A RANDOM WALK IN REPRESENTATIONS

Shanshan Ding

A DISSERTATION

in

Mathematics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2014

Robin Pemantle, Merriam Term Professor of Mathematics
Supervisor of Dissertation

David Harbater, C. H. Browne Distinguished Professor of Mathematics
Graduate Group Chairperson

Dissertation Committee:
Jonathan Block, Professor of Mathematics
Robin Pemantle, Merriam Term Professor of Mathematics
Martha Yip, Lecturer of Mathematics

A RANDOM WALK IN REPRESENTATIONS

Acknowledgments

In the sunset of dissolution, everything is illuminated by the aura of nostalgia, even the guillotine.

– Milan Kundera, *The Unbearable Lightness Of Being*

First I must thank my advisor Robin Pemantle for, among many other things, introducing me to probability at a time when I was lost for a direction, suggesting the immensely satisfying topic of this dissertation, and keeping me on track while supporting whatever I wish to pursue. I am also indebted to Martha Yip for her indispensable technical assistance and general awesomeness; she always has answers to my questions and is invariably generous with her time. Moreover, I am grateful for Jim Haglund's advice and encouragement in the later stages of this work.

The graduate chairs during my time at Penn have all been very sympathetic. Tony Pantev and David Harbater helped me navigate the beginning and the end of grad school, while Jonathan Block refused to let me give up on myself through all the years in the middle.

I would like to thank Persi Diaconis for pioneering a beautiful field of mathematics at the intersection of probability and representation theory. Just as importantly, I thank Ehrhard Behrends, whom I have never met, but without whose wonderfully clear introduction to this field I may still be searching for a thesis topic.

I want to acknowledge the professors who inspired me to go to grad school and whose recommendation letters made it possible: Robert Friedman, Dorian Goldfeld, John Morgan, and Ken Ono. In fact, I am deeply grateful for all of the outstanding teachers and mentors, in and out of math, that I have encountered from grade school to grad school.

At the same time, I also thank my students, who have collectively taught me much more than I can ever teach them. Their joy in moments of breakthrough reminded me of why I chose to study math on those occasions when it was otherwise difficult to remember.

The department's administrative staff—Janet Burns, Monica Pallanti, Paula Scarborough, and Robin Toney—steadfastly protects us and holds the department together. They bring sense and warmth into a building that I think has been proved to be topologically baffling and bleak. I wish Janet the very best in her upcoming retirement, though a DRL without her is an unfathomable place.

I am proud to have been part of the Penn math department's graduate student community, which has consistently been comprised of friendly, talented, and all-around wonderful people. It is impossible to name everyone as the requisite margins of this thesis are too large, but I would be remiss to not at least reference a subset.

In particular, I would like to thank Ying Zhao for being my officemate during an unforgettable year. I am also much indebted to Hilaf Hasson for his wisdom and patience. Michael Lugo almost single-handedly taught me to think like a probabilist. Deborah Crook, from Montreal to Marrakech, and from California to Cappadocia, it has been fun traveling half of the world with you, and here's to hoping that we get to the other half at some point. Paul Levande, I already miss your insight and wit, and I would be so very sad to miss watching in person your reaction to Ben Affleck's Batman. Additionally, I am grateful for the friendship and extensive mathematical assistance over the years from Adam Topaz, David Lonoff, Elaine So, Matthew Wright, Matti Astrand, Ryan and Better Ryan (Eberhart and Manion, not necessarily in that order), Taisong Jing, Torin Greenwood, Tyler Kelly, Ying Zhang, and Zhentao Lu.

I would like to thank the Philadelphia Orchestra for affording students in the area the weekly opportunity to attend world-class concerts. These concerts have been reminders on many Saturdays that, high or low, light or heavy, all notes eventually come to pass as a rich, transformative melody is left behind.

Finally, it's *not* always sunny in Philadelphia, and I extend my sincerest gratitude to all those whose optimism and support, in good times and bad, sustained me through the storms.

The writing of this dissertation is partially supported by NSF grant DMS-1209117.

ABSTRACT

A RANDOM WALK IN REPRESENTATIONS

Shanshan Ding

Robin Pemantle

The unifying objective of this thesis is to find the mixing time of the Markov chain on S_n formed by applying a random n-cycle to a deck of n cards and following with repeated random transpositions. This process can be viewed as a Markov chain on the partitions of n that starts at (n), making it a natural counterpart to the random transposition walk, which starts at (1^n). By considering the Fourier transform of the increment distribution on the group representations of S_n and then computing the characters of the representations, Diaconis and Shahshahani showed in [DS81] that the order of mixing for the random transposition walks is $n \ln n$. We adapt this approach to find an upper bound for the mixing time of the n-cycle-to-transpositions shuffle. To obtain a lower bound, we derive the distribution of the number of fixed points for the chain using the method of moments. In the process, we give a nice closed-form formula for the irreducible representation decomposition of tensor powers of the defining representation of S_n. Along the way, we also look at the more general m-cycle-to-transpositions chain ($m \leq n$) and give an upper bound for the mixing time of the $m = n - 1$ case as well as characterize the expected number of fixed points in the general case where m is an arbitrary function of n.

Contents

List of Figures

Chapter 1

Introduction

The goals we pursue are always veiled.

– *The Unbearable Lightness Of Being*, "Words Misunderstood"

1.1 A mathematical history of card shuffling

We start with a question of centuries-old interest to diviners, gamblers, and magicians: how many shuffles does it take to mix a deck of cards?

Naturally, the answer depends on what we mean by "shuffle" and "mix". Broadly speaking, a shuffle on n cards is a permutation of the set $\{1, 2, \ldots n\}$ by an element σ of the symmetric group S_n. The outcome of a sequence of shuffles $\sigma_1, \sigma_2, \ldots, \sigma_k$ is then permutation by the composition $\sigma_k \cdots \sigma_2 \sigma_1$. We presume that each σ_i is chosen according to some probability distribution on S_n, so that the sequence $\sigma_1, \sigma_2\sigma_1, \sigma_3\sigma_2\sigma_1, \ldots$ forms a Markov chain on S_n. If, furthermore, each σ_i is chosen from the same distribution, then this chain is a random walk on S_n. The distribution of $\sigma_k \cdots \sigma_2\sigma_1$ is a probability measure on S_n for each k, and the deck is mixed when the total variation distance between this measure and the uniform measure on S_n is small. Intuitively, mixing means that one can no longer infer the positions of the cards from their initial order.

Since the early 1900s and especially during the past 30 years, mathematical analyses of card shuffling have inspired significant progress in the theory of Markov chain mixing times, particularly in revealing its rich connections with algebraic combinatorics. Markov himself had cited card shuffling as a leading example of his eponymous processes, and his 1906 proof in [Mar06] for the convergence of finite-state Markov chains implies that shuffling eventually mixes the deck. Poincaré then supplied a Fourier-analytic proof in [Poi12].

Of course, *eventual* mixing has always been the implicit premise of card shuffling, so the more pertinent question is *how soon*. The first significant breakthrough in this topic came in 1981, when Diaconis (a former professional magician) and Shahshahani showed in [DS81] that the order of mixing for the random transposition shuffle, where one repeatedly chooses two random cards and exchanges them, is $n \ln n$. Though this shuffle is unlikely to be employed by card players in real life, [DS81] is a landmark development in probability theory for introducing techniques from representation theory. A very high-level summary of its ideas is as follows: Fourier transforms convert convolutions of probability distributions in the time domain to pointwise products in the frequency domain, and the "frequency domain" for a non-abelian group is given by its group representations, so we can track the mixing of a Markov chain by observing the Fourier transform of the increment distribution on the representations of the underlying group, which in turn can be quantified by computing and summing the characters of the representations. This approach is applicable to all random walks generated by conjugacy classes of finite groups, and it was used by Hildebrand ([Hil92]: random transvections in $SL_n(F_q)$), Pemantle ([Pem94]: 3-cycles), and Lulov ([Lul96]: a wide class of fixed-point-free permutations, including fixed-point-free involutions) to obtain the mixing times of various other random walks that fit the description.

Meanwhile, further techniques arose from studies of Markov chains that more realistically model human card shuffling. Aldous and Diaconis [AD86] introduced the concept of strong stationary time to prove that the order of mixing for the top-to-random shuffle, where the top card is removed and inserted into the deck at a random position, is $n \ln n$. Using a coupling construction, Pemantle [Pem89] proved an upper bound of $O(n^2 \ln n)$ for the overhand shuffle, where one shaves off packets of cards from the top of the deck and stacks each packet on top of the previous one until all cards have been transferred to the new pile. This bound was ultimately shown to be tight by Jonasson [Jon06] using a method for establishing sharp lower bounds due to Wilson [Wil04]. As for the riffle shuffle, the most common shuffling technique where one divides a deck into two piles and interlaces them together, Bayer and

Diaconis [BD92] concluded that seven shuffles are necessary and sufficient to mix a 52-card deck and in the process related the underlying Markov chain to Solomon's descent algebra and Hochschild homology. Throughout the 1990s, extensions of the techniques used to study card shuffling have led to substantial progress in the general study of stochastic processes on groups, including diffusions on Lie groups (see [S-C01]). For comprehensive surveys of the works produced, refer to [Dia01] and [S-C04].

Research in card shuffling and related topics is active and ongoing. Recent areas of focus have included systematic scan versions of well-understood shuffles, whereby the location of each update is deterministic ([MPS04], [MNP12]), and randomization of only selected features, such as card values but not suits [CV06] or the location of the original bottom card [ADS11]. Extensive effort has also been devoted to exploring and leveraging the symbiotic connections between card shuffling and the theories of Lie type groups ([Ful00], [Ful01]), quasi-symmetric functions ([Sta01], [DF09]), and Hopf algebras [DPR12]. The pervasive theme in this line of research since Diaconis and Shahshahani's analysis of the random transposition shuffle has been the marrying of spectral and probabilistic phenomena and techniques, a theme that reverberates in the modern studies of expander graphs (see [HLW06]) and random matrices (see [Tao12]).

As Markov chains have a wide range of applications, any new development in the field has built-in ramifications for potentially multiple areas of applied math.

Two standout applications derived specifically from the work on card shuffling are in cryptography ([Mir02], [HMR12]), where shuffles are exploited as enciphering schemes, and in genetics ([SG89], [Dur03]), where shuffles model rearrangements of DNA segments. Of course, we should not overlook the implications of card shuffling research for card playing itself ([Tho73], [CH10]). Vegas certainly paid attention and even invited Diaconis, the renegade magician, for a homecoming of sorts to assess some new automated shuffling machines. For the findings of the said investigation, see [DFH13], though we take this opportunity to put forth the disclaimer that no knowledge of gambling will be endorsed or imparted, here and throughout.

1.2 Scope and organization of this thesis

Nearly the entirety of the the card shuffling literature that we just surveyed deals with time-homogeneous Markov chains, where the same method of shuffling is repeated until the deck is mixed. The present thesis, on the other hand, is motivated by a time-inhomogeneous Markov chain: after a single application of an n-cycle to a deck of n cards, how many transpositions are needed to mix the deck?

This chain is a natural counterpart to the random transposition walk on S_n in the following sense: a transposition changes the cycle type of a permutation by either splitting a cycle in two (if the two objects transposed are in the same cycle) or joining two cycles as one (if the two objects are in different cycles), so random transpositions in fact induce a Markov chain on the set of partitions of n;

the time-homogeneous random transposition walk is one such chain that starts at the partition (1^n), whereas the process we proposed is one that starts at the other extreme, (n). Markov chains formed on partitions under random transpositions are examples of coagulation-fragmentation processes, the profound mathematics and applications of which are surveyed in [Ald99], and a related chain whose eigenfunctions give probabilistic interpretations for the Macdonald polynomials is constructed in [DR12].

The focus of this thesis is the n-cycle-to-transpositions chain viewed as a process on S_n instead of on the partitions of n, though we do hope that our work can lead to new insight on coagulation-fragmentation processes. We will in fact consider the more general process of a random m-cycle ($m \leq n$) followed by random transpositions. Formally:

Question. Fix m as a function $m(n)$ of n with $2 \leq m \leq n$ for all n. Form a Markov chain $\{X_k\}$ on the symmetric group S_n as follows: let X_0 be the identity[1], set $X_1 = \pi X_0$, where π is a uniformly selected m-cycle, and for $k \geq 2$ set $X_k = \tau_k X_{k-1}$, where τ_k is a uniformly selected transposition. Observe that $X_k \in A_n$ when m and k are of the same parity. Otherwise, $X_k \in S_n \backslash A_n$. Let μ_k be the law of X_k, and let U_k be the uniform measure on A_n if $X_k \in A_n$ and the uniform measure on $S_n \backslash A_n$ if $X_k \in S_n \backslash A_n$. What is the total variation distance between μ_k and U_k?

The increment distributions of these Markov chains are conjugacy-invariant, so

[1]Markov chains on finite groups are translation-invariant, so setting X_0 to some other element of S_n may affect parity, but not mixing time.

we follow Diaconis and Shahshahani's approach and adapt their analysis of the random transposition shuffle to obtain upper bounds for the mixing times of the $m = n$ and $m = n - 1$ cases. The relevant concepts and tools from probability, Fourier analysis, and representation theory are introduced and modified as necessary in Chapter 2, while the computations are carried out in Chapter 3. The lower bound was much trickier, but we ultimately obtain one for the $m = n$ case in Chapter 4 by deriving the distribution of the number of fixed points using the method of moments. Putting the two together gives the main result:

Theorem 3.1.1 and Corollary 4.3.5. *For any $c > 0$, after one n-cycle and cn transpositions,*

$$\frac{e^{-2c}}{e} - o(1) \leq \|\mu_{cn+1} - U_{cn+1}\|_{\text{TV}} \leq \frac{e^{-2c}}{2\sqrt{1 - e^{-4c}}} + o(1)$$

as n goes to infinity.

Our arguably most significant contribution is that, while trying to compute the moments of the fixed point distribution, we discovered a neat (in all senses of the word) formula for the decomposition of tensor powers of the defining representation (see Definition 4.2.1) ϱ of S_n:

Theorem 4.2.3. *Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$ be a partition of n, and let S^λ denote the irreducible representation of S_n corresponding to the shape λ. For $1 \leq r \leq n - \lambda_2$, the multiplicity $a_{\lambda,r}$ of S^λ in the irreducible representation decomposition of $\varrho^{\otimes r}$ is* *given by*

$$a_{\lambda,r} = f^{\bar{\lambda}} \sum_{i=|\bar{\lambda}|}^{r} \binom{i}{|\bar{\lambda}|} \left\{ {r \atop i} \right\},$$

where $\bar{\lambda}$ *is the truncated partition* $(\lambda_2, \ldots, \lambda_l)$ *of weight* $|\bar{\lambda}|$, $f^{\bar{\lambda}}$ *is the number of standard Young tableaux of shape* $\bar{\lambda}$, *and* $\left\{ {r \atop i} \right\}$ *is a Stirling number of the second kind.*

In Chapter 5 we give two more results on expected numbers of fixed points, one about the m-cycle-to-transpositions chain for arbitrary m, and the other the following little gem:

Proposition 5.1.1. *If a Markov chain on* S_n *whose increment distributions are class measures starts with one fixed point, then it will always average exactly one fixed point.*

We then conclude by reflecting on what could have been and what could still be, enumerating questions that seem just out of reach and suggesting related topics that may be within grasp.

Chapter 2

Technical Preparations

Without realizing it, the individual composes his life according to the laws of beauty even in times of greatest distress.

– *The Unbearable Lightness Of Being*, "Soul and Body"

2.1 Markov chains

As the King asked of the White Rabbit, we begin at the beginning. Specifically, we begin with a very brief introduction to the central objects of this thesis: Markov chains. For comprehensive treatises, check out [LPW08] or [Beh00].

Definition 2.1.1. A sequence of random variables $(X_0, X_1, \ldots)$ is a *Markov chain* on a finite set Ω if, for all $x_i \in \Omega$ and $k \geq 1$,

$$\mathbf{P}(X_{k+1} = x_{k+1} \mid X_k = x_k) = \mathbf{P}(X_{k+1} = x_{k+1} \mid X_0 = x_0, \ldots, X_k = x_k). \tag{2.1.1}$$

In words, given the present, the future is independent of the past.

When a Markov chain is at state x, the next position is chosen according to a fixed probability distribution $P(x, \cdot)$. If every step is chosen according to the same transition matrix P, then the chain is said to be *time-homogeneous*, and the k-step transition probabilities are given by P^k.

If Ω is a finite group, a probability distribution μ on Ω induces a Markov chain with transition probabilities $P(x, yx) := \mu(y)$. This means that the chain moves via left multiplication by a random element of Ω selected according to μ. The measure μ is called the *increment distribution* on Ω.

Definition 2.1.2. A chain is *irreducible* if it is possible to get from any state to any other state.

Definition 2.1.3. Let $\mathcal{T}(x)$ be the set of times when it is possible for a chain starting at state x to return to x. The *period* of x is the gcd of $\mathcal{T}(x)$. The chain is *aperiodic* if all states have period one.

Definition 2.1.4. For a time-homogeneous Markov chain on Ω with transition matrix P, a distribution π on Ω satisfying $\pi P = \pi$ is a *stationary distribution* of the chain.

If Ω is a finite group, then for a chain on Ω with increment distribution μ, the uniform distribution U_Ω satisfies

$$\sum_{y \in \Omega} U_\Omega(y) P(y, x) = \frac{1}{|\Omega|} \sum_{y \in \Omega} P(y, x) = \frac{1}{|\Omega|} \sum_{y \in \Omega} \mu(xy^{-1}) = \frac{1}{|\Omega|} = U_\Omega(x) \qquad (2.1.2)$$

for all $x \in \Omega$, as the second to last equality is from the observation that the operation $y \to xy^{-1}$ re-indexes Ω. Thus the uniform distribution is a stationary distribution for Markov chains on finite groups. Note that as a result, the distance to stationarity does not depend on the initial state: a chain that starts at x is simply a translation by x of a chain starting at the identity element, and the uniform distribution is translation-invariant.

Most of the theory on finite-state Markov chains is concerned with the long-term behavior of the chains. In particular, we would like to know whether a chain converges to a stationary distribution and, if so, how quickly. To quantify the speed of convergence, we need an appropriate metric for measuring the distance between probability distributions.

Definition 2.1.5. The *total variation distance* between measures μ and ν on Ω is

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|. \tag{2.1.3}$$

Theorem 2.1.6 (Markov chain convergence theorem)**.** *Every time-homogeneous, irreducible, and aperiodic Markov chain has a unique stationary distribution* π*. Furthermore, there exist constants* $0 < \alpha < 1$ *and* $C > 0$ *such that*

$$\max_{x \in \Omega} \|P^k(x, \cdot) - \pi\|_{\mathrm{TV}} \leq C\alpha^k. \tag{2.1.4}$$

Proof. See Theorem 4.9 of [LPW08]. □

The convergence theorem states the sufficient condition for mixing and even specifies that mixing is exponentially fast. However, it gives no information on how to determine the actual rate of convergence, which typically needs to be handled on a case-by-case basis.

Before moving on, we should note that the Markov chain defined in Chapter 1 is time-inhomogeneous and periodic. While periodicity will present complications in the next section, it is not difficult to see that this chain alternates between A_n and $S_n \backslash A_n$, and that each of the two subsequences converges to the uniform distribution on the corresponding coset. It is also clear that inhomogeneity does not affect *whether* a chain converges as long as the chain is time-homogeneous after a finite number of steps.

2.2 Harmonic analysis on finite groups

In this section we present an overview of Diaconis and Shahshahani's approach to analyzing Markov chain mixing times. A detailed and accessible treatment of the material can be found in Chapters 15 and 16 of [Beh00]. Another helpful resource is Chapter 3 of [CST08].

In what follows, let G be a finite group.

Definition 2.2.1. A d-dimensional *(unitary) representation* ρ of G is a group homomorphism from G to the set of d-by-d unitary matrices, that is, $\rho(gh) = \rho(g)\rho(h)$ for all $g, h \in G$. The 1-dimensional representation that sends every $g \in G$ to 1 is the *trivial representation* ρ_{triv} of G.

Example. The 1-dimensional representation of S_n which is 1 on A_n and -1 on $S_n \backslash A_n$ is the *sign representation* of S_n.

Definition 2.2.2. (1) Representations ρ' and ρ'' of the same dimension d are *equivalent* if there exists a d-by-d unitary matrix M such that $\rho_2(g) = M\rho_1(g)M^{-1}$ for all $g \in G$.

(2) A representation ρ is *irreducible*, an *irrep* for short, if it is not equivalent to a representation of the form $\rho_1 \oplus \rho_2$.

Remark. By Maschke's theorem (see, for instance, Theorem 1.5.3 of [Sag01]), every representation of a finite group is equivalent to a direct sum of irreps.

An alternative way to characterize representations of G is in terms of the vector spaces that elements of G act on.

Definition 2.2.3. A vector space V is a G-module if there is a G-action $\cdot$ on V such that $g \cdot (av + bw) = a(g \cdot v) + b(g \cdot w)$ for all $g \in G$, $v, w \in V$, and $a, b \in \mathbb{C}$. We say that a G-module V *carries a representation* of G. Two representations are *equivalent* if their associated G-modules are isomorphic, and a representation is *irreducible* if its associated G-module contains no non-trivial G-submodule.

Remark. To go back and forth between Definitions 2.2.1 and 2.2.3, define the group action $g \cdot v$ to be $(\rho(g))(v)$.

We use $\hat{G}$ to denote a collection[2] of representations of G that contains precisely

[2] If G is abelian, then all representations of G are 1-dimensional and $\hat{G}$ is a group itself, commonly referred to as the *Pontryagin dual* of G.

one representative from each equivalence class of irreps of G.

Definition 2.2.4. Let f be a function on G. The *Fourier transform* of f is the matrix-valued map on $\hat{G}$ defined by $\hat{f}(\rho) = \sum_{g\in G} f(g)\rho(g)$.

The key idea of Diaconis and Shahshahani's approach is to translate the question of "how close to uniformity is μ" to "how close to $\mathbf{0}$ (the zero matrix) is $\hat{\mu}$ on the non-trivial[3] irreps of G". The following helps to start making this idea precise:

Theorem 2.2.5 (Plancherel's formula)**.** *For any function f on G,*

$$\sum_{g\in G} |f(g)|^2 = \frac{1}{|G|}\sum_{\rho\in\hat{G}} d_\rho \mathrm{tr}[\hat{f}(\rho)(\hat{f}(\rho))^\dagger], \tag{2.2.1}$$

where d_ρ is the dimension of ρ and $(\hat{f}(\rho))^\dagger$ is the conjugate transpose of $\hat{f}(\rho)$.

Proof. See Proposition 16.16 of [Beh00]. □

Remark. Theorem 2.2.5 is a consequence of the celebrated Peter-Weyl theorem, which says that the collection of normalized coordinate functions

$$\left\{ \sqrt{d_\rho/|G|}\varphi^\rho_{ij} : \rho\in\hat{G}, 1\le i,j\le d_\rho \right\}, \tag{2.2.2}$$

where φ^ρ_{ij} is defined by assigning to $\varphi^\rho_{ij}(g)$ the ij-th entry of $\rho(g)$, is an orthonormal basis for the space[4] of L^2 functions on G. The Peter-Weyl theorem applies to all compact topological groups; a proof for the case of finite groups is given in Theorem 16.11 of [Beh00].

[3] As we will see, the sign representation is also excluded for the m-cycle-to-transpositions chain due to the parity of the chain.

[4] This is a Hilbert space with the inner product $\langle f_1, f_2\rangle_G = \sum_{g\in G} f_1(g)\overline{f_2(g)}$.

Let μ and ν be measures on G. Rewriting Theorem 2.2.5 with $f = \mu - \nu$ gives

$$\sum_{g \in G} (\mu(g) - \nu(g))^2 = \frac{1}{|G|} \sum_{\rho \in \hat{G}} d_\rho \mathrm{tr}[(\hat{\mu}(\rho) - \hat{\nu}(\rho))(\hat{\mu}(\rho) - \hat{\nu}(\rho))^\dagger], \tag{2.2.3}$$

and the connection to mixing begins to emerge.

If a Markov chain $\{X_0, X_1, \ldots\}$ on G is time-homogeneous with increment distribution υ, then the k-step move from X_0 to X_k is governed by υ^{*k}, the k-fold convolution of υ.

Definition 2.2.6. Let υ and η be measures on G. Their *convolution* is the measure defined by $(\upsilon * \eta)(g) = \sum_{h \in G} \upsilon(gh^{-1})\eta(h)$.

Proposition 2.2.7. *For any υ and η on G, $\widehat{\upsilon * \eta} = \hat{\upsilon}\hat{\eta}$. Thus $\widehat{\upsilon^{*k}} = \hat{\upsilon}^k$.*

Proof. See Proposition 16.19 of [Beh00]. □

Proposition 2.2.8. *If μ is a symmetric measure, i.e. if $\mu(g) = \mu(g^{-1})$ for all $g \in G$, then $\hat{\mu}(\rho) = (\hat{\mu}(\rho))^\dagger$ for all $\rho \in \hat{G}$.*

Proof. See Lemma 16.23 of [Beh00]. □

Proposition 2.2.9. *If ρ is any non-trivial irrep of G, then $\sum_{g \in G} \rho(g) = \mathbf{0}$, and hence $\widehat{U_G}(\rho) = \mathbf{0}$ for the uniform measure U_G on G.*

Proof.[5] Since ρ is non-trivial, there exists $g_0 \in G$ such that $\rho(g_0) \neq I_{d_\rho}$, and

$$\sum_{g \in G} \rho(g) = \sum_{g \in G} \rho(g_0 g) = \rho(g_0) \sum_{g \in G} \rho(g). \tag{2.2.4}$$

[5] Despite being widely used, we have not found a coherent proof of this proposition in any text. The proof given here is adapted from the proof of Lemma 15.3 of [Beh00], which is for the special case where G is abelian.

Consider V, the G-module that carries the representation ρ. It is straightforward to verify that $W = \{(\sum_{g\in G}\rho(g))(v) : v \in V\}$ is a G-submodule of V. Since ρ is irreducible, W must be either trivial or V itself. If W is trivial, then $\sum_{g\in G}\rho(g) = \mathbf{0}$, and if $W = V$, then $\sum_{g\in G}\rho(g)$ is invertible. But $\sum_{g\in G}\rho(g)$ cannot be invertible because $\rho(g_0) \neq I_{d_\rho}$ in (2.2.4), so it must be that $\sum_{g\in G}\rho(g) = \mathbf{0}$. □

Suppose that υ is symmetric. Furthermore, suppose that $\{X_k\}$ is aperiodic and irreducible, so that υ^{*k} converges to U_G. Applying Propositions 2.2.7-2.2.9 and the observation that $\hat{\mu}(\rho_{\text{triv}}) = 1$ for any μ to (2.2.3) gives the L^2 distance

$$\sum_{g\in G}\left(\upsilon^{*k}(g) - U_G(g)\right)^2 = \frac{1}{|G|}\sum_{\substack{\rho\in\hat{G}\\ \rho\neq\rho_{\text{triv}}}} d_\rho \mathrm{tr}[(\hat{\upsilon}(\rho))^{2k}] \tag{2.2.5}$$

between υ^{*k} and U_G.

For arbitrary $x_1, \ldots, x_j$, the Cauchy-Schwarz inequality implies that

$$(\sum_{i=1}^{j} x_i)^2 \leq j\sum_{i=1}^{j} x_i^2, \tag{2.2.6}$$

which, applied to Definition 2.1.5, gives that

$$4\|\mu - \nu\|_{\text{TV}}^2 = \left(\sum_{g\in G}|\mu(g) - \nu(g)|\right)^2 \leq |G|\sum_{g\in G}(\mu(g) - \nu(g))^2. \tag{2.2.7}$$

This extracts from (2.2.5) the upper bound

$$4\|\upsilon^{*k} - U_G\|_{\text{TV}}^2 \leq \sum_{\substack{\rho\in\hat{G}\\ \rho\neq\rho_{\text{triv}}}} d_\rho \mathrm{tr}[(\hat{\upsilon}(\rho))^{2k}] \tag{2.2.8}$$

for the total variation distance between υ^{*k} and U_G.

Before we work on the right hand sides of (2.2.5) and (2.2.8), let us note a couple of things. First of all, strictly speaking *our* Markov chain is not time-homogeneous. This is not a big deal: let υ_m be the uniform measure on the m-cycles of S_n and υ_2 be uniform on the transpositions, then the law μ_{k+1} of X_{k+1} is given by $\mu_{k+1} = \upsilon_2^{*k} * \upsilon_m$, with Fourier transform $\widehat{\mu_{k+1}} = \widehat{\upsilon_2}^k \widehat{\upsilon_m}$.

Secondly, the limiting distribution of our Markov chain is not uniform on the whole group S_n, but rather alternates between the uniform measure on A_n and the uniform measure on $S_n \backslash A_n$. This is slightly more problematic. Diaconis and Shahshahani, as well as most of those who followed, avoided parity by making their chain lazy. The trade-off is a small amount of precision in the ensuing computations. We consider instead the restrictions of the representations of S_n to A_n. The result is the following proposition, which we will prove at the end of the chapter:

Lemma 2.2.10. *Let μ be a measure on S_n with support in A_n, and let U be uniform on A_n. Then*

$$\sum_{g \in S_n} (\mu(g) - U(g))^2 = \frac{1}{n!} \sum_{\substack{\rho \in \widehat{S_n} \\ \rho \neq \rho_{\mathrm{triv}}, \rho_{\mathrm{sign}}}} d_\rho \mathrm{tr}[\hat{\mu}(\rho)(\hat{\mu}(\rho))^\dagger]. \tag{2.2.9}$$

The same holds if the support of μ is in $S_n \backslash A_n$ and U is uniform on $S_n \backslash A_n$.

Corollary 2.2.11. *With μ_k and U_k as defined in Chapter 1, we have the L^2 equality*

$$\sum_{g \in S_n} (\mu_{k+1}(g) - U_{k+1}(g))^2 = \frac{1}{n!} \sum_{\substack{\rho \in \widehat{S_n} \\ \rho \neq \rho_{\mathrm{triv}}, \rho_{\mathrm{sign}}}} d_\rho \mathrm{tr}[((\widehat{\upsilon_2}(\rho))^k \widehat{\upsilon_m}(\rho))^2] \tag{2.2.10}$$

and the total variation bound

$$4\|\mu_{k+1} - U_{k+1}\|_{\mathrm{TV}}^2 \leq \frac{1}{2} \sum_{\substack{\rho\in\widehat{S_n} \\ \rho\neq\rho_{\mathrm{triv}},\rho_{\mathrm{sign}}}} d_\rho \mathrm{tr}[((\widehat{\upsilon_2}(\rho))^k \widehat{\upsilon_m}(\rho))^2]. \tag{2.2.11}$$

Proof. Equation (2.2.10) is clear from Propositions 2.2.7-2.2.9 and Lemma 2.2.10. To see (2.2.11), observe that $\mu_{k+1}(g) - U_{k+1}(g) = 0$ for half of S_n, so

$$\left(\sum_{g\in S_n} |\mu_k(g) - U_k(g)|\right)^2 \leq \frac{n!}{2} \sum_{g\in S_n} (\mu_{k+1}(g) - U_{k+1}(g))^2 \tag{2.2.12}$$

by Cauchy-Schwarz. □

If G is abelian, then any irrep of G is 1-dimensional, so that the matrices $\upsilon(\rho)$ in (2.2.5) and (2.2.8) are all just scalars. Fortunately, even for a non-abelian G, a certain type of measures on G mimics measures on abelian groups.

Definition 2.2.12. A measure υ on G is a *class measure* if it is constant on the conjugacy classes of G. Note that class measures are clearly symmetric.

Lemma 2.2.13. *Let υ be a class measure. For every $\rho \in \hat{G}$, we have that*

$$\hat{\upsilon}(\rho) = \left(\frac{1}{d_\rho} \sum_g \upsilon(g)\chi_\rho(g)\right) I_{d_\rho}, \tag{2.2.13}$$

where $\chi_\rho(g) = \mathrm{tr}(\rho(g))$ is the character *of ρ at g.*

Proof. See Lemma 16.24 of [Beh00]. □

Remark. Since traces are similarity-invariant, $\chi_\rho(g) = \chi_\rho(h)$ whenever g and h are in the same conjugacy class. For elements of the symmetric group, this happens when g and h have the same cycle type.

Example (Diaconis and Shahshahani, [DS81]). Consider the (lazy) random transposition shuffle on n cards, the time-homogeneous Markov chain on S_n with increment measure υ that assigns mass $\frac{1}{n}$ to the identity and $\frac{2}{n^2}$ to each of the $\frac{n(n-1)}{2}$ transpositions τ. By Lemma 2.2.13,

$$(\hat{\upsilon}(\rho))^k = \left(\frac{1}{n} + \frac{(n-1)\chi_\rho(\tau)}{nd_\rho}\right)^k I_{d_\rho}, \tag{2.2.14}$$

which turns (2.2.8) into

$$4\|\mu_k - U\|_{\mathrm{TV}}^2 \leq \sum_{\substack{\rho \in \widehat{S_n} \\ \rho \neq \rho_{\mathrm{triv}}}} d_\rho^2 \left(\frac{1}{n} + \frac{(n-1)\chi_\rho(\tau)}{nd_\rho}\right)^{2k}. \tag{2.2.15}$$

The spectral interpretation of the right hand side of (2.2.15) is that the eigenvalues of the transition matrix associated with the shuffle are

$$\frac{1}{n} + \frac{(n-1)\chi_\rho(\tau)}{nd_\rho}, \ \rho \in \widehat{S_n}, \tag{2.2.16}$$

each occuring with multiplicity d_ρ^2. For more on the spectral theory of Markov chains, see Chapters 12 and 13 of [LPW08].

Analogously, for our Markov chain,

$$(\widehat{\upsilon_2}(\rho))^k \widehat{\upsilon_m}(\rho) = \left(\frac{\chi_\rho(\tau)}{d_\rho}\right)^k \left(\frac{\chi_\rho(\pi)}{d_\rho}\right) I_{d_\rho}, \tag{2.2.17}$$

where π is any m-cycle and τ is any transposition. Corollary 2.2.11 then gives

$$\sum_{g \in S_n} (\mu_{k+1}(g) - U_{k+1}(g))^2 = \frac{1}{n!} \sum_{\substack{\rho \in \widehat{S_n} \\ \rho \neq \rho_{\mathrm{triv}}, \rho_{\mathrm{sign}}}} d_\rho^2 \left(\frac{\chi_\rho(\tau)}{d_\rho}\right)^{2k} \left(\frac{\chi_\rho(\pi)}{d_\rho}\right)^2 \tag{2.2.18}$$

and

$$4\|\mu_{k+1} - U_{k+1}\|_{\mathrm{TV}}^2 \leq \frac{1}{2} \sum_{\substack{\rho \in \widehat{S_n} \\ \rho \neq \rho_{\mathrm{triv}}, \rho_{\mathrm{sign}}}} d_\rho^2 \left(\frac{\chi_\rho(\tau)}{d_\rho} \right)^{2k} \left(\frac{\chi_\rho(\pi)}{d_\rho} \right)^2. \tag{2.2.19}$$

Expressions of the form $\frac{\chi_\rho}{d_\rho}$ are called *normalized characters*. The next step is to compute the relevant ones of these for S_n.

2.3 Representation theory of S_n

We now turn our attention to the representations and characters of S_n. For a thorough introduction, see [Sag01] or Part I of [FH91].

Recall that $\hat{G}$ is, roughly speaking, a collection of the non-redundant irreducible representations of G. Such a collection is in general not unique, so it would be helpful to establish a canonical $\hat{G}$. It is well-known (e.g. see Proposition 1.10.1 of [Sag01]) that the number of equivalence classes of irreps is equal to the number of conjugacy classes of G. While an explicit correspondence has not been achieved for arbitrary groups, for S_n we can index both the conjugacy classes and the irreps with the partitions of n. As we describe below, the partitions of n give rise to a canonical $\widehat{S_n}$.

Definition 2.3.1. A *Young diagram* of size n is a configuration of n boxes, arranged in left-justified rows, such that the row lengths are weakly decreasing. For each partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$ of n, the Young diagram (of shape) λ contains λ_i

boxes in its i^{th} row.

Example. Figure 2.1 displays the Young diagrams corresponding to the partitions of 4.

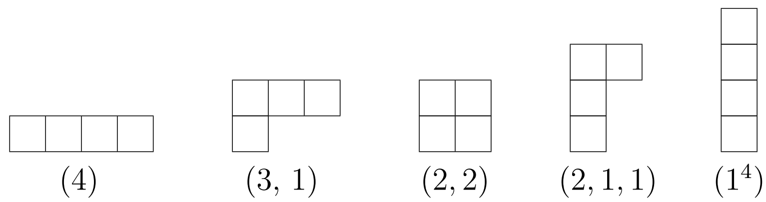

Figure 2.1: Young diagrams corresponding to the partitions of 4

Definition 2.3.2. Let $\lambda \vdash n$. A *Young tableau of shape* λ is obtained from the Young diagram of shape λ by filling its boxes with the numbers $1, 2, \ldots, n$ bijectively. A Young tableau is *standard* if the entries in each row and each column are increasing.

At this point we shall briefly describe the construction of *Specht modules*, which are indexed by partitions of n and form a complete set of irreps of S_n. See Chapter 2 of [Sag01] for the details.

Definition 2.3.3. Two Young tableaux t_1 and t_2 of the same shape are *row equivalent* if corresponding rows of the two tableaux contain the same elements. For a Young tableau t, the λ-*tabloid* $\{t\}$ is the set of all Young tableaux that are row equivalent to t.

A permutation σ acts on a Young tableau by replacing each number x in the tableau with $\sigma(x)$. This action gives rise to an S_n-module.

Definition 2.3.4. Let $\lambda \vdash n$. The vector space over $\mathbb{C}$ whose basis is the list of λ-tabloids, denoted as M^λ, is the *permutation module corresponding to* λ.

Definition 2.3.5. Suppose that the Young tableau t has columns $C_1, C_2, \ldots, C_k$. Then the *column-stabilizer* of t is

$$C_t = S_{C_1} \times S_{C_2} \times \ldots \times S_{C_k}, \tag{2.3.1}$$

i.e. the subgroup of S_n that permutes only the elements within each column of t.

Definition 2.3.6. For a Young tableau t, define $\kappa_t = \sum_{\sigma \in C_t} \operatorname{sign}(\sigma)\sigma$. Then the associated *polytabloid* of t is given by $e_t = \kappa_t\{t\}$.

Definition 2.3.7. For each partition λ, the corresponding *Specht module*, S^λ, is the submodule of M^λ spanned by all polytabloids e_t with t of shape λ.

Theorem 2.3.8. *The Specht modules S^λ for $\lambda \vdash n$ form a complete set of irreps of S_n over $\mathbb{C}$.*

Proof. See Theorem 2.4.6 of [Sag01]. □

We note here that $S^{(n)}$ is the trivial representation of S_n and that $S^{(1^n)}$ is the sign representation of S_n. These are the only canonical 1-dimensional representations of S_n. In general, the dimension of S^λ is the number of distinct standard Young tableaux of shape λ, which can be computed with the elegant hook length formula of Frame, Robinson, and Thrall:

Definition 2.3.9. Let (i, j) denote the j^{th} box in the i^{th} row of a Young diagram. Its *hook* is the set of all boxes directly below and directly to the right (including itself), i.e.

$$H_{i,j} = \{(i', j) : i' \geq i\} \cup \{(i, j') : j' \geq j\}, \tag{2.3.2}$$

with corresponding *hook length* $h_{i,j} = |H_{i,j}|$.

Theorem 2.3.10 (Hook length formula, [FRT54])**.** *For any partition λ of n,*

$$\dim S^\lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h_{i,j}}. \tag{2.3.3}$$

Proof. See Theorem 3.10.2 of [Sag01]. □

Example. Consider the Young diagram of shape $(4, 4, 3)$. On the left of Figure 2.2, the dotted boxes constitute the hook $H_{1,2}$. On the right, the number in each box is the length of the hook of the box, from which we see that the dimension of $S^{(4,4,3)}$ is $\frac{11!}{6 \cdot 5^2 \cdot 4^2 \cdot 3^2 \cdot 2^2 \cdot 1^2}$.

Figure 2.2: $H_{1,2}$ and the array of hook lengths for $(4, 4, 3)$

In addition to being useful for finding the dimensions of representations, Young diagrams are helpful for computing characters.

Definition 2.3.11. A *rim hook* ξ of a Young diagram λ is an edge-connected set of boxes, containing no subset of 2-by-2 blocks, that can be removed from λ to leave a proper Young diagram with the same top left corner as λ. The *leg length* of ξ, $ll(\xi)$, is the number of rows of ξ minus one.

Example. The top half of Figure 2.3 shows several rim hooks of $(4, 4, 3)$ along with their leg lengths, while the bottom half gives several non-examples of rim hooks.

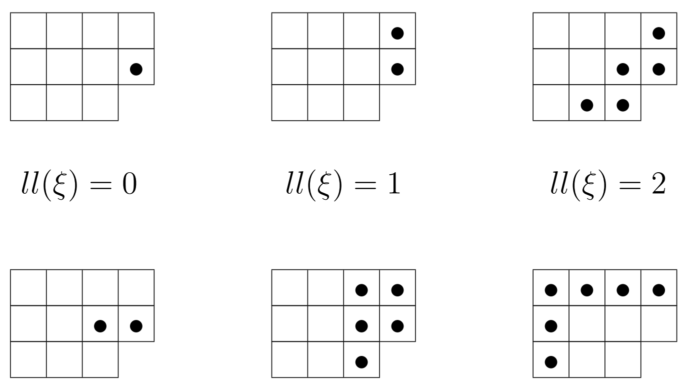

Figure 2.3: Examples and non-examples of rim hooks

We use $\lambda\backslash\xi$ to denote the Young diagram obtained from λ by removing the rim hook ξ. In the top right diagram of Figure 2.3, for instance, we have that $(4, 4, 3)\backslash\xi = (3, 2, 1)$. Also, for cycle type $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_r)$, we use the notation that $\gamma\backslash\gamma_1 = (\gamma_2, \ldots, \gamma_r)$. Moreover, we denote by χ_γ^λ the character of S^λ on the conjugacy class (of cycle type) γ.

Theorem 2.3.12 (Murnaghan-Nakayama rule, [Mur37] and [Nak40])**.** *If λ is a partition of n and γ is the cycle type of an element of S_n, then*

$$\chi_\gamma^\lambda = \sum_\xi (-1)^{ll(\xi)} \chi_{\gamma\backslash\gamma_1}^{\lambda\backslash\xi}, \tag{2.3.4}$$

where the sum is over all rim hooks ξ of λ with γ_1 boxes.

Proof. See Theorem 4.10.2 of [Sag01]. $\square$

Remark. This is a recursive formula. The first iteration is to remove from λ a rim hook with γ_1 boxes in all possible ways, the next iteration is to remove from each remaining diagram a rim hook with γ_2 boxes in all possible ways, and so on. The process terminates either when it is impossible to remove a rim hook of designated size, so that the contribution of the corresponding character is zero, or when all boxes have been deleted, leaving a contribution of ± 1.

Example. Figure 2.4 illustrates how to compute the character $\chi^{(4,4,3)}_{(5,4,2)}$ using the Murnaghan-Nakayama rule. The sign of the rim hook being removed (± 1 depending on $(-1)^{ll(\xi)}$, or 0 if no rim hook can be removed) is indicated below each diagram. We multiply together the signs along each path and add the products, so that

$$\chi^{(4,4,3)}_{(5,4,2)} = -\chi^{(4,2)}_{(4,2)} + \chi^{(3,2,1)}_{(4,2)} = (-1)^2\chi^{(1,1)}_{(2)} + 0 = (-1)^3 + 0 = -1. \qquad (2.3.5)$$

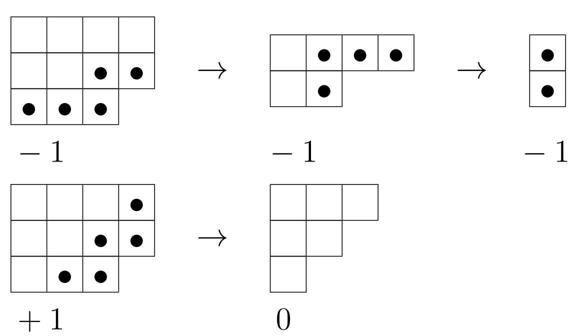

Figure 2.4: Computing $\chi^{(4,4,3)}_{(5,4,2)}$ with the Murnaghan-Nakayama rule

The stage is now set. We have the tools we need to compute the dimensions and characters of the representations of S_n. Before we do, however, we prove Lemma 2.2.10 as promised.

Definition 2.3.13. For a partition λ, its *conjugate partition*, λ', is the partition corresponding to the Young diagram obtained by switching the rows and columns of λ. If $\lambda = \lambda'$, then λ is said to be *self-conjugate*.

Example. The partitions $(4, 4, 3)$ and $(3, 3, 3, 2)$ are conjugates, and the partition $(4, 3, 3, 1)$ is self-conjugate.

Remark. Note that by the hook length formula, $\dim S^\lambda = \dim S^{\lambda'}$. Furthermore, 6.6 of [Jam78] implies that $\chi_\gamma^\lambda = \pm\chi_\gamma^{\lambda'}$, depending on the sign of γ.

There is a natural correspondence between the self-conjugate partitions of n and the conjugacy classes of S_n that split in A_n, which have cycle types with all odd cycle lengths: the cycle type $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_r)$ corresponds to the self-conjugate Young diagram whose diagonal boxes have hook lengths $\gamma_1, \gamma_2, \ldots, \gamma_r$. For instance, the cycle type $(7, 3, 1)$ corresponds to the partition $(4, 3, 3, 1)$.

Proposition 2.3.14. (1) *If λ is not self-conjugate, then $S^\lambda|_{A_n} = S^{\lambda'}|_{A_n}$, and this is irreducible as a representation of A_n.*

(2) *If λ is self-conjugate, then $S^\lambda|_{A_n} = \rho_1 \oplus \rho_2$, where ρ_1 and ρ_2 are irreps of A_n of dimension $\frac{\dim S^\lambda}{2}$. For conjugacy classes γ of S_n that do not correspond to λ as*

described above (even if γ also splits in A_n),

$$\chi_{\rho_1}(\gamma) = \chi_{\rho_2}(\gamma) = \frac{\chi_\gamma^{S^\lambda}}{2}. \tag{2.3.6}$$

For the conjugacy class $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_r)$ that corresponds to λ, let ξ and ξ' be the classes of A_n that it splits into, then $\chi_{\rho_1}(\xi) = \chi_{\rho_2}(\xi')$ and $\chi_{\rho_1}(\xi') = \chi_{\rho_2}(\xi)$, and furthermore these characters are given by

$$\frac{1}{2}\left((-1)^q \pm \sqrt{(-1)^q \gamma_1\gamma_2 \ldots \gamma_r}\right), \text{ where } q = \frac{n-r}{2}. \tag{2.3.7}$$

Proof. See Propositions 5.1 and 5.3 of [FH91]. □

Proof of Lemma 2.2.10. First, observe that $\hat{\mu}(S^{(1^n)})$ and $\hat{U}(S^{(1^n)})$ are equal to 1 if μ and U are supported on A_n and -1 if μ and U are supported on $S_n \backslash A_n$, so by (2.2.3) it suffices to show that $\hat{U}(S^\lambda) = \mathbf{0}$ for all $\lambda \neq (n), (1^n)$.

Suppose that λ is not self-conjugate. By the first part of Proposition 2.3.14, $S^\lambda|_{A_n}$ is a non-trivial irrep of A_n, so by Proposition 2.2.9 with $\rho = S^\lambda|_{A_n}$ and $G = A_n$, we have that $\sum_{g\in A_n} S^\lambda(g) = \mathbf{0}$. But Proposition 2.2.9 also implies that $\sum_{g\in S_n} S^\lambda(g) = \mathbf{0}$, so that $\sum_{g\in S_n\backslash A_n} S^\lambda(g) = \mathbf{0}$ as well! Thus $\hat{U}(S^\lambda) = \mathbf{0}$ whether U is uniform on A_n or on $S_n \backslash A_n$.

Now suppose that λ is self-conjugate. The second part of Proposition 2.3.14 tells us that $S^\lambda|_{A_n} = \rho_1 \oplus \rho_2$, where ρ_1 and ρ_2 are non-trivial irreps of A_n. Since $\sum_{g\in A_n} \rho_1(g) = \mathbf{0}$ and $\sum_{g\in A_n} \rho_2(g) = \mathbf{0}$, we again have that $\sum_{g\in A_n} S^\lambda(g) = \mathbf{0}$ and, analogously to above, that $\sum_{g\in S_n\backslash A_n} S^\lambda(g) = \mathbf{0}$. □

Chapter 3

Upper Bound

Einmal ist keinmal*, says Tomas to himself. What happens but once, says the German adage, might as well not have happened at all.*

– *The Unbearable Lightness Of Being*, "Lightness and Weight"

3.1 The $m = n$ case

The goal of this section is to prove the following:

Theorem 3.1.1. *For any $c > 0$, after one n-cycle and cn transpositions,*

$$4\|\mu_{cn+1} - U_{cn+1}\|_{\mathrm{TV}}^2 \leq \frac{e^{-4c}}{1-e^{-4c}} + o(1) \tag{3.1.1}$$

as n goes to infinity.

The first and most critical step of the proof is the observation that, discounting (n) and (1^n), $\chi^\lambda_{(n)} = 0$ for all λ except the L-shaped ones, for which $\lambda_2 = 1$. This is an almost trivial consequence of the Murnaghan-Nakayama rule, as it is impossible to remove a rim hook of size n from a Young diagram of size n unless the Young diagram itself is the rim hook; we will discuss later what this means probabilistically. Moreover, for an L-shaped λ, it is clear that $\chi^\lambda_{(n)}$ is equal to 1 if λ has an odd number of rows and -1 if λ has an even number of rows. Thus we arrive at a significant simplication of (2.2.19), namely that

$$4\|\mu_{k+1} - U_{k+1}\|^2_{\mathrm{TV}} \leq \frac{1}{2} \sum_{\lambda \in \Lambda_n} \left(\frac{\chi^\lambda_{(2,1^{n-2})}}{\dim S^\lambda} \right)^{2k}, \tag{3.1.2}$$

where

$$\Lambda_n = \{\lambda \vdash n : \lambda_1 > 1 \text{ and } \lambda_2 = 1\}. \tag{3.1.3}$$

The normalized characters $\frac{\chi^\lambda_{(2,1^{n-2})}}{\dim S^\lambda}$ have a simple description when $\lambda \in \Lambda_n$:

Proposition 3.1.2. *Let $\lambda \in \Lambda_n$, and let j be one less than the number of rows of λ. For $1 \leq j \leq \left\lfloor \frac{n-1}{2} \right\rfloor$,*

$$\frac{\chi^{(n-j,1^j)}_{(2,1^{n-2})}}{\dim S^{(n-j,1^j)}} = \frac{n-1-2j}{n-1}. \tag{3.1.4}$$

Proof. By the hook length formula,

$$\dim S^{(n-j,1^j)} = \frac{n!}{n \cdot j!(n-j-1)!} = \binom{n-1}{j}. \tag{3.1.5}$$

If $j > 1$, the first iteration of the Murnaghan-Nakayama rule, where we remove a rim hook with two boxes, results in

$$\chi^{(n-j,1^j)}_{(2,1^{n-2})} = \chi^{(n-j-2,1^j)}_{(1^{n-2})} - \chi^{(n-j,1^{j-2})}_{(1^{n-2})}. \tag{3.1.6}$$

Let $\tilde{n}$ be the number of remaining boxes, i.e. $n-2$. Observe that, for any partition $\tilde{\lambda}$ of $\tilde{n}$, the character of $\tilde{\lambda}$ at $(1^{\tilde{n}})$ is exactly the number of standard Young tableaux of shape $\tilde{\lambda}$, or the dimension of $\tilde{\lambda}$, which again can be computed with the hook length formula:

$$\chi^{(n-j-2,1^j)}_{(1^{n-2})} = \frac{(n-2)!}{(n-2)\cdot j!(n-j-3)!} = \binom{n-3}{j} \tag{3.1.7}$$

and

$$\chi^{(n-j,1^{j-2})}_{(1^{n-2})} = \frac{(n-2)!}{(n-2)\cdot (j-2)!(n-j-1)!} = \binom{n-3}{j-2}. \tag{3.1.8}$$

Putting (3.1.5)-(3.1.8) together and simplifying, we get that

$$\begin{aligned}
\frac{\chi^{(n-j,1^j)}_{(2,1^{n-2})}}{\dim S^{(n-j,1^j)}} &= \left(\frac{(n-3)!}{j!(n-j-3)!} - \frac{(n-3)!}{(j-2)!(n-j-1)!}\right)\cdot\frac{j!(n-j-1)!}{(n-1)!} \\
&= \frac{(n-3)![(n-j-1)(n-j-2)-j(j-1)]}{j!(n-j-1)!} \\
&\quad\cdot\frac{j!(n-j-1)!}{(n-1)!} \\
&= \frac{n^2-3n-2nj+4j+2}{(n-1)(n-2)} \\
&= \frac{(n-1-2j)(n-2)}{(n-1)(n-2)} = \frac{n-1-2j}{n-1}
\end{aligned} \tag{3.1.9}$$

for $j > 1$.

For $j=1$, $\dim S^{(n,1)} = n-1$, and since there is only one way to remove a rim hook of size two from $(n-1,1)$, we see that $\chi^{(n-1,1)}_{(2,1^{n-2})} = n-3$. □

Remark. When $\tilde{\lambda}$ is an L-shaped partition of $\tilde{n}$, we can actually skip the hook length formula and derive $\chi^{\tilde{\lambda}}_{(1^{\tilde{n}})}$ with the following simple combinatorial argument:

Let $\tilde{j}$ be one less than the number of boxes in the first column of $\tilde{\lambda}$. Removing one box at a time according to the Murnaghan-Nakayama rule, $\tilde{j}$ boxes in the first column are removed before we are left with a single row of boxes, at which point there is only one way to remove the remaining boxes. The number of ways to get to that point is the number of ways to pace the removal of the $\tilde{j}$ boxes throughout the removal of an overall $\tilde{n}-1$ boxes (the upper left box must be removed last), that is, $\binom{\tilde{n}-1}{\tilde{j}}$.

Proof of Theorem 3.1.1. Fix any $c > 0$. By calculus, for $n-1-2j > 0$,

$$\lim_{n\to\infty}\left(\frac{n-1-2j}{n-1}\right)^{2cn} = e^{-4cj}. \tag{3.1.10}$$

Thus Proposition 3.1.2 and the fact that $\chi^{\lambda}_{\gamma} = \pm\chi^{\lambda'}_{\gamma}$ imply that, for large n,

$$\sum_{\lambda\in\Lambda_n}\left(\frac{\chi^{\lambda}_{(2,1^{n-2})}}{\dim S^{\lambda}}\right)^{2cn} \sim \begin{cases} 2\displaystyle\sum_{j=1}^{(n-2)/2} e^{-4cj} & n \text{ is even} \\ 2\displaystyle\sum_{j=1}^{(n-3)/2} e^{-4cj} & n \text{ is odd.} \end{cases} \tag{3.1.11}$$

Summing the geometric series gives

$$4\|\mu_{cn+1} - U_{cn+1}\|^2_{\mathrm{TV}} \leq \frac{1}{2}\sum_{\lambda\in\Lambda_n}\left(\frac{\chi^{\lambda}_{(2,1^{n-2})}}{\dim S^{\lambda}}\right)^{2cn} \sim \frac{e^{-4c}}{1-e^{-4c}}, \tag{3.1.12}$$

as was to be shown. $\square$

3.2 The $m = n-1$ case

Next we prove an upper bound for the $m = n-1$ case.

Theorem 3.2.1. *For any $c > 0$, after one $(n-1)$-cycle and cn transpositions,*

$$4\|\mu_{cn+1} - U_{cn+1}\|_{\mathrm{TV}}^2 \leq \frac{e^{-8c}}{1-e^{-4c}} + o(1) \tag{3.2.1}$$

as n goes to infinity.

The proof is similar to the $m = n$ case. We start with the observation that $\chi_{(n-1,1)}^{\lambda} = 0$ for all λ except the ones with a 2-by-2 block of boxes in the upper left, for which $\lambda_2 = 2$ and $\lambda_3 = 0$ or 1 (see Figure 3.1).

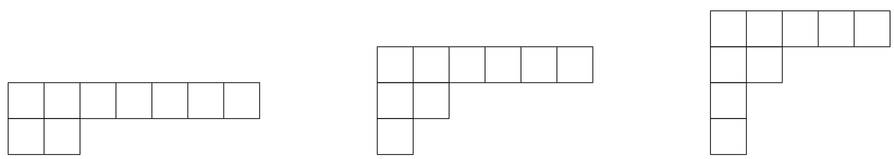

Figure 3.1: Examples of λ for which $\chi_{(n-1,1)}^{\lambda} \neq 0$

For such λ, we again have that $\chi_{(n-1,1)}^{\lambda} = \pm 1$, which gives

$$4\|\mu_{k+1} - U_{k+1}\|_{\mathrm{TV}}^2 \leq \frac{1}{2} \sum_{\lambda \in \Lambda_{n-1}} \left(\frac{\chi_{(2,1^{n-2})}^{\lambda}}{\dim S^{\lambda}} \right)^{2k}, \tag{3.2.2}$$

where

$$\Lambda_{n-1} = \{\lambda \vdash n : \lambda_2 = 2 \text{ and } \lambda_3 = 0 \text{ or } 1\}. \tag{3.2.3}$$

Proposition 3.2.2. *Let $\lambda \in \Lambda_{n-1}$, and let j be two less than the number of rows of λ. For $0 \leq j \leq \left\lfloor \frac{n-4}{2} \right\rfloor$,*

$$\frac{\chi_{(2,1^{n-2})}^{(n-2-j,2,1^j)}}{\dim S^{(n-2-j,2,1^j)}} = \frac{n-4-2j}{n}. \tag{3.2.4}$$

Proof. By the hook length formula,

$$\dim S^{(n-2-j,2,1^j)} = \frac{n!}{(n-1)(2+j)(n-2-j)\cdot j!(n-4-j)!}. \tag{3.2.5}$$

For $j = 0$, e.g. the leftmost diagram in Figure 3.1, there are two ways to remove a rim hook of size two: from the first row, or from the second. The latter leaves a single row and therefore contributes $+1$ to the value of $\chi^{(n-2,2)}_{(2,1^{n-2})}$, whereas the former contributes

$$\chi^{(n-4,2)}_{(1^{n-2})} = \frac{(n-2)!}{2(n-3)(n-4)\cdot(n-6)!} = \frac{(n-2)(n-5)}{2}. \tag{3.2.6}$$

Thus

$$\begin{aligned}\frac{\chi^{(n-2,2)}_{(2,1^{n-2})}}{\dim S^{(n-2,2)}} &= \frac{((n-2)(n-5)+2)}{2}\cdot\frac{2(n-1)(n-2)\cdot(n-4)!}{n!}\\ &= \frac{n^2-7n+12}{n(n-3)} = \frac{(n-4)(n-3)}{n(n-3)} = \frac{n-4}{n}.\end{aligned} \tag{3.2.7}$$

For $j = 1$, e.g. the middle diagram in Figure 3.1, there is only one way to remove a rim hook of size two, namely from the first row, so that

$$\begin{aligned}\chi^{(n-3,2,1)}_{(2,1^{n-2})} = \chi^{(n-5,2,1)}_{(1^{n-2})} &= \frac{(n-2)!}{3(n-3)(n-5)\cdot(n-7)!}\\ &= \frac{(n-2)(n-4)(n-6)}{3},\end{aligned} \tag{3.2.8}$$

and

$$\begin{aligned}\frac{\chi^{(n-3,2,1)}_{(2,1^{n-2})}}{\dim S^{(n-3,2,1)}} &= \frac{(n-2)(n-4)(n-6)}{3}\cdot\frac{3(n-1)(n-3)\cdot(n-5)!}{n!}\\ &= \frac{(n-2)(n-4)(n-6)}{n(n-2)(n-4)} = \frac{n-6}{n}.\end{aligned} \tag{3.2.9}$$

For $j > 1$, there are two ways to remove a rim hook of size two: from the first row, or from the first column. This implies that

$$\chi^{(n-2-j,2,1^j)}_{(2,1^{n-2})} = \chi^{(n-4-j,2,1^j)}_{(1^{n-2})} - \chi^{(n-2-j,2,1^{j-2})}_{(1^{n-2})}, \tag{3.2.10}$$

with

$$\chi^{(n-4-j,2,1^j)}_{(1^{n-2})} = \frac{(n-2)!}{(n-3)(2+j)(n-4-j)\cdot j!(n-6-j)!} \tag{3.2.11}$$

and

$$\chi^{(n-2-j,2,1^{j-2})}_{(1^{n-2})} = \frac{(n-2)!}{j(n-3)(n-2-j)\cdot (j-2)!(n-4-j)!}. \tag{3.2.12}$$

Combining and simplifying,

$$\begin{aligned}
\chi^{(n-2-j,2,1^j)}_{(2,1^{n-2})} &= \frac{(n-2)![j(n-2-j)(n-5-j) - j(2+j)(j-1)]}{j(n-3)(2+j)(n-2-j)\cdot j!(n-4-j)!} \\
&= \frac{(n-2)!j(n-3)(n-4-2j)}{j(n-3)(2+j)(n-2-j)\cdot j!(n-4-j)!} \\
&= \frac{(n-2)!(n-4-2j)}{(2+j)(n-2-j)\cdot j!(n-4-j)!},
\end{aligned} \tag{3.2.13}$$

and

$$\begin{aligned}
\frac{\chi^{(n-2-j,2,1^j)}_{(2,1^{n-2})}}{\dim S^{(n-2-j,2,1^j)}} &= \frac{(n-2)!(n-4-2j)}{(2+j)(n-2-j)\cdot j!(n-4-j)!} \\
&\quad \cdot \frac{(n-1)(2+j)(n-2-j)\cdot j!(n-4-j)!}{n!} \\
&= \frac{n-4-2j}{n},
\end{aligned} \tag{3.2.14}$$

as promised. □

Proof of Theorem 3.2.1. Fix any $c > 0$. For $n-4-2j > 0$,

$$\lim_{n\to\infty} \left(\frac{n-4-2j}{n}\right)^{2cn} = e^{-2c(4+2j)}, \tag{3.2.15}$$

and thus for large n,

$$\sum_{\lambda \in \Lambda_{n-1}} \left(\frac{\chi^{\lambda}_{(2,1^{n-2})}}{\dim S^{\lambda}} \right)^{2cn} \sim \begin{cases} 2 \sum_{j=0}^{(n-6)/2} e^{-2c(4+2j)} & n \text{ is even} \\ 2 \sum_{j=0}^{(n-5)/2} e^{-2c(4+2j)} & n \text{ is odd.} \end{cases} \tag{3.2.16}$$

This gives

$$4\|\mu_{cn+1} - U_{cn+1}\|^2_{\mathrm{TV}} \leq \frac{1}{2} \sum_{\lambda \in \Lambda_{n-1}} \left(\frac{\chi^{\lambda}_{(2,1^{n-2})}}{\dim S^{\lambda}} \right)^{2cn} \sim \frac{e^{-8c}}{1 - e^{-4c}} \tag{3.2.17}$$

as an upper bound. $\square$

We pause here for a few remarks. First, it is worth pointing out just how good Theorems 3.1.1 and 3.2.1 are, in the sense that the only source of inequality comes from Cauchy-Schwarz. This is the payoff of Lemma 2.2.10.

Secondly, the proofs of Propositions 3.1.2 and 3.2.2, while messy, are satisfying in that only the hook length formula and the Murnaghan-Nakayama rule are used. On the other hand, the results turned out to be essentially special cases of the identity

$$\frac{\chi^{\lambda}_{(2,1^{n-2})}}{\dim S^{\lambda}} = \frac{\sum_i (\lambda_i^2 - (2i-1)\lambda_i)}{n(n-1)}, \tag{3.2.18}$$

known as early as to Frobenius in [Fro00].

Thirdly, representation theory confirms what seems intuitive, that moving a lot of cards in the beginning leads to the cards being mixed sooner. In particular, the initial m-cycle promotes mixing by nullifying the contributions of some representations and lessening the contributions of the rest. However, we have also uncovered something counterintuitive, that moving $n-1$ cards in the beginning seems to lead to even faster mixing than moving all n cards! We will verify this and propose an explanation as we tackle the lower bound.

Chapter 4

Fixed Points and Lower Bound

A question is like a knife that slices through the stage backdrop and gives us a look at what lies hidden behind it.

– *The Unbearable Lightness Of Being*, "The Grand March"

4.1 Fixed points

For measures μ and ν on a set G, one of the classic approaches to finding a lower bound for $\|\mu - \nu\|_{\mathrm{TV}}$ is to identify a subset A of G where $|\mu(A) - \nu(A)|$ is close to maximal. In many mixing problems involving the symmetric group, it is convenient to make A either the set of fixed-point-free permutations or its complement, since it is well-known (e.g. to Montmort three centuries ago in [Mon08]) that the distribution of the number of fixed points with respect to the uniform measure on S_n is asymptotically $\mathcal{P}(1)$, the Poisson distribution of mean one.

Slightly less well-known[6], though unsurprising, is that the distribution of fixed points with respect to the uniform measure on either A_n or $S_n \backslash A_n$ is also asymptotically $\mathcal{P}(1)$. We will give an original proof for all of the Poisson limit laws mentioned here in Section 4.3. For a brute-force combinatorial proof of the weaker result that the mass, with respect to the uniform measure on S_n, A_n, as well as $S_n \backslash A_n$, of fixed-point-free permutations approaches $\frac{1}{e}$ as n approaches infinity, consult [AU08].

For Diaconis and Shahshahani's random transposition shuffle, A is the set of permutations with one or more fixed points, and finding $\mu_k(A)$ boils down to a coupon collector's problem. Let B be the event that, after k transpositions, at least one card is untouched. It is not difficult to see that $\mu_k(A) \geq \mathbf{P}(B)$, where $\mathbf{P}(B)$ is equal to the probability that at least one of n coupons is still missing after $2k$ trials. The coupon collector's problem is well-studied (see, for instance, Section IV.2 of [Fel68]), so this immediately gives a lower bound for $\mu_k(A)$, which in turn[7] produces a lower bound for $\|\mu_k(A) - U(A)\|_{\mathrm{TV}}$.

The above argument is so short and simple that it was tagged on to the end of the introduction of [DS81], as if an afterthought. Unfortunately, it is completely inapplicable to our problem, since the initial (large) m-cycle obliterates the core of the argument. To delve more deeply into the behavior of fixed points, we again turn to representation theory.

---

[6]We have not actually found this documented anywhere but presume that it is known.

[7]As $\mu_k(A) - U(A) \geq 0$, the inequality is in the desired direction.

4.2 The defining representation

Definition 4.2.1. The *defining*, or *permutation*, *representation* of S_n is the n-dimensional representation ϱ where $(\varrho(\sigma))_{i,j}$ is 1 if $\sigma(j) = i$ and 0 otherwise.

Example. For S_3,

$$\varrho(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \varrho(1,2) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\varrho(1,3) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \varrho(2,3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{4.2.1}$$

$$\varrho(1,2,3) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \varrho(1,3,2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

The significance of ϱ should be apparent: the fixed points can be read off of the matrix diagonal, so that $\chi_\varrho(\sigma)$ is precisely the number of fixed points of σ. We should also point out that ϱ is reducible and decomposes as $S^{(n-1,1)} \oplus S^{(n)}$ (see Examples 1.4.3, 1.9.5, and 2.3.8 of [Sag01]), so that the character of $S^{(n-1,1)}$ at σ is one less than the number of fixed points of σ. The representation $S^{(n-1,1)}$ is often referred to as the *standard representation* of S_n.

Heuristically, the connection between $S^{(n-1,1)}$ and fixed points vouches for the quality of the lower bound obtained via fixed points, since $S^{(n-1,1)}$ is in some sense the representation closest to the trivial representation and usually contributes the largest normalized character to the sum in (2.2.19). Moreover, this connection sheds light on why the $m = n-1$ case seems to converge faster: it is an atypical case where the contribution from $S^{(n-1,1)}$ is zero! Informally, this is the representation-theoretic analogue of the probabilistic intuition that, since the expected number of fixed points is one under the uniform distribution, a chain that starts with exactly one fixed point is closer to uniformity than a chain that starts with none.

Now, heuristics aside, we would like to find the mass of fixed-point-free permutations under μ_{k+1}. Since this cannot be done directly, we will in fact prove something more general: we will fully characterize the distribution of χ_ϱ with respect to μ_{k+1} by deriving all moments of χ_ϱ with respect to μ_{k+1}. The pivotal observation, inspired by Remark 1 in Chapter 3D of [Dia88], is the following, which relates raw moments of the fixed point distribution to tensor powers of ϱ:

Proposition 4.2.2. *Let E_μ denote expectation with respect to μ, and let $a_{\lambda,r}$ be the multiplicity of S^λ in the decomposition of $\varrho^{\otimes r}$ into a direct sum of irreducible representations, i.e. let*

$$\varrho^{\otimes r} = \bigoplus_{\lambda \vdash n} a_{\lambda,r} S^\lambda := \bigoplus_{\lambda \vdash n} (S^\lambda)^{\oplus a_{\lambda,r}}. \tag{4.2.2}$$

Then, viewing χ_ϱ as a random variable on S_n,

$$E_\mu((\chi_\varrho)^r) = \sum_{\lambda \vdash n} a_{\lambda,r} \mathrm{tr}(\hat{\mu}(S^\lambda)) \tag{4.2.3}$$

for any positive integer r.

Proof. Since the tensor product has the property that the trace of the product is equal to the product of the traces,

$$\begin{aligned}
E_\mu((\chi_\varrho)^r) &= \sum_{\sigma\in S_n} \mu(\sigma)[\mathrm{tr}(\varrho(\sigma))]^r = \sum_{\sigma\in S_n} \mu(\sigma)\mathrm{tr}(\varrho^{\otimes r}(\sigma)) \\
&= \sum_{\sigma\in S_n} \mu(\sigma)\mathrm{tr}\left(\bigoplus_{\lambda\vdash n} a_{\lambda,r} S^\lambda(\sigma)\right) = \sum_{\sigma\in S_n}\left(\mu(\sigma)\sum_{\lambda\vdash n} a_{\lambda,r}\mathrm{tr}(S^\lambda(\sigma))\right) \quad (4.2.4)\\
&= \sum_{\lambda\vdash n}\left(a_{\lambda,r}\sum_{\sigma\in S_n}\mu(\sigma)\mathrm{tr}(S^\lambda(\sigma))\right) = \sum_{\lambda\vdash n} a_{\lambda,r}\mathrm{tr}(\hat{\mu}(S^\lambda)),
\end{aligned}$$

where the last equality is due to the linearity of the trace. □

Remark. The first line of (4.2.4) is clearly true for any representation ρ of S_n, and hence the equality $E_\mu((\chi_\rho)^r) = \mathrm{tr}(\hat{\mu}(\rho^{\otimes r}))$ holds for all ρ.

Recall that by (2.2.17),

$$\mathrm{tr}(\widehat{\mu_{k+1}}(S^\lambda)) = \mathrm{tr}[(\widehat{\upsilon_2}(S^\lambda))^k\widehat{\upsilon_m}(S^\lambda)] = \chi^\lambda_{(m,1^{n-m})}\left(\frac{\chi^\lambda_{(2,1^{n-2})}}{\dim S^\lambda}\right)^k, \tag{4.2.5}$$

which we have already computed for all λ in the $m = n$ and $m = n-1$ cases while working on the upper bound! Thus in light of Proposition 4.2.2, if we find $a_{\lambda,r}$ for all λ and r, then we would know all moments of χ_ϱ with respect to μ_{k+1}.

We shall do just that.

Theorem 4.2.3. *Let $\lambda\vdash n$ and $1\le r\le n-\lambda_2$. The multiplicity of S^λ in the irrep decomposition of $\varrho^{\otimes r}$ is given by*

$$a_{\lambda,r} = f^{\bar{\lambda}}\sum_{i=|\bar{\lambda}|}^{r}\binom{i}{|\bar{\lambda}|}\left\{ {r \atop i} \right\}, \tag{4.2.6}$$

where $\bar{\lambda}$ is the truncated partition $(\lambda_2,\dots,\lambda_l)$ of weight $|\bar{\lambda}|$, $f^{\bar{\lambda}}$ is the number of standard Young tableaux of shape $\bar{\lambda}$, and $\left\{ {r \atop i} \right\}$ is a Stirling number of the second *kind, i.e. the number of ways to partition r objects into i non-empty subsets.*

Remark. Since $f^{\bar{\lambda}} = \dim S^{(\lambda_2,\ldots,\lambda_l)}$ can be computed with the hook length formula and the Stirling numbers can be explicitly defined as

$$\left\{ {r \atop i} \right\} = \frac{1}{i!} \sum_{j=1}^{i} (-1)^{i-j} \binom{i}{j} j^r, \tag{4.2.7}$$

we can rewrite (4.2.6) as

$$a_{\lambda,r} = \frac{n!}{\prod_{(x,y)\in\lambda} h_{x,y}} \sum_{i=|\bar{\lambda}|}^{r} \left(\binom{i}{|\bar{\lambda}|} \frac{1}{i!} \sum_{j=1}^{i} (-1)^{i-j} \binom{i}{j} j^r \right), \tag{4.2.8}$$

which defines $a_{\lambda,r}$ in terms of elementary expressions and factorials.

Proof. Theorem 4.2.3 owes its existence to the recent work of Goupil and Chauve, who derived in [GC06] the generating function

$$\sum_{r\geq|\bar{\lambda}|} a_{\lambda,r} \frac{x^r}{r!} = \frac{f^{\bar{\lambda}}}{|\bar{\lambda}|!} e^{e^x-1} (e^x-1)^{|\bar{\lambda}|} \tag{4.2.9}$$

for $\lambda \vdash n$ and $n \geq r + \lambda_2$.

By (24b) and (24f) in Chapter 1 of [Sta97],

$$\sum_{s\geq j} \left\{ {s \atop j} \right\} \frac{x^s}{s!} = \frac{(e^x-1)^j}{j!} \tag{4.2.10}$$

and

$$\sum_{t\geq 0} B_t \frac{x^t}{t!} = e^{e^x-1}, \tag{4.2.11}$$

where $B_0 := 1$ and $B_t = \sum_{q=1}^{t} \left\{ {t \atop q} \right\}$ is the t-th Bell number, so we obtain from (4.2.9) that

$$\frac{a_{\lambda,r}}{r!} = f^{\bar{\lambda}} \sum_{s+t=r} \frac{B_t}{s!t!} \left\{ \begin{matrix} s \\ |\bar{\lambda}| \end{matrix} \right\}, \tag{4.2.12}$$

and thus

$$\begin{aligned} \frac{a_{\lambda,r}}{f^{\bar{\lambda}}} &= \sum_{t=0}^{r-|\bar{\lambda}|} B_t \binom{r}{t} \left\{ \begin{matrix} r-t \\ |\bar{\lambda}| \end{matrix} \right\} \\ &= \left\{ \begin{matrix} r \\ |\bar{\lambda}| \end{matrix} \right\} + \sum_{t=1}^{r-|\bar{\lambda}|} \sum_{q=1}^{t} \left\{ \begin{matrix} t \\ q \end{matrix} \right\} \binom{r}{t} \left\{ \begin{matrix} r-t \\ |\bar{\lambda}| \end{matrix} \right\} \\ &= \left\{ \begin{matrix} r \\ |\bar{\lambda}| \end{matrix} \right\} + \sum_{q=1}^{r-|\bar{\lambda}|} \sum_{t=q}^{r-|\bar{\lambda}|} \left\{ \begin{matrix} t \\ q \end{matrix} \right\} \binom{r}{t} \left\{ \begin{matrix} r-t \\ |\bar{\lambda}| \end{matrix} \right\}. \end{aligned} \tag{4.2.13}$$

By (24.1.3, II.A) of [AS65],

$$\sum_{t=q}^{r-|\bar{\lambda}|} \left\{ \begin{matrix} t \\ q \end{matrix} \right\} \binom{r}{t} \left\{ \begin{matrix} r-t \\ |\bar{\lambda}| \end{matrix} \right\} = \binom{q+|\bar{\lambda}|}{|\bar{\lambda}|} \left\{ \begin{matrix} r \\ q+|\bar{\lambda}| \end{matrix} \right\}, \tag{4.2.14}$$

so that

$$\begin{aligned} \frac{a_{\lambda,r}}{f^{\bar{\lambda}}} &= \left\{ \begin{matrix} r \\ |\bar{\lambda}| \end{matrix} \right\} + \sum_{q=1}^{r-|\bar{\lambda}|} \binom{q+|\bar{\lambda}|}{|\bar{\lambda}|} \left\{ \begin{matrix} r \\ q+|\bar{\lambda}| \end{matrix} \right\} \\ &= \left\{ \begin{matrix} r \\ |\bar{\lambda}| \end{matrix} \right\} + \sum_{i=|\bar{\lambda}|+1}^{r} \binom{i}{|\bar{\lambda}|} \left\{ \begin{matrix} r \\ i \end{matrix} \right\} = \sum_{i=|\bar{\lambda}|}^{r} \binom{i}{|\bar{\lambda}|} \left\{ \begin{matrix} r \\ i \end{matrix} \right\}, \end{aligned} \tag{4.2.15}$$

as we rejoice. □

On a side note, let $b_{\lambda,r}$ be the multiplicity of S^λ in the irrep decomposition of $(S^{(n-1,1)})^{\otimes r}$, so that

$$(S^{(n-1,1)})^{\otimes r} = \bigoplus_{\lambda \vdash n} b_{\lambda,r} S^\lambda := \bigoplus_{\lambda \vdash n} (S^\lambda)^{\oplus b_{\lambda,r}}. \tag{4.2.16}$$

Goupil and Chauve also derived the generating function

$$\sum_{r\geq|\bar{\lambda}|} b_{\lambda,r}\frac{x^r}{r!} = \frac{f^{\bar{\lambda}}}{|\bar{\lambda}|!}e^{e^x-x-1}(e^x-1)^{|\bar{\lambda}|}, \tag{4.2.17}$$

so from Theorem 4.2.3 we can obtain a decent formula for the irrep decomposition of $(S^{(n-1,1)})^{\otimes r}$ as well.

Corollary 4.2.4. *Let $\lambda \vdash n$ and $1 \leq r \leq n-\lambda_2$. The multiplicity of S^λ in the irrep decomposition of $(S^{(n-1,1)})^{\otimes r}$ is given by*

$$b_{\lambda,r} = f^{\bar{\lambda}}\sum_{s=|\bar{\lambda}|}^{r}(-1)^{r-s}\binom{r}{s}\left(\sum_{i=|\bar{\lambda}|}^{s}\binom{i}{|\bar{\lambda}|}\left\{{s \atop i}\right\}\right). \tag{4.2.18}$$

Proof. Comparing (4.2.17) with (4.2.9) gives

$$\sum_{r\geq|\bar{\lambda}|} b_{\lambda,r}\frac{x^r}{r!} = \left(\sum_{s\geq|\bar{\lambda}|} a_{\lambda,s}\frac{x^s}{s!}\right)e^{-x} = \left(\sum_{s\geq|\bar{\lambda}|} a_{\lambda,s}\frac{x^s}{s!}\right)\left(\sum_{t\geq 0}\frac{(-x)^t}{t!}\right), \tag{4.2.19}$$

so that

$$\frac{b_{\lambda,r}}{r!} = \sum_{s+t=r}\frac{(-1)^t a_{\lambda,s}}{s!t!} = \sum_{s=|\bar{\lambda}|}^{r}\frac{(-1)^{r-s}}{s!(r-s)!}\left(f^{\bar{\lambda}}\sum_{i=|\bar{\lambda}|}^{s}\binom{i}{|\bar{\lambda}|}\left\{{s \atop i}\right\}\right), \tag{4.2.20}$$

and the result follows. $\square$

Remark. Corollary 4.2.4 is very similar to Proposition 2 of [GC06], but our result is a bit cleaner, as it does not involve associated Stirling numbers of the second kind.

4.3 Lower bound for the $m = n$ case

Before unleashing the power of Theorem 4.2.3, we need to clear up a technicality: not all probability distributions are uniquely determined by their moments. For instance, a distribution all of whose moments match those of the log-normal is not necessarily log-normal. Fortunately, there is a simple sufficient condition for uniqueness.

Theorem 4.3.1. *Let m_r denote the r-th moment of the distribution of a random variable Y. If the moment-generating function $\mathbf{E}(e^{tY}) = \sum_{r\geq 0} m_r \frac{t^r}{r!}$ has a positive radius of convergence, then there is no other distribution with the same moments.*

Proof. See Theorem 30.1 of [Bil95]. □

In Theorems 4.3.3 and 4.3.4, we will argue that a sequence of distributions converges to a Poisson. By definition, the moment-generating function for a Poisson of mean ν is

$$\sum_{j\geq 0} e^{tj} \frac{\nu^j e^{-\nu}}{j!} = e^{-\nu} \sum_{j\geq 0} \frac{(e^t \nu)^j}{j!} = e^{-\nu} e^{e^t \nu} = e^{\nu(e^t - 1)}, \tag{4.3.1}$$

which satisfies the uniqueness condition, so Poisson distributions are indeed determined by their moments.

Theorem 4.3.2. *Suppose that the distribution of Y is determined by its moments, that Y has moments of all orders, and that $\mathbf{E}(Y_i^r)$ tends to $\mathbf{E}(Y^r)$ for all r, then Y_i converges in distribution to Y.*

Proof. See Theorem 30.2 of [Bil95]. □

We are now ready to prove several Poisson limit laws. First, as promised, we give a new proof for an ancient result:

Theorem 4.3.3. (1) *Let U_{S_n} denote the uniform measure on S_n. As n approaches infinity, the distribution of the number of fixed points of a permutation randomly chosen according to U_{S_n} converges to $\mathcal{P}(1)$.*

(2) *The above statement holds if we replace S_n with either A_n or $S_n \backslash A_n$.*

Proof. For (1), recall Proposition 2.2.9, which implies that $\widehat{U_{S_n}}(S^\lambda)$ is 1 if $\lambda = (n)$ and 0 otherwise. Thus the combination of Proposition 4.2.2 and Theorem 4.2.3 gives that, for $1 \leq r \leq n$,

$$E_{U_{S_n}}((\chi_\varrho)^r) = a_{(n),r} = \sum_{i=0}^{r} \left\{ {r \atop i} \right\} = B_r, \tag{4.3.2}$$

which by (4.2.11) and (4.3.1) is exactly the r-th moment of $\mathcal{P}(1)$. This means that the first n moments of χ_ϱ with respect to U_{S_n} match those of $\mathcal{P}(1)$, and therefore convergence follows from Theorem 4.3.2.

For (2), recall from the proof of 2.2.10 that $\widehat{U_{A_n}}(S^\lambda)$ is 1 if λ is (n) or (1^n) and 0 otherwise. Moreover, $\widehat{U_{S_n \backslash A_n}}(S^\lambda)$ is 1 if $\lambda = (n)$, -1 if $\lambda = (1^n)$, and 0 otherwise. Hence

$$\begin{aligned} E_{U_{A_n}}((\chi_\varrho)^r) &= a_{(n),r} + a_{(1^n),r} \\ \text{and} \quad E_{U_{S_n \backslash A_n}}((\chi_\varrho)^r) &= a_{(n),r} - a_{(1^n),r}. \end{aligned} \tag{4.3.3}$$

As before, $a_{(n),r} = B_r$ for $1 \leq r \leq n$. Meanwhile, for $1 \leq r \leq n-1$,

$$a_{(1^n),r} = \sum_{i=n-1}^{r} \binom{i}{n-1} \left\{ {r \atop i} \right\}, \tag{4.3.4}$$

which is 0 for $1 \leq r \leq n-2$. Thus the first $n-2$ moments of χ_ϱ with respect to either U_{A_n} or $U_{S_n \setminus A_n}$ match those of $\mathcal{P}(1)$. $\square$

Returning to the Markov chain mixing rate problem, the next Poisson limit law will finally give a satisfactory lower bound for the mixing rate of the n-cycle-to-transpositions shuffle.

Theorem 4.3.4. *Fix any $c > 0$. As n approaches infinity, the distribution of the number of fixed points after one n-cycle and cn transpositions converges to $\mathcal{P}(1 - e^{-2c})$.*

Proof. One can deduce from the moment-generating function, or just look up in [Rio37], that the r-th moment of $\mathcal{P}(\nu)$ is $\sum_{i=1}^{r} \left\{ {r \atop i} \right\} \nu^i$. As it went with the proof of Theorem 4.3.3, $\widehat{\mu_{cn+1}}(S^{(n)}) = 1$, and we will ignore the alternating representation because it suffices to consider the first $n-2$ moments. For the non-trivial and non-alternating representations, we take advantage of previous computations and synthesize Proposition 3.1.2, (3.1.10) with n instead of $2n$, and (4.2.5) with the recollection that $\chi^\lambda_{(n)} = (-1)^{|\bar{\lambda}|}$ to obtain

$$\widehat{\mu_{cn+1}}(S^\lambda) \sim \begin{cases} (-1)^{|\bar{\lambda}|} e^{-2c|\bar{\lambda}|} & \lambda \in \Lambda_n \\ 0 & \text{otherwise.} \end{cases} \tag{4.3.5}$$

By Theorem 4.2.3 (second line below) and (4.3.5) (fourth line), for $1 \le r \le n-2$,

$$\begin{aligned}
E_{\mu_{cn+1}}((\chi_\varrho)^r) &= a_{(n),r} + \sum_{\lambda \in \Lambda_n} a_{\lambda,r} \widehat{\mu_{cn+1}}(S^\lambda) \\
&= \sum_{i=1}^{r} \left\{ {r \atop i} \right\} + \sum_{|\bar\lambda|=1}^{n-2} \sum_{i=|\bar\lambda|}^{r} \left\{ {r \atop i} \right\} \binom{i}{|\bar\lambda|} \widehat{\mu_{cn+1}}(S^\lambda) \\
&= \sum_{i=1}^{r} \left\{ {r \atop i} \right\} + \sum_{i=1}^{r} \sum_{|\bar\lambda|=1}^{i} \left\{ {r \atop i} \right\} \binom{i}{|\bar\lambda|} \widehat{\mu_{cn+1}}(S^\lambda) \\
&\sim \sum_{i=1}^{r} \left\{ {r \atop i} \right\} + \sum_{i=1}^{r} \sum_{|\bar\lambda|=1}^{i} \left\{ {r \atop i} \right\} \binom{i}{|\bar\lambda|} (-e^{-2c})^{|\bar\lambda|} \\
&= \sum_{i=1}^{r} \left\{ {r \atop i} \right\} \left(1 + \sum_{|\bar\lambda|=1}^{i} \binom{i}{|\bar\lambda|} (-e^{-2c})^{|\bar\lambda|} \right) \\
&= \sum_{i=1}^{r} \left\{ {r \atop i} \right\} (1 - e^{-2c})^i.
\end{aligned} \tag{4.3.6}$$

This shows that the first $n-2$ moments of χ_ϱ with respect to μ_{cn+1} approach those of $\mathcal{P}(1 - e^{-2c})$, and once again convergence follows from Theorem 4.3.2. $\square$

Corollary 4.3.5. *For any $c > 0$, after one n-cycle and cn transpositions,*

$$\|\mu_{cn+1} - U_{cn+1}\|_{\mathrm{TV}} \ge \frac{e^{-2c}}{e} - o(1) \tag{4.3.7}$$

as n goes to infinity.

Proof. Let A be the set of fixed-point-free permutations. Then

$$\begin{aligned}
\|\mu_{cn+1} - U_{cn+1}\|_{\mathrm{TV}} &\ge |\mu_{cn+1}(A) - U_{cn+1}(A)| \\
&\sim e^{e^{-2c}-1} - \frac{1}{e} = \frac{1}{e}\left(e^{-2c} + \frac{(e^{-2c})^2}{2!} + \cdots \right) \ge \frac{e^{-2c}}{e},
\end{aligned} \tag{4.3.8}$$

as was to be shown. $\square$

Remark. Together with Theorem 3.1.1, we have that

$$\frac{e^{-2c}}{e} - o(1) \leq \|\mu_{cn+1} - U_{cn+1}\|_{\mathrm{TV}} \leq \frac{e^{-2c}}{2\sqrt{1-e^{-4c}}} + o(1) \tag{4.3.9}$$

as n goes to infinity. The gap is especially respectable if e^{-4c} is small. Also, recall from Theorem 3.2.1 that an upper bound for the $m = n-1$ case is

$$\|\mu_{cn+1} - U_{cn+1}\|_{\mathrm{TV}} \leq \frac{e^{-4c}}{2\sqrt{1-e^{-4c}}} + o(1), \tag{4.3.10}$$

which is smaller than even the lower bound for the $m = n$ case so long as c is at least approximately 0.262. Hence we can reasonably say that starting with an $(n-1)$-cycle is indeed more beneficial for mixing than starting with an n-cycle.

Chapter 5

Further Considerations

[V]ertigo is something other than the fear of falling. It is the voice of the emptiness below us which tempts and lures us, it is the desire to fall, against which, terrified, we defend ourselves.

– *The Unbearable Lightness of Being*, "Soul and Body"

5.1 Miscellaneous results

In this section we use $S^{(n-1,1)}$ to derive two more results about expected numbers of fixed points. Recall that the character of $S^{(n-1,1)}$ at σ is one less than the number of fixed points of σ.

First, we present the following martingale-like property for Markov chains on S_n whose increment distributions are class measures: if a chain starts with one fixed point, then it will always average exactly one fixed point.

Proposition 5.1.1. *Let X_0 be the identity, and set $X_1 = \tau_1 X_0$, where τ_1 is selected according to any class measure supported on the set of permutations with one fixed* *point. For $k \geq 2$, set $X_k = \tau_k X_{k-1}$, where τ_k is selected according to any class measure on S_n (the measure can be different for each k). Then the expected number of fixed points of X_k is one for all $k \geq 1$.*

Proof. Let ν_1 be a class measure supported on the set of permutations with one fixed point, $\nu_2, \nu_3, \ldots, \nu_k$ be class measures on S_n, and define $\mu_k = \nu_k * \cdots * \nu_2 * \nu_1$. By the remark following Proposition 4.2.2,

$$\begin{aligned} E_{\mu_k}(\chi_{S^{(n-1,1)}}) &= \operatorname{tr}[\widehat{\mu_k}(S^{(n-1,1)})] \\ &= \operatorname{tr}[\widehat{\nu_1}(S^{(n-1,1)})\widehat{\nu_2}(S^{(n-1,1)}) \cdots \widehat{\nu_k}(S^{(n-1,1)})], \end{aligned} \tag{5.1.1}$$

where

$$\widehat{\nu_1}(S^{(n-1,1)}) = \left(\frac{1}{n-1} \sum_{\sigma \in S_n} \nu_1(\sigma) \chi_{S^{(n-1,1)}}(\sigma) \right) I_{n-1} \tag{5.1.2}$$

by Lemma 2.2.13. Consider the anatomy of the partition $(n-1,1)$: under the Murnaghan-Nakayama rule, the only way for a single box to remain at the end is for the box in the second row to have been removed as a singleton, which requires a cycle type with at least two fixed points. This means that $\chi_{S^{(n-1,1)}}(\sigma) = 0$ if σ has exactly one fixed point. On the other hand, if σ does not have exactly one fixed point, then $\nu_1(\sigma) = 0$. Thus $\widehat{\nu_1}(S^{(n-1,1)}) = \mathbf{0}$, which in turn implies that $E_{\mu_k}(\chi_{S^{(n-1,1)}}) = 0$, and hence the expected number of fixed points with respect to μ_k is one for all $k \geq 1$. $\square$

Returning to the m-cycle-to-transpositions chain, we can now characterize the expected number of fixed points for the general case where m is defined by an arbitrary function $m(n)$ of n.

Theorem 5.1.2. *After one $m(n)$-cycle and k transpositions,*

$$E_{\mu_{k+1}}(\chi_\varrho)\begin{cases}\sim 1-e^{-2c} & m(n)=n, k=cn\\[2ex] =1 & m(n)=n-1,\ \text{any } k\\[2ex] \sim 1+e^{-2c} & m(n)\neq n, n-1,\\ & k=cn+\frac{n}{2}\ln(n-m(n)-1),\end{cases}\tag{5.1.3}$$

where, as in Chapter 4, $\varrho = S^{(n-1,1)} \oplus S^{(n)}$.

Proof. The $m(n)=n$ (Theorem 4.3.4) and $m(n)=n-1$ (special case of Proposition 5.1.1) cases have already been shown. For $m(n)\neq n, n-1$,

$$\chi^{(n-1,1)}_{(m(n),1^{n-m(n)})} = n-m(n)-1,\tag{5.1.4}$$

so by (4.2.5),

$$\begin{aligned}E_{\mu_k}(\chi_{S^{(n-1,1)}}) &= \chi^{(n-1,1)}_{(m(n),1^{n-m(n)})}\left(\frac{\chi^{(n-1,1)}_{(2,1^{n-2})}}{\dim S^{(n-1,1)}}\right)^k\\ &= (n-m(n)-1)\left(\frac{n-3}{n-1}\right)^k.\end{aligned}\tag{5.1.5}$$

Setting $k=cn+\frac{n}{2}\ln(n-m(n)-1)$, we have that

$$\begin{aligned}\lim_{n\to\infty} E_{\mu_k}(\chi_{S^{(n-1,1)}}) &= \lim_{n\to\infty}(n-m(n)-1)\left(\frac{n-3}{n-1}\right)^{cn+\frac{n}{2}\ln(n-m(n)-1)}\\ &=(n-m(n)-1)e^{-2c}e^{-\ln(n-m(n)-1)} = e^{-2c},\end{aligned}\tag{5.1.6}$$

and the result follows. □

5.2 Open questions

We conclude with a list of open questions related to our work.

Question (1)**.** What is the lower bound for the mixing time of the $m = n - 1$ case of the m-cycle-to-transpositions chain?

The $m = n - 1$ case is trickier than the $m = n$ case because, unlike the $m = n$ case, the distribution of the number of fixed points is not quite Poisson. Indeed, after an initial $(n - 1)$-cycle, the expected number of fixed points is always one. On the other hand, we can compute from either Corollary 4.3.5 or, as a more fun exercise, Proposition 1 of [GC06] that

$$S^{(n-1,1)} \otimes S^{(n-1,1)} = S^{(n)} \oplus S^{(n-1,1)} \oplus S^{(n-2,2)} \oplus S^{(n-2,1,1)}, \tag{5.2.1}$$

which along with Proposition 3.2.2 implies that the variance of the number of fixed points after one $(n-1)$-cycle and cn transpositions is asymptotically $1-e^{-4c}$. As the mean does not match the variance, the distribution is not Poisson. Nevertheless, it must be very close to Poisson, and one may be able to compute a few more moments and use brute force to bound the mass of fixed-point-free permutations, which in turn will give a lower bound for the mixing time.

In general, when finding a lower bound for the mixing time of a Markov chain, the method of moments is powerful because it produces robust results without relying on convenient but narrowly-scoped combinatorial arguments. Theorem 4.2.3, in particular, enables a wide class of Markov chains on S_n to be analyzed this way. On the other hand, we got lucky with Theorems 4.3.3 and 4.3.4 in the sense that we happened to recognize each sequence of moments as that of a Poisson. When the moments do not match up with those of any well-known distributions, there is the additional task of extracting information about the distribution from its moments.[8]

Question (2)**.** In the general case where m is an arbitrary function $m(n)$ of n (excluding n and $n-1$), is $n\ln(n-m(n)-1)$ the correct order of mixing time?

This question is motivated by Theorem 5.1.2. To see that it is at least plausible, consider that $O(n\ln(n-m(n)-1))$ is the right mixing time for $m(n)=2$, i.e. the random transposition shuffle. Proving a general upper bound with only the techniques from this thesis is likely difficult, but the order of a lower bound may be within reach. In particular, from (5.2.1) we should be able to completely charaterize the variance of the number of fixed points for arbitrary $m(n)$ like Theorem 5.1.2 did for the expected value. If we have both the first and the second moments, then we may be able to derive the order of a lower bound using Chebyshev's inequality or Proposition 7.8 of [LPW08], a method of procuring lower bounds from distinguishing statistics.

Question (3)**.** For a Markov chain on S_n whose increment distributions are class measures, what conditions are sufficient for its fixed point distribution to be asymptotically (in n) Poisson?

[8] The classical moment problem is oft-studied, but predominantly for determinancy conditions, and most of the work on reconstruction has been for continuous distributions. See the introduction of [MH09] for a survey of results.

A necessary condition appears to be that the initial step does not create exactly one fixed point. Is it also sufficient? By simply playing around with (5.2.1), we may be able to identify a class of Markov chains whose fixed point distributions have asymptotically the same mean and variance, which would be a small step toward proving Poisson-ness but worthwhile heuristic evidence nonetheless.

Question (4)**.** What is the contribution, if any, of Theorem 4.2.3 and Corollary 4.2.4 to related topics in algebraic combinatorics?

In particular, can Theorem 4.2.3 and Corollary 4.2.4 shed any insight on the notoriously difficult-to-compute Kronecker coefficients? Kronecker coefficients are the multiplicities in the tensor product decomposition of two irreps; see [BI08] for a survey of the subject as well as a complexity-theoretic implication. Decompositions of higher tensor powers are related to plethysms of symmetric functions, and plethystic computations have led to remarkable recent advances in the theory of Macdonald polynomials. See Appendix 2 and Exercise 7.74 following the Chapter 7 of [Sta99] for an introduction to plethysms and [LR11] for their connections to the Macdonald polynomials, connections that delve into some of the deepest and most active areas of algebraic combinatorics.

We have now ventured into a field of intricately connected ideas with much potential for further exploration. Any of these topics is sure to lead us down a wondrous rabbit hole. However, to quote Dostoevsky, that might be the subject of a new story, but our present story is ended.

Bibliography

[AS65] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*, Dover, New York, 1965.

[Ald99] D. J. Aldous, *Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists*, Bernoulli **5** (1999), no. 1, 3-48.

[AD86] D. J. Aldous and P. Diaconis, *Shuffling cards and stopping times*, Amer. Math. Monthly **93** (1986), no. 5, 333-348.

[AU08] B. Ali and A. Umar, *Some combinatorial properties of the alternating group*, Southeast Asian Bull. Math. **32** (2008), no. 5, 823-830.

[ADS11] S. Assaf, P. Diaconis, and K. Soundararajan, *A rule of thumb for riffle shuffling*, Ann. Appl. Probab. **21** (2011), no. 3, 843-875.

[BD92] D. Bayer and P. Diaconis, *Trailing the dovetail shuffle to its lair*, Ann. Appl. Probab. **2** (1992), no. 2, 294-313.

[Beh00] E. Behrends, *Introduction to Markov Chains (with Special Emphasis on Rapid Mixing)*, Vieweg Verlag, Braunschweig/Wiesbaden, 2000.

[Bil95] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley, New York, 1995.

[BI08] P. Bürgisser and C. Ikenmeyer, *The complexity of computing Kronecker coefficients*, FPSAC 2008, DMTCS proc. **AJ** (2008), 357-368.

[CST08] T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli, *Harmonic Analysis on Finite Groups: Representation Theory, Gelfand Pairs and Markov Chains*, Cambridge Stud. Adv. Math. 108, Cambridge Univ. Press, Cambridge, 2008.

[CH10] M. A. Conger and J. A. Howald, *A better way to deal the cards*, Amer. Math. Monthly **117** (2010), no. 8, 686-700.

[CV06] M. A. Conger and D. Viswanath, *Riffle shuffles of decks with repeated cards*, Ann. Prob. **34** (2006), no. 2, 804-819.

[Dia88] P. Diaconis, *Group Representations in Probability and Statistics*, IMS Lecture Notes Monogr. Ser. 11, Inst. Math. Statist., Hayward, CA, 1988.

[Dia01] P. Diaconis, *Mathematical developments from the analysis of riffle shuffling*, in *Groups, Combinatorics and Geometry: Durham 2001* (A. A. Ivanov, M. W. Liebeck, and J. Saxl., eds.), 73-97, World Scientific, River Edge, NJ, 2001.

[DF09] P. Diaconis and J. Fulman, *Carries, shuffling, and symmetric functions*, Adv. in Appl. Math. **43** (2009), no.2, 176-196.

[DFH13] P. Diaconis, J. Fulman, and S. P. Holmes, *Analysis of casino shelf shuffling machines*, Ann. Appl. Probab. **23** (2013), no. 4, 1692-1720.

[DPR12] P. Diaconis, C. Y. A. Pang, and A. Ram, *Hopf algebras and Markov chains: two examples and a theory*, preprint, 2012.

[DR12] P. Diaconis and A. Ram, *A probabilistic interpretation of the Macdonald polynomials*, Ann. Prob. **40** (2012), no. 5, 1861-1896.

[DS81] P. Diaconis and M. Shahshahani, *Generating a random permutation with random transpositions*, Z. Wahrsch. verw. Geb. **57** (1981), no. 2, 159-179.

[Dur03] R. Durrett, *Shuffling chromosomes*, J. Theor. Probab. **16** (2003), no. 3, 725-750.

[Fel68] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. I*, 3rd ed., Wiley, New York, 1968.

[Ful00] J. Fulman, *Semisimple orbits of Lie algebras and card shuffling measures on Coxeter groups*, J. Algebra **224** (2000), no. 1, 151-165.

[Ful01] J. Fulman, *Applications of the Brauer complex: card shuffling, permutation statistics, and dynamical systems*, J. Algebra **243** (2001), no. 1, 96-122.

[FH91] W. Fulton and J. Harris, *Representation Theory: A First Course*, GTM 129, Springer-Verlag, New York, 1991.

[Fro00] F. G. Frobenius, *Über die Charaktere der symmetrischen Gruppen*, Sitz. Konig. Preuss. Akad. Wissen. (1900), 516-534.

[FRT54] J. S. Frame, G. de B. Robinson, and R. M. Thrall, *The hook graphs of the symmetric group*, Canad. J. Math. **6** (1954), 316-325.

[GC06] A. Goupil and C. Chauve, *Combinatorial operators for Kronecker powers of representations of S_n*, Sem. Lothar. Combin. **54** (2006), B54j.

[Hil92] M. V. Hildebrand, *Generating random elements in $SL_n(F_q)$ by random transvections*, J. Alg. Comb. **1** (1992), no. 2, 133-150.

[HMR12] V. T. Hoang, B. Morris, and P. Rogaway, *An enciphering scheme based on a card shuffle*, in *Advances in Cryptology – CRYPTO 2012* (R. Safavi-Naini and R. Canetti, eds.), LNCS 7417, 1-13, Springer, Berlin/Heidelberg, 2012.

[HLW06] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. **43** (2006), no. 4, 439-561.

[Jam78] G. D. James, *The Representation Theory of the Symmetric Groups*, LNM 682, Springer-Verlag, Berlin, 1978.

[Jon06] J. Jonasson, *The overhand shuffle mixes in* $\Theta(n^2 \ln n)$ *steps*, Ann. Appl. Probab. **16** (2006), no. 1, 231-243.

[LPW08] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society, Providence, RI, 2008.

[LR11] N. A. Loehr and J. B. Remmel, *A computational and combinatorial exposé of plethystic calculus*, J. Alg. Comb. **33** (2011), no. 2, 163-198.

[Lul96] N. Lulov, *Random Walks on the Symmetric Group Generated by Conjugacy Classes*, Ph.D. Thesis, Harvard University, 1996.

[Mar06] A. A. Markov, *Extension of the law of large numbers to dependent events*, Bull. Soc. Math. Kazan **2** (1906), 135-156.

[Mir02] I. Mironov, *(Not so) random shuffles of RC4*, in *Advances in Cryptology – CRYPTO 2002* (M. Yung, ed.), LNCS 2442, 304-319, Springer, Berlin/Heidelberg, 2002.

[MH09] R. M. Mnatsakanov and A. S. Hakobyan, *Recovery of distributions via moments*, IMS Lecture Notes Monogr. Ser. 57, 252-265, Inst. Math. Statist., Beachwood, OH, 2009.

[Mon08] P. R. de Montmort, *Essay d'Analyse sur les Jeux de Hazard*, Jacques Quillau, Paris, 1708, reprinted by Chelsea, New York, 1980.

[MNP12] B. Morris, W. Ning, and Y. Peres, *Mixing time of the card-cyclic-to-random shuffle*, preprint, 2012.

[MPS04] E. Mossel, Y. Peres, and A. Sinclair, *Shuffling by semi-random transpositions*, Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (2004), 572-581.

[Mur37] F. E. Murnaghan, *The characters of the symmetric group*, Amer. J. Math. **59** (1937), 739-753.

[Nak40] T. Nakayama, *On some modular properties of irreducible representations of a symmetric group, I and II*, Jap. J. Math. **17** (1940), 165-184 and 411-423.

[Pem89] R. Pemantle, *Randomization time for the overhand shuffle*, J. Theor. Probab. **2** (1989), no. 1, 37-49.

[Pem94] R. Pemantle, *A shuffle that mixes sets of any fixed size much faster than it mixes the whole deck*, Rand. Struct. Alg. **5** (1994), no. 5, 609-625.

[Poi12] H. Poincaré, *Calcul des Probabilités*, 2nd ed., Gauthier-Villars, Paris, 1912.

[Rio37] J. Riordan, *Moment recurrence relations for binomial, poisson and hypergeometric frequency distributions*, Ann. Math. Stat. **8** (1937), no. 2, 103-111.

[Sag01] B. E. Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, 2nd ed., GTM 203, Springer-Verlag, New York, 2001.

[S-C01] L. Saloff-Coste, *Probability on groups: random walks and invariant diffusions*, Notices Amer. Math. Soc. **48** (2001), no. 9, 968-977.

[S-C04] L. Saloff-Coste, *Random walks on finite groups*, in *Probability on Discrete Structures* (H. Kesten, ed.), Encyclopaedia Math. Sci. 110, 263-346, Springer-Verlag, Berlin, 2004.

[SG89] D. Sankoff and M. Goldstein, *Probabilistic models of genome shuffling*, Bull. Math. Biol. **51** (1989), no. 1, 117-124.

[Sta97] R. P. Stanley, *Enumerative Combinatorics, Vol. I*, Wadsworth, Monterey CA, 1986, Cambridge Stud. Adv. Math. 49, reprinted by Cambridge Univ. Press, Cambridge, 1997.

[Sta99] R. P. Stanley, *Enumerative Combinatorics, Vol. II*, Cambridge Stud. Adv. Math. 62, Cambridge Univ. Press, Cambridge, 1999.

[Sta01] R. P. Stanley, *Generalized riffle shuffles and quasisymmetric functions*, Ann. Comb. **5** (2001), no. 3-4, 479-491.

[Tao12] T. Tao, *Topics in Random Matrix Theory*, American Mathematical Society, Providence, RI, 2012.

[Tho73] E. O. Thorp, *Nonrandom shuffling with applications to the game of Faro*, J. Amer. Statist. Assoc. **68** (1973), no. 344, 842-847.

[Wil04] D. B. Wilson, *Mixing times of Lozenge tiling and card shuffling Markov chains*, Ann. Appl. Probab. **14** (2004), no. 1, 274-325.