace of spades at the bottom of the deck. If $A \subset S_{52}$ are the arrangements of the deck with the ace of spades at the bottom, then $\nu(A) = 1$ but $\pi(A) = 1/52$ and $\|\nu - \pi\| \ge 1 - 1/52$; the deck is very far from random in variation distance! Similarly suppose that after shuffling by ν that the ace of spades is in the bottom half of the deck. By letting $B \subset S_{52}$ be all such arrangements it is clear $\|\nu - \pi\| \ge 1/2$. So for any shuffle the entire deck must be well shuffled; it won't do to have even coarse information on a single card.

To illustrate further, consider the top-to-random shuffle. This is the shuffle that takes the top card of the deck and inserts it back into the deck randomly[3]. Suppose the initial arrangement has the ace of spades at the bottom of the deck. Initially it will take a while for a card from the top to be placed underneath the ace of spades but eventually one will be and the ace of spades will be second from bottom. After a great number of shuffles the ace of spades will eventually surface at the top of the deck. At every stage up to this point, to within a statistical deviation, the ace of spades is in a specific portion of the deck, dependent on the number of shuffles. Hence up to this point the deck will be far from random. After this step however the ace of spades shall be placed at a random position in the deck and there is every chance the deck is random. It will be seen in the next chapter that the time for the bottom card to come to the top is essentially the time to random and hence the cut-off time.

The survey article by Diaconis [13] suggests a number of reasons why cut-off may occur. Diaconis claims that high-multiplicity of second eigenvalue implies cut-off after a remark of Aldous & Diaconis [4]. The result from [27]

$$\|\nu^{\star k} - \pi\|_2 \geq m_\star \lambda_\star \tag{4.4}$$

has some implications for this claim in the two norm (see Chen [8]). However, in this thesis, cut-offs in variation norm are the subject of study. One might fear 'folklore heuristic' failure here. Indeed the claim of Diaconis is almost cited as fact by Hora [20, 21]. Perhaps a more measured statement would be that to show cut-off the random walk may have to exhibit a high degree of symmetry which can imply high multiplicity of the second largest eigenvalue. In the extreme case of almost all eigenvalues equal to $\lambda_\star$ (remembering the average of the eigenvalues is $\nu(e)$), the variation distance looks like $C\lambda_\star^k$ and this doesn't look like cut-off.

---

[3]i.e. driven by the measure constant on the cycles $(1, m, m-1, \ldots, 3, 2)$, $m = 1, \ldots, 52$

Chen [8] discusses a conjecture of Peres that a general Markov chain exhibits the cut-off phenomenon if and only if $\tau_n(1 - \lambda_{n\star}) \underset{n \to \infty}{\to} \infty$. Any Markov chain with cut-off will satisfy this condition. Chen & Saloff-Coste [9] have proved this conjecture in the p-norm case for $1 < p < \infty$ however Aldous has given a Markov chain which is a counterexample in variation distance [8]. Presently there is no known counterexample to Peres' conjecture in the case of random walks on groups.

Theorem 2.6.2 is relevant for family of groups (G_n, ν_n) of moderate growth with $|\Sigma|$, A, d fixed as $n \to \infty$. These random walks take large multiple of Δ_n^2 to get random. While a small multiple of Δ_n^2 is not sufficient for randomness, the transition from 1 to 0 as the number of steps grows is smooth so that the cut-off is not exhibited [16]. Diaconis [13] notes that — via Gromov's Theorem for nilpotent groups of finite index — this result is generic. For random walks on families of nilpotent groups where $|\Sigma|$ and the index are bounded as $n \to \infty$, order Δ^2 steps are necessary for convergence and there is no cut-off. Two examples of such walks are the simple walk on the circle and the walk on the Heisenberg groups, and indeed these are the canonical examples where cut-off does not occur.

Chapter 5

Probabilistic Methods

5.1 Stopping Times

In previous chapters the convergence behaviour of a random walks has been ex-
amined. It is natural to ask questions of the type *from which time T onwards
does ξ_T have a particular property*. As a simple example of such a random time,
consider a random walk ξ. The lowest T_0 such that $\xi_{T_0} = e$ is such a random
time, namely the *first return time.*

To make precise, let $\mathcal{A}_k$ be the σ-algebra generated by the random variables
$\{\xi_j : j \leq k\}$, for $j, k \in \mathbb{N}_0$. Then the σ-algebra generated by the σ-algebras
$\{\mathcal{A}_k : k \in \mathbb{N}_0\}$, $\mathcal{A}$, canonically admits an increasing sequence:

$$\mathcal{A}_0 \subset \mathcal{A}_1 \subset \cdots \subset \mathcal{A}_k \subset \cdots \subset \mathcal{A}$$

of sub-σ-algebras of $\mathcal{A}$ (i.e. a *filtration*). If $S(G)$ is the set of sequences in G,
then a *stopping time* is a map $T : S(G) \to \mathbb{N} \cup \{\infty\}$ which satisfies $\{T \leq k\} \in \mathcal{A}_k$
for all $k \in \mathbb{N}$.

To formalise the first example of a stopping time, the first return time, write
$T_0 = \min\{k \geq 1 : \xi_k = e\}$. Of course this generalises easily to another example
of a stopping time, namely the *first hitting time*, $T_g = \min\{k \geq 0 : \xi_k = g\}$.
More generally, a subset $A \subset G$ has first hitting time $T_A = \min\{k \geq 0 : \xi_k \in A\}$

New stopping times may be constructed from old. If T and S are stopping times for a random walk ξ, then so are $\min\{T, S\}$, $\max\{T, S\}$, and $T + n$, $n \in \mathbb{N}$ (see [28] for proof). The standard analysis of stopping times involves an examination of their expectation, E_μ. There is a strong relationship between the random distribution π and stopping times which is given in the following proposition.

5.1.1 Proposition

Let ξ be a random walk on a group G. Let $T \in \mathbb{N}$ be a non-zero stopping time such that $\xi_T = e$ and $E_\mu T < \infty$. Let $g \in G$. Then

$$E_\mu(\text{number of visits to } g \text{ before time } T) = E_\mu T/|G|$$

Proof. Taking the approach of [5] (Proposition 4, Chapter 2), write $\rho(g) = E_\mu(\text{number of visits to } g \text{ before time } T)$. Now

$$\lambda(g) := \frac{\rho(g)}{\sum_t \rho(t)} = \frac{\rho(g)}{E_\mu T}$$

is a probability measure on G. Next it is claimed that

$$\sum_{t \in G} \lambda(t) p(t, g) = \lambda(g). \tag{5.1}$$

To see this note that

$$\lambda(g) = \frac{1}{E_\mu T} \sum_{k=0}^{\infty} \mu(\xi_k = g, T > k).$$

If $g = e$, then $\mu(\xi_0 = e) = \mu(\xi_T = e) = 1$. Also, for $g \neq e$, by hypothesis, $\mu(\xi_0 = g) = \mu(\xi_T = g) = 0$. Therefore, in the reindexing $\xi_k \to \xi_{k+1}$, the term $\mu(\xi_0 = g)$ is replaced by $\mu(\xi_T = g)$ (in the event $T = k + 1$). Thus

$$\lambda(g) = \frac{1}{E_\mu T} \sum_{k=0}^{\infty} \mu(\xi_{k+1} = g, T > k)$$

$$= \frac{1}{E_\mu T} \sum_{k=0}^{\infty} \sum_{t \in G} \mu(\xi_k = t, T > k, \xi_{k+1} = g)$$

By the Markov property,

$$\lambda(g) = \frac{1}{E_\mu T} \sum_{k=0}^{\infty} \sum_{t \in G} \mu(\xi_k = t, T > k) p(t, g)$$

$$= \sum_{t \in G} \frac{\rho(t)}{E_\mu T} p(t, g) = \sum_{t \in G} \lambda(t) p(t, g)$$

Thus it is shown that $\lambda P = \lambda$, and so λ is in fact the unique stationary distribution. Consequently

$$\lambda(g) = \frac{\rho(g)}{E_\mu T} = \pi(g) \Rightarrow \rho(g) = \pi(g) E_\mu T \quad \bullet$$

5.2   Strong Uniform Times

Consider the following shuffling scheme. Given a deck of n cards in order remove a random card and place it on the top of the deck. Repeat this shuffle until the random time T when every card in the deck has been touched. This T is a stopping time and further every arrangement of the deck is equally likely at this time. Call such a stopping time a *strong uniform time*: a stopping time T such that $\mu(\xi_T = g) = 1/|G|$. Diaconis [12] remarks that this is equivalent to $\mu(\xi_k = g | T \leq k) = 1/|G|$.

Aldous & Diaconis [3] gives a classic account of strong uniform times. For many applications, including the random to top shuffle, the classical coupon collector's problem is required knowledge. Consider a random sample with replacement from a collection of n coupons. Let T be the number of samples required until each coupon has been chosen at least once.

5.2.1 Coupon Collector's Bound

In the notation above, let $k = n \log n + cn$, with $c > 0$. Then

$$\mu(T > k) \le e^{-c} \tag{5.2}$$

Proof. The proof is standard but this is taken from [12]. For each coupon b, let A_b be the event *coupon b is not drawn in the first k draws*. The probability of not picking b once is $1 - 1/n$, hence $\mu(A_b) = (1 - 1/n)^k$. Thence

$$\mu(T > k) = \mu\left(\bigcup_b A_b\right) \le \sum_b \mu(A_b) = n\left(1 - \frac{1}{n}\right)^k \le ne^{-k/n} = e^{-c} \quad \bullet$$

Recall the separation distance $s(k)$. The separation distance is related to strong uniform times via the following theorem:

5.2.2 Theorem

If T is a strong uniform time for a random walk driven by $\nu \in M_p(G)$, then for all k

$$\|\nu^{\star k} - \pi\| \le s(k) \le \mu(T > k) \tag{5.3}$$

Conversely there exists a strong uniform time such that the rightmost inequality holds with equality.

Proof. Variation distance is controlled by separation distance so it suffices to prove the rightmost inequality. Again taking the approach of [12], let k_0 be the smallest k such that $\mu(T \le k_0) > 0$. The inequality (5.3) holds if $k = \infty$ and for $k < k_0$. For $k \ge k_0$, $s \in G$:

$$s(k) \le 1 - |G|\nu^{\star k}(s) \le 1 - |G|\mu(\xi_k = s , T \le k)$$

$$s(k) \le 1 - \underbrace{|G|\mu(\xi_k = s | T \le k)}_{=1} \cdot \mu(T \le k)$$

$$\le 1 - \mu(T \le k) = \mu(T > k)$$

See [12] (Theorem 4, Chapter 4C) for the converse result $\bullet$

This result along with the coupon collector's bound applies immediately to the random to top shuffle. The upper bound proved here is supplemented by the (tricky) second result from [12] to yield another example of a random walk exhibiting cut-off:

5.2.3 Theorem

For the random to top shuffle, let $k = n \log n + cn$. Then

$$\|\nu^{\star k} - \pi\| \leq e^{-c} \text{ for } c \geq 0, \qquad (5.4)$$

$$\|\nu^{\star k} - \pi\| \to 1 \text{ as } n \to \infty, \text{ for negative } c = c(n) \to -\infty \qquad (5.5)$$

5.3 Coupling

Coupling is a theoretically stronger method than that of strong uniform times. A *coupling* takes a random walk ξ along with the random walk Π (with random distribution) and couples them as a product process (ξ, Π). The interpretation being that the two random walks evolve until they are equal, at which time they couple, and thereafter remain equal. More formally a coupling of a random walk ξ (with stochastic operator P) takes a 'random' operator Γ on $M_p(G) \times M_p(G)$ and uses it as an input into (ξ, Π) such that the marginal distribution of the first factor is precisely the distribution of ξ. The operator must be random in the sense that $\Gamma(\mu, \pi) = (\mu P, \pi)$. Hence $\Gamma(\nu^{\star k}, \pi) = (\nu^{\star k+1}, \pi)$. The operator must act on $M_p(G) \times M_p(G)$ in such a way that the ξ_k begin to match up with the Π_k until all the elements lie along the diagonal: $\xi_T = \Pi_T$. That is after T steps the process will have the same distribution as the second process: that is after the stopping time $k = T$ steps the walk will be random. Call such a T a *coupling time*. For appropriate couplings, the coupling time, T, may be calculated. To make this argument precise a lemma from [12] about marginal distributions is required.

5.3.1   Lemma

Let G be a finite group. Let μ_1, $\mu_2 \in M_p(G)$. Let $\mu \in M_p(G \times G)$ with margins μ_1, μ_2. Let $\Delta = \{(s,s) : s \in G\}$ be the diagonal. Then

$$\|\mu_1 - \mu_2\| \leq \mu(\Delta^C)$$

Proof. Following Diaconis [12], let $A \subset G$. Thus

$$
\begin{aligned}
|\mu_1(A) - \mu_2(A)| &= |\mu(A \times G) - \mu(G \times A)| \\
&= \big|\mu((A \times G) \cap \Delta) + \mu((A \times G) \cap \Delta^C) \\
&\quad - \mu((G \times A) \cap \Delta) - \mu((G \times A) \cap \Delta^C)\big|
\end{aligned}
$$

The first and third quantities in the absolute sign are equal. The second and fourth give a difference of two numbers, both smaller than $\mu(\Delta^C)$ •

5.3.2   Corollary: Coupling Inequality

If T is a coupling time for a random walk driven by $\nu \in M_p(G)$, then for all k

$$\|\nu^{\star k} - \pi\| \leq \mu(T > k) \tag{5.6}$$

Conversely there exists a coupling such that the inequality holds with equality.

Proof. Let μ be the distribution of (ξ_k, Π). Then μ has marginal distributions $\nu^{\star k}$ and π. Lemma 5.3.1 implies that

$$\|\nu^{\star k} - \pi\| \leq \mu(\Delta^C) = \mu(T > k)$$

See [10] for a proof and discussion of the existence of an optimal coupling time •

5.3.3 Example: A Walk on the n-Cube [25]

Consider the walk on $\mathbb{Z}_2^n$ driven by the measure:

$$\nu_n(s) := \begin{cases} 1/2 & \text{if } s = e \\ 1/2n & \text{if } s = e_i \text{ for some } i \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

An equivalent formulation is that a coordinate is chosen independently from $\{1, \ldots, n\}$ and a coin flip determines whether the coordinate is flipped or not. Consider the following coupling operator Γ. Suppose $\xi_k = \sum_i \alpha_i e_i$ and coordinate j is chosen at random. If the coin is heads, then $\xi_{k+1} = \sum_{i \neq j} \alpha_i e_i + (1 - \alpha_j) e_j$ and the jth coordinate of $\Pi_{k+1} = (1 - \alpha_j)$. If the coin is tails, $\xi_{k+1} = \xi_k$ but the jth coordinate of $\Pi_{k+1} = \alpha_j$. From the marginal viewpoint of ξ, Γ is identical to sampling by ν_n. It remains to show that the coupling is suitably random (as described above). Suppose coordinate j is chosen. The distribution of each coordinate of Π_k is uniform on $\{0, 1\}$. Suppose without loss of generality that the jth coordinate of ξ_k is 1. With equal probability the jth coordinate of Π_{k+1} will be 0 or 1 by the coin flip, hence the coupling operator is suitably random. Hence the coupling time is when all of the coordinates $\{1, \ldots, n\}$ have been chosen. The bound on the coupon collector's bound and the coupling inequality implies the walk is random after $n \log n$ steps.

Chapter 6

Some New Heuristics

6.1 The Random Walk as a Dynamical System

Although the dynamics of a particle in a random walk are indeed random, the dynamics of its probability distribution certainly are not. Indeed note the probability distributions $\{\nu^{\star k}\}_{k \in \mathbb{N}}$ evolve deterministically as $\{\delta^e P^k : k \in \mathbb{N}\}$. Thus the random walk has the structure of a dynamical system $\{M_p(G), P\}$ with fixed point attractor $\{\pi\}$. The two canonical categories of dynamical systems (for which there is an existing literature of powerful methods e.g. [30]) are *topological* and *measure preserving* dynamical systems. Unfortunately at first remove $\{M_p(G), P\}$ appears too coarse and structureless to apply any of these powerful methods. Also the mapping function P is not necessarily invertible and this poses further problems. Indeed in many examples of walks exhibiting cut-off, P may be seen to be singular. Hence the assumption that needs to be made on P to put a structure on $\{M_p(G), P\}$ sufficient for application of dynamical systems methods to the cut-off phenomenon is overly strict. A more fundamental problem occurs in trying to put the structure of a measure preserving dynamical system on the walk in that if a meaningful[1] measure is put on $M_p(G)$, the fact that $(M_p(G))P^k \underset{k \to \infty}{\to} \{\pi\}$ would imply that P is in fact not measure preserving.

---

[1] a measure κ wouldn't be very meaningful if $\kappa(M_p(G)) = \kappa(\{\pi\})$

6.2 Charge Theory

Two features of the ergodic random walk suggest an obvious generalisation. The first is that a stochastic operator conserves the unit weight of $\mu \in M_p(G)$. Suppose $u \in \mathbb{R}^{|G|}$ is a row vector of weight q in the positive orthant. A normalisation ensures $u/q \in M_p(G)$ hence uP/q has weight 1 and thus uP has weight q. A simple calculation shows that given *any* row vector $u \in \mathbb{R}^{|G|}$ of weight q, uP also has weight q. Therefore stochastic operators are weight preserving. This immediately implies that the left eigenvectors of an ergodic stochastic operator are of weight zero: $u_i P = \lambda_i u_i$ (except u_1 of course).

Secondly an ergodic stochastic operator converges to $U = [1/|G|]$ (the matrix with all entries equal to $1/|G|$), so that given a weight 1 row vector u, uP^n converges to π. In particular, if ξ_0 is distributed as any signed probability measure (or *charge*: a signed measure on G such that $\rho(G) = 1$) ρ, the random walk will still converge to the random distribution. This allows an all manner of generalisations. For example, consider the *signed* stochastic operator $Q = [\rho(ht^{-1})]_{th}$ generated by a signed probability measure ρ. Under what conditions will $\delta^e Q^n$ converge to the random distribution?

6.3 Invertible Stochastic Operators

In general a random walk need not start deterministically at e, but rather in an initial distribution $\mu = \sum_t \alpha_t \delta^t$. However $\mu P^n = \sum_t \alpha_t (\delta^t P^n)$. By right-invariance all the $\delta^t P \to \pi$ and hence $\mu P^n \to \pi$ for any initial distribution. In this sense there is a loss of information about initial conditions: the walk forgets where it began, where it was and is totally random. The dynamical systems community make distinctions between the behaviour of invertible and non-invertible maps, however this approach has not been exploited for the case of a random walk on a group.

It would be desirable to quantify the 'folklore thesis' that [23]:

> *The loss of information about initial conditions, as the iteration process proceeds in a chaotic regime, is associated with the non-invertibility of the mapping function... Hence system memory of initial conditions becomes blurred.*

Consider the case of a singular and symmetric stochastic operator P. The spectral theorem implies $\mathbb{R}^{|G|}$ has a basis of (left) eigenvectors of P. Hence $\mathbb{R}^{|G|}$ has an eigenspace decomposition $\bigoplus V_t$, where $V_t := \ker(\lambda_t I - P)$, where $\{\lambda_t : t \in G\}$ are the eigenvalues of P (with the convention $\lambda_1 = 1$). Consider $M_p(G) \subset \mathbb{R}^{|G|} = \bigoplus_t V_t$. With a non-trivial kernel P, can '*destroy information*' and the naïve reaction to this would be to consider $\nu^{\star k} \in V_1 \oplus \ker P$ such that $\|\nu^{\star k} - \pi\| \approx 1$. Then $\|\nu^{\star k+1} - \pi\| = 0$ and there is cut-off. However given $\delta^e \in \bigoplus V_t$, clearly P kills the $\ker P$ terms at the very first iterate, $\delta^e P$, so this heuristic is incorrect. However in contrived examples the sampling could be done by ν_1 until $\nu_1^{\star k} \in V_1 \oplus \ker P_2$ but far from random then sampling by ν_2 (or multiplying by P_2) would project onto V_1. See Section 6.4 for more.

6.3.1   Proposition

A stochastic operator P is invertible if and only if the equation $uP = \pi$ has the unique solution $u = \pi$.

If P is an invertible stochastic operator then the following hold:

(i) If u is an eigenvector of P, then u is an eigenvector of P^{-1}. In particular, $\pi P^{-1} = \pi$ and $P^{-1} k = k$ for any constant function $k \in F(G)$.

(ii) If $\{\lambda_t : t \in G\}$ are the eigenvalues of P, then $\{1/\lambda_t : t \in G\}$ are the eigenvalues of P^{-1}. In particular, 1 is an eigenvalue of P^{-1}, and all other eigenvalues of P^{-1} have modulus greater than 1.

(iii) The signed probability measures on G, $M_1(G)$, are stable under P^{-1}.

(iv) For $k \in \mathbb{N}$, $\delta^e P^{-k} \in M_1(G) \backslash M_p(G)$.

Proof. If P is invertible $uP = \pi$ has unique solution. If P is singular then the kernel is non-trivial. Let $u_1 \neq u_2 \in \ker P$ be normalised such that $\nu_i := \pi + u_i \in M_p(G)$, then $\nu_i P = \pi$.

(i) and (ii) are basic linear algebra facts.

(iii) From (i) the row and column sums of P^{-1} are 1. Thence let $v \in M_1(G)$;

$$vP^{-1}(G) = \sum_{s \in G} \left(\sum_{t \in G} v(t)p^{-1}(t,s) \right)$$
$$= \sum_{t \in G} v(t) \underbrace{\left(\sum_{s \in G} p^{-1}(t,s) \right)}_{=1} \bullet$$

(iv) From (iii), $\delta^e P^{-1} \in M_1(G)$. Assume there exists $\nu \in M_p(G)$ such that $\nu P = \delta^e$. Now $\nu P(s) = \langle \nu, p_s \rangle$ must equal $\delta^e(s)$ where p_s is the row vector equal to the s-column of P. By Cauchy-Schwarz:

$$|\langle \nu, p_s \rangle| \leq \|\nu\|_2 \|p_s\|_2 \leq \|\nu\|_1 \|p_s\|_1 \tag{6.1}$$

Because

$$\nu P(e) = \langle \nu, p_e \rangle = 1 = \|\nu\|_1 \|p_e\|_1$$

the second and third inequalities are equalities for $s = e$. The first equality implies that ν and p_e are linearly dependent, $\nu = kp_e$. As probability measures must have weight 1, this implies $\nu = p_e$. The second equality implies that ν and p_e are Dirac measures. Hence ν is a Dirac measure, say δ^g, and thus P is not ergodic (as Σ is a subset of the coset $\{e\}g$, of the proper normal subgroup $\{e\}$). Inductively given $v \in M_1(G) \backslash M_p(G)$, there does not exist $\nu \in M_p(G)$ such that $\nu P = v$ as v must have negative entries but both ν and P are positive $\bullet$

6.4 Convolution Factorisations of π

Take a deck of cards and transpose the top card with a random card. Next transpose the second card with a random card (at or underneath the second) and continue inductively until all but the second from bottom card has been transposed. Apply the same shuffle to the 51st card ((51,51) or (51,52)). The first card is random, the second is random and inductively all the cards are random. Hence considering the group S_n and the measures ν_i uniform on the transpositions $\{(i,i),(i,i+1),\ldots,(i,n)\}$ the random distribution factorises as:

$$\pi = \nu_{n-1} \star \cdots \star \nu_2 \star \nu_1 \qquad (6.2)$$

Urban [31] considers the question: given a group G and a symmetric set of generators Σ, does there exist a finite number of convolutions of symmetric measures $\{\nu_i \in M_p(G) : i = 1,\ldots,m\}$ supported on Σ such that (6.2) holds (with m rather than $n-1$ terms)? Urban uses Diaconis-Fourier theory (particularly Lemma 3.2.3) to show that if, at a non-trivial irreducible representation of G, ρ, the Fourier transform of $\nu_m \star \cdots \star \nu_1$ is non-zero then (6.2) cannot hold. Briefly, Lemma 3.2.3 states that at any non-trivial irreducible representation, $\widehat{\pi}(\rho) = 0$; and the Fourier transform of $\nu_m \star \cdots \star \nu_1$ is easily computed via the convolution theorem.

If $\nu^{\star k} = \pi$ for some finite $k \in \mathbb{N}$ then the results of Section 2.3 shows that $\nu = \pi$. In particular, as ν is symmetric, P has an eigenbasis, and 1 is an eigenvalue of P with multiplicity 1. Suppose for contradiction that $\nu^{\star k} = \pi$ for some $k \in \mathbb{N}$, but $\nu \neq \pi$. Suppose $\delta^e \in V_1 \oplus \ker P$; then $\delta^e P = \pi$. However $\delta^e P = \nu \star \delta^e$, however $\nu \star \delta^e = \nu$ and thus $\nu = \pi$. Hence at least one of the eigenvectors in the eigenbasis expansion of δ^e is associated with a non-zero eigenvalue. Thus hence $\nu^{\star k} \neq \pi$ for any $k \in \mathbb{N}$. Note that each of the ν_i induces a stochastic operator P_i and (6.2) is equivalent to

$$U = P_m P_{m-1} \cdots P_2 P_1 \qquad (6.3)$$

Note that U is singular. If each of the P_i are invertible then so is U, a contradiction. Therefore (6.3) cannot be true if each of the P_i are invertible. Theorem 6 on page 49 of Diaconis [12] implies that each eigenvalue of $\widehat{\nu}(\rho)$, where ρ is an

irreducible representation, is an eigenvalue of multiplicity d_ρ. In the case of an Abelian group, the eigenvalues of P are simply given by $\{\widehat{\nu}(\rho_i) : \rho_i \text{ irreducible}\}$ and the analysis breaks down to that of Urban's as $\widehat{\nu}(\rho_i) \neq 0$ is equivalent to 0 is not an eigenvalue of P; i.e. P is invertible.

Example: Simple Walk on the Circle

Let n be odd and consider the set $\mathcal{M}$ of not-necessarily symmetric measures with support $\Sigma = \{\pm 1\}$ (i.e. $\mathcal{M} = \{\nu_p \in M_p(G) : \nu_p(1) = p, \nu_p(-1) = 1 - p; p \in (0,1)\}$). Does π admit a finite convolution factorisation of measures from $\mathcal{M}$? For convenience denote $q := 1 - p$ and $\alpha := p/q$. Consider the stochastic operator associated to ν_p:

$$
P_p = \begin{pmatrix}
0 & p & 0 & 0 & \cdots & 0 & q \\
q & 0 & p & 0 & \cdots & 0 & 0 \\
0 & q & 0 & p & \cdots & 0 & 0 \\
\vdots & \vdots & & \ddots & & \vdots \\
p & 0 & 0 & 0 & \cdots & q & 0
\end{pmatrix}
$$

Apply the elementary row operation $r_i \to r_i/q$ to each row and permute the rows by $(r_n r_{n-1} r_{n-2} \cdots r_1)$:

$$
P_p \equiv \begin{pmatrix}
1 & 0 & \alpha & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & \alpha & & 0 & 0 \\
& & \vdots & & & \vdots \\
\alpha & 0 & 0 & 0 & & 1 & 0 \\
0 & \alpha & 0 & 0 & \cdots & 0 & 1
\end{pmatrix}
$$

Now[2] eliminate by $r_{n-1} \to r_{n-1} - \alpha r_1$ and $r_n \to r_n - \alpha r_2$:

$$
P_p \equiv \begin{pmatrix}
1 & 0 & \alpha & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & \alpha & & 0 & 0 \\
\vdots & & & & & \vdots \\
0 & 0 & -\alpha^2 & 0 & & 1 & 0 \\
0 & 0 & 0 & -\alpha^2 & & 0 & 1
\end{pmatrix}
$$

---

[2]if $p < q$ then $\alpha < 1$ and Gershgorin's Theorem implies that P_p is invertible. If $p > q$, then $\alpha > 1$ and elementary row operations give P_p invertible similarly. Gershgorin cannot deal with the case $p = q$ however. Gershgorin can show P_p is invertible with n even when $p \neq q$, but on this support, the walk is not ergodic.

Now suppose $n = 2m + 1$ and continue inductively until:

$$P_p \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ & & & \vdots \\ 0 & (-1)^{m+1}\alpha^m & 1 & 0 \\ 0 & \cdots & 0 & (-1)^{m+1}\alpha^m & 1 \end{pmatrix}$$

A final application of $r_{n-1} \to r_{n-1} - (-1)^{m+1}\alpha^m r_{n-2}$ and $r_n \to r_n - (-1)^{m+1}\alpha^m r_{n-1}$ yields:

$$P_p \equiv \begin{pmatrix} 1 & 0 & 0 \\ & \ddots & \\ 0 & 1 & (-1)^{m+2}\alpha^{m+1} \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

Hence the P_p have n pivots and are thus invertible so a finite convolution of measures from $\mathcal{M}$ is never random.

Urban proves a stronger result using the Diaconis-Fourier theory; namely if $\mathcal{M}$ is a set of measures symmetric on $\{s \in \mathbb{Z}_n : |s| < n/4\}$ then there is no π-factorisation. A quick look at the representation theory of $\mathbb{Z}_n$ shows that the Fourier transform of these measures is bounded away from 0 and hence so are the eigenvalues.

Example: Urban's Transposition Shuffle

Consider the convolution described by at the start of this section. The final driving measure $\nu_{n-1} = (\delta^e + \delta^{(n-1,n)})/2$ generates a singular stochastic operator P_{n-1} by Proposition 6.3.1 (v) and a slight rearrangement shows that all of the ν_i generate singular stochastic operators.

Open Problem

This leads onto the interesting question:

> *For what measures $\nu \in M_p(G)$ is the associated stochastic operator invertible?*

A sufficient condition for invertibility guaranteed by Gershgorin's circle theorem is that $\nu(e) > 1/2$.

6.5 Geometry of the $\|\nu^{\star k} - \pi\|$ Graph

Consider an invertible symmetric ergodic stochastic operator P. Due to the fact that the eigenvalues of P^{-1} (except 1) are all modulus greater than 1, the sequence $\|\nu^{\star(-k)} - \pi\|$ is monotonically increasing to infinity as $k \to \infty$. Hence the graph looks something like:

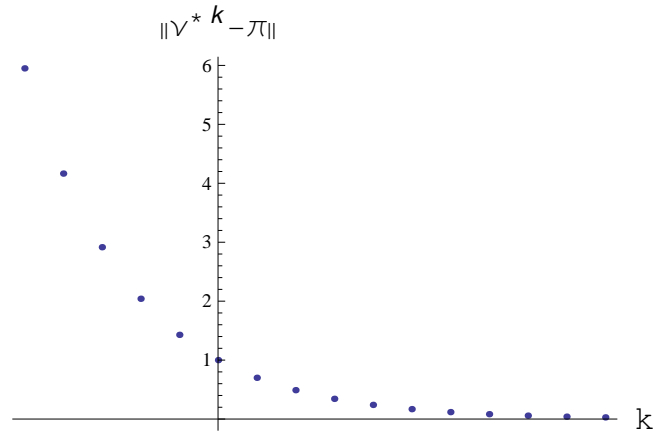Figure 6.1: As $k \to -\infty$, $\nu^{\star k}$ leaves $M_p(G)$ and becomes a 'big' signed measure.

The assumption could be made that in this case the graph *must* be 'concave up' and similarly to $g(x)$ in Figure 4.3, does *not* exhibit cut-off. Suppose an invertible stochastic operator did show cut-off:
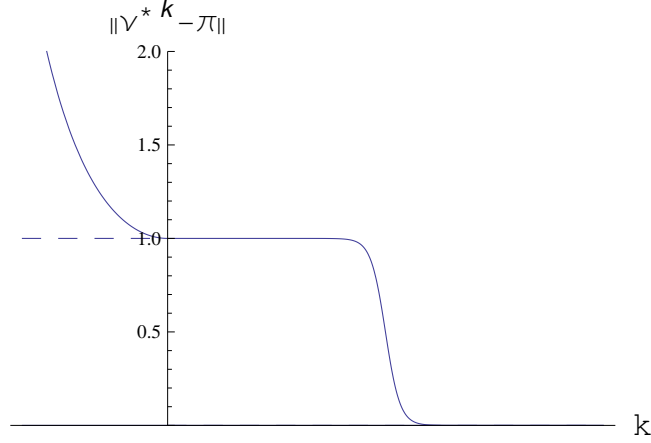

Figure 6.2: One could conjecture that the non-dashed line behaviour, supposedly corresponding to an invertible stochastic operator, with two 'turning points' be impossible.

Instead one might think that somehow the dashed line behaviour is necessary for cut-off to hold — and of course this behaviour cannot hold when P is invertible. This leads to the conjecture: P invertible implies no cut-off. However, in general, $P^{-m}(\{\delta^e\})$ is non-empty, and if a representative u_m from this set is chosen the graph of $\|u_m P^{m+k} - \pi\|$ will exhibit the 'non-dashed line' behaviour. Note that for the random walk on the cube with loops there is no charge that is sent to δ^e by P. This leads onto another interesting question:

Open Problem

> *For what singular stochastic operators P generated by $\nu \in M_p(G)$ does there exist a charge $u \in M_1(G)$ such that $uP = \delta^e$?*

Unfortunately the stochastic operator for the simple walk with loops on $\mathbb{Z}_2^n$ with n even is invertible. If true the conjecture would have placed the problem in a very precarious position. Suppose (G_n, ν_n) is a family exhibiting the cut-off phenomenon (so that the stochastic operator is singular), such that $e \in \text{supp}(\nu_n) = \Sigma_n$. Let $\varepsilon \in (0, 1/2)$, and transform the ν_n as:

$$
\nu_n'(s) := \begin{cases} \frac{1}{2} + \varepsilon & \text{if } s = e \\ \frac{1/2 - \varepsilon}{|\Sigma_n| - 1} & \text{if } s \in \Sigma_n \backslash \{e\} \end{cases} \tag{6.4}
$$

Then by Gershgorin's circle theorem P' would be invertible and hence two random walks with the same support need not exhibit the same behavior: the condition for cut-off to hold would not be on the support only. Unfortunately for those active in the field one would assume the condition is indeed this complex.

Chapter 7

Appendix

7.1 Proof of Lemma 3.4.1

1. **Claim:**

$$\left| \cos\left(j\frac{\pi}{n}\right) \right| = \left| \cos\left(l\frac{\pi}{n}\right) \right| \quad \text{for any } l \in [j]_n \tag{7.1}$$

Suppose $j \equiv l \bmod n$, where $l \in \{0, 1, \ldots, n-1\}$, so that $j = l + mn$ for some $m \in \mathbb{Z}$. Then

$$\begin{aligned}
\cos\left(j\frac{\pi}{n}\right) &= \cos\left((l + mn)\frac{\pi}{n}\right) \\
&= \cos\left(\frac{l\pi}{n} + m\pi\right) \\
&= \cos\frac{l\pi}{n}\cos m\pi - \sin\frac{l\pi}{n}\underbrace{\sin m\pi}_{=0} \\
&= (-1)^m \cos\frac{l\pi}{n}
\end{aligned}$$

Now let $a_t = \cos(\pi t/n)$ and $b_t = \cos(2\pi t/n)$, and note that for $t = 1, 2, \ldots, (n-1)/2$:

$$|a_t| = \begin{cases} \left| b_{(n+t)/2} \right| \quad \text{and} \quad \left| b_{(n-t)/2} \right| & \text{if } t \text{ odd} \\[2mm] \left| b_{t/2} \right| \quad \text{and} \quad \left| b_{n-t/2} \right| & \text{if } t \text{ even} \end{cases} \tag{7.2}$$

Hence as $(x)^2 = |x|^2$:

$$\sum_{t=1}^{n-1} \cos^{2k}(2\pi t/n) = 2 \sum_{t=1}^{(n-1)/2} \cos^{2k}(\pi t/n) \quad \bullet$$

2. Let $h(x) = \log\left(e^{x^2/2}\cos x\right)$; so that $h'(x) = x - \tan x$ and $h''(x) = 1 - \sec^2 x$. Thus $h''(x) \leq 0$ on $[0, \pi/2]$ and so with $h'(0) = 0$, $h(x)$ is a decreasing function in x. In particular, $h(x) \leq h(0) = 0$ and as log is an increasing function, $e^{x^2/2}\cos x \leq 1$, for $x \in [0, \pi/2]$ •

3. In the first instance:

$$\sum_{j=0}^{\infty} e^{-3jx} = \frac{1}{1 - e^{-3x}}$$

is a convergent geometric series when $x > 0$. Now

$$\sum_{j=0}^{\infty} e^{-3jx} = \sum_{j=1}^{\infty} e^{-3(j-1)x}.$$

Also $j^2 - 1 \geq 3(j-1)$ for each $j \in \mathbb{N}_0$. Hence, as e^x is increasing, for all $j \in \mathbb{N}_0$, $e^{-(j^2-1)x} \leq e^{-3(j-1)x}$, and so

$$\sum_{j=1}^{\infty} e^{-(j^2-1)x} \leq \sum_{j=1}^{\infty} e^{-3(j-1)x} = \sum_{j=0}^{\infty} e^{-3jx} \quad •$$

4. Taking the approach of [7], let $h(x) = \log\left(e^{x^2/2+x^4/2}\cos x\right)$;

$$h(0) = 0$$
$$h'(x) = x + x^3 - \tan x\big|_{x=0} = 0$$
$$h''(x) = 3x^2 - \tan^2 x\big|_{x=0} = 0$$
$$h'''(x) = 6x - 2\sec^2 x \tan x\big|_{x=0} = 0$$
$$h^{iv}(x) = 6 + 4\sec^2 x - 6\sec^4 x$$

This is a quadratic in $\sec^2 x$ which is positive when $|\sec x| \leq \sqrt{1 + \sqrt{10}/3}$. This translates into better than $x \in [0, \pi/6]$ •

7.2 Proof of Lemma 3.5.1

1. In the first instance:

$$1 - \frac{2(n+1-l)}{n+1} = 1 - 2 + \frac{2l}{n+1} = -\left(1 - \frac{2l}{n+1}\right)$$

So that

$$\left(1 - \frac{2l}{n+1}\right)^{2k} = \left(1 - \frac{2(n+1-l)}{n+1}\right)^{2k}$$

Secondly,

$$\binom{n}{l} - \binom{n}{n+1-l} = \frac{n!}{l!(n-l!)} - \frac{n!}{(n+1-l)!(\not n - (\not n +1-l))!}$$

$$= \frac{n!}{(l-1)!(n-l)!}\left[\frac{1}{l} - \frac{1}{n+1-l}\right]$$

$$= \frac{n!}{(l-1)!(n-l)!}\left[\frac{n+1-l-l}{l(n+1-l)}\right]$$

$$= \frac{n!}{(l-1)!(n-l)!}\left[\frac{n+1-2l}{l(n+1-l)}\right] \underset{l \leq n/2}{\geq} 0$$

That is, if $l \leq n/2$,

$$\binom{n}{l} \geq \binom{n}{n+1-l} \quad \bullet$$

2. By definition,

$$\binom{a}{b} = \frac{a!}{b!(a-b)!} = \frac{a(a-1)(a-2)\cdots(a-b+1)}{b!} \leq \frac{a^b}{b!} \quad \bullet$$

3. It suffices to show

$$f(j) := \log\left(1 - \frac{2j}{n+1}\right)^{2k} \leq -j\log n - jc =: g(j) \qquad (7.3)$$

as exp is an increasing function. Now writing $k = (n+1)(\log n + c)$,

$$f(1) = \frac{1}{2}(n+1)(\log n + c)\log\left(1 - \frac{2j}{n+1}\right) \text{ , and}$$

$$g(1) = -(c + \log n)$$

Now $c = 4k/(n+1) - \log n$ so $c + \log n = 4k/(n+1)$. Therefore

$$f(1) - g(1) = \left(\frac{4k}{n+1}\right)\left[1 + \frac{1}{2}(n+1)\log\left(\frac{n-1}{n+1}\right)\right]$$

77

This is negative $(f(1) \leq g(1))$ if

$$1 + \frac{1}{2}(n+1) \log \left(\frac{n-1}{n+1} \right) \leq 0$$

$$\Leftrightarrow \log \frac{n-1}{n+1} \leq -\frac{2}{n+1}$$

$$\Leftrightarrow h(n) = \log \left(\frac{n+1}{n-1} \right) - \frac{2}{n+1} \geq 0$$

Now $h(2) = \log 3 - 1 > 0$ and

$$\lim_{n \to \infty} \left[\log \underbrace{\left(\frac{n+1}{n-1} \right)}_{\to 1} - \frac{2}{n+1} \right] = 0.$$

Differentiating with respect to n,

$$h'(n) = -\frac{2}{n^2 - 1} + \frac{2}{(1+n)^2} = -\frac{4}{(n+1)^2(n-1)} \leq 0.$$

Hence $h(n)$ is monotone decreasing from $h(2) > 0$ to 0 so is positive. Hence $f(1) \leq g(1)$. Now differentiating with respect to j,

$$f'(j) = -\frac{(n+1)(c + \log n)}{n+1-2j} = -\frac{4k}{n+1-2j} \underset{j \leq n/2}{\leq} 0$$

Also

$$g'(j) = -c - \log n = -\frac{4k}{n+1}$$

Finally as $j \geq 0$, $f'(j) \leq g'(j)$, for all $j \leq n/2$ ●

Bibliography

[1] K.M. Abadir and J. R. Magnus. *Matrix Algebra.* Cambridge University Press, New York, 2005.

[2] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. *Seminar on probability, XVII, 243-297, Lecture Notes in Math., 986, Springer, Berlin*, 1983.

[3] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly* **93***, 333-348*, 1986.

[4] D. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.* **8***, 69-97*, 1987.

[5] D. Aldous and J.A. Fill. *Preliminary version of a book on finite Markov chains.* http://www.stat.berkeley.edu/users/aldous, 2010.

[6] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2***, 294-313*, 1992.

[7] F. Ceccherini-Silberstein, T. Scarabotti and F. Tolli. *Harmonic Analysis on Finite Groups.* Cambridge University Press, New York, 2008.

[8] G. Chen. *The Cutoff Phenomenon for finite Markov chains.* PhD thesis, Cornell University, 2006.

[9] G. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electronic Journal of Probability, 13(3), 2678*, 2008.

[10] S. Connor. *Coupling: Cutoffs, CFTP and Tameness.* PhD thesis, Warwick University, 2007.

[11] T.M. Cover and J.A. Thomas. *Elements of Information Theory, 2nd Edition.* Wiley, NJ, 2006.

[12] P. Diaconis. *Group Representations in Probability and Statistics.* IMS: Hayward, CA, 1988.

[13] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci U.S.A. 93-94, 1659-1664*, 1996.

[14] P. Diaconis, J.A. Fill, and J. Pitman. Analysis of top to random shuffles. *Combin. Probab. Comput. **1**, 135-155*, 1992.

[15] P. Diaconis and L. Saloff-Coste. Comparison techniques for random walks on finite groups. *Ann. Probab. **21**, 2131-2156*, 1993.

[16] P. Diaconis and L. Saloff-Coste. Moderate growth and random walk on finite groups. *Geom. Funct. Anal. **4**, no. 1, 1-36*, 1994.

[17] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete **57**, 159-179*, 1981.

[18] W. Feller. *An Introduction to Probability Theory and its Applications,* **Vol. 1**, *3rd Edition.* Wiley, New York, 1968.

[19] N. Fountoulakis. *Random Walks on Finite Groups: An Introduction.* http://web.mat.bham.ac.uk/nikolaos/rwg.pdf, 2010.

[20] A. Hora. A critical phenomenon appearing in the process of particle diffusion in classical statistical mechanics. *Journal of The Faculty of Environmental Science and Technology, Okayama University **2**, (1) 1-8*, 1997.

[21] A. Hora. An axiomatic approach to the cut-off phenomenon for random walks on large distance-regular graphs. *Hiroshima Math. J. **30**, 271-299*, 2000.

[22] T.W. Hungerford. *Abstract Algebra: An Introduction.* Brooks-Cole: U.S.A., 1997.

[23] T. W. B. Kibble and F. H. Berkshire. *Classical Mechanics, 5th Edition.* Imperial College Press, 2004.

[24] D.W. Lewis. *Matrix Theory.* Cambridge University Press, Reprint, 2007.

[25] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomised Algorithms and Probabilistic Analysis.* Cambridge University Press, 2005.

[26] J.S. Rosenthal. *A First Look at Rigorous Probability Theory, 2nd Edition.* World Scientific, 2006.

[27] L. Saloff-Coste. *Random Walks on Finite Groups (Probability on Discrete Structures), p.263-346.* Springer, Berlin, 2004.

[28] R.L. Schilling. *Measures, Integrals and Martingales.* World Scientific, 1991.

[29] F. Schmitt and F. Rothlauf. On the importance of the second largest eigenvalue on the convergence rate of genetic algorithms. *Illinois Genetic Algorithms Laboratory, Report No. 2001021*, 2001.

[30] T. Tao. *MATH 254A : Topics in Ergodic Theory (Lecture Notes).* http://terrytao.wordpress.com/category/254a-ergodic-theory/, 2010.

[31] R. Urban. Some remarks on the random walk on finite groups. *Colloq Math* ***74****, No.2 287-298*, 1997.