The Ramanujan Journal
https://doi.org/10.1007/s11139-021-00482-z

Hahn polynomials and the Burnside process

Persi Diaconis[1] · **Chenyang Zhong**[2]

Dedicated to the memory of Richard Askey.

Received: 26 December 2020 / Accepted: 26 June 2021

Abstract
We study a natural Markov chain on $\{0, 1, \ldots, n\}$ with eigenvectors the Hahn polynomials. This explicit diagonalization makes it possible to get sharp rates of convergence to stationarity. The process, the Burnside process, is a special case of the celebrated 'Swendsen–Wang' or 'data augmentation' algorithm. The description involves the beta-binomial distribution and Mallows model on permutations. It introduces a useful generalization of the Burnside process.

Keywords Hahn polynomials · Burnside process · Beta-binomial distribution · Mallows model on permutations · Orthogonal polynomial eigenfunctions · Auxiliary variables

Mathematics Subject Classification 33C45 · 33D45 · 60J10

1 Introduction

Over the past 100 years, physicists, geneticists, statisticians, and probabilists have found positive symmetric operators (Markov chains) with orthogonal polynomial eigenfunctions. This allows precise asymptotic analysis of high powers of the operator. Orthogonal polynomials come in 'families' and an explicit diagonalization suggests

Persi Diaconis's and Chenyang Zhong's research is supported by NSF grant DMS 1954042.
Data availability statement: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

✉ Chenyang Zhong
cyzhong@stanford.edu

Persi Diaconis
diaconis@math.stanford.edu

[1] Department of Statistics and Mathematics, Stanford University, Stanford, CA, USA

[2] Department of Statistics, Stanford University, Stanford, CA, USA

Published online: 17 September 2021

'deforming' the operator so that the full family of eigenfunctions appears. These deformations can turn out to be useful and natural. This paper gives an example with these characteristics.

We start off with a high-level description of our main results. This uses the language of Markov chains and the Burnside process. A tutorial containing background on these topics is in Sects. 2.1 and 2.2. See also Sect. 4 for more details.

1.1 The Burnside process

Let $\mathcal{X}$ be a finite set. Let G be a finite group acting on $\mathcal{X}$. This splits $\mathcal{X}$ into disjoint orbits. Let $O_x = \{x^g : g \in G\}$ be the orbit containing x. The Burnside process gives a practical way to choose an orbit uniformly at random. This captures a familiar problem: enumerating *unlabeled* objects. For example, there are n^{n-2} labeled trees on n vertices. There is no formula for the number of unlabeled trees. Indeed, there is an emerging literature for choosing a random spanning tree of a graph. As a second example, let $\mathcal{X} = G$ with G acting on itself by conjugation. Now the orbits are conjugacy classes of G. For instance, if $G = S_n$, the symmetric group, the conjugacy classes are indexed by partitions of n and the Burnside algorithm gives a novel way to choose a random partition.

Let

$$\mathcal{X}_g = \{y \in \mathcal{X} : y^g = y\}, \quad G_x = \{g \in G : x^g = x\}.$$

The Burnside process is a Markov chain on $\mathcal{X}$:

- From x, choose $g \in G_x$ uniformly at random
- From g, choose $y \in \mathcal{X}_g$ uniformly at random

One step of the chain goes from x to y. The transition matrix of this chain is

$$K(x, y) = \sum_{g \in G_x \cap G_y} \frac{1}{|G_x|} \frac{1}{|\mathcal{X}_g|}, \quad \text{an } |\mathcal{X}| \times |\mathcal{X}| \text{ matrix.}$$

It is easy to see that K has stationary distribution

$$\pi(x) = \frac{z^{-1}}{|O_x|} \quad (z = \text{number of orbits}).$$

This means, if the chain is run from any starting state x_0, after 'a long time' the orbit containing the current state is close to uniformly distributed.

In this generality, useful quantitative analysis of the Burnside process is an open problem, even for the symmetric group.

We study a special case where analysis can be pushed through.

Let

- $\mathcal{X} = C_2^n$—the binary n-tuples
- $G = S_n$ acting by permuting coordinates

If $x \in \mathcal{X}$ has j ones (denoted $|x| = j$), the orbit

$$O_x = \{y \in \mathcal{X} : |y| = j\}. \tag{1.1}$$

Here $0 \le |x| \le n$ and the Burnside process gives a complicated way of choosing an element uniformly from $\{0, 1, \ldots, n\}$. It is instructive to see how the general algorithm specializes:

- For $x \in \mathcal{X}$, the permutations fixing x form the subgroup $S_j \times S_{n-j}$ (for $j = |x|$) permuting the ones among themselves and the zeros among themselves. It is easy to choose $\sigma \in G_x$ at random.
- For $\sigma \in S_n$, write σ as a product of cycles. Label the entries of each cycle, randomly, by a zero or a one ($2^{c(\sigma)}$ choices if σ has $c(\sigma)$ cycles). Installing this zero/one pattern gives y. It is easy to choose y given σ.

A first result of this paper shows that, for this example, a finite number of steps are necessary and sufficient for convergence no matter how large n is. To say this carefully, define

$$\|K_x^l - \pi\| = \frac{1}{2} \sum_{y \in \mathcal{X}} |K^l(x, y) - \pi(y)|.$$

Theorem 1 *For $\mathcal{X} = C_2^n$, $G = S_n$, the Burnside process defined above satisfies, for all $n \ge 2$, starting state $x_0 = (1, 1, \ldots, 1)$ and all $l \in \{1, 2, 3, \ldots\}$,*

$$\frac{1}{4}\left(\frac{1}{4}\right)^l \le \|K_{x_0}^l - \pi\| \le 4\left(\frac{1}{4}\right)^l. \tag{1.2}$$

The development thus far does not seem to have much to do with orthogonal polynomials. Indeed, we do not really understand where they come from or just when they will appear in other versions of the Burnside process. Let us explain their appearance here.

Let $X_0 = x, X_1, X_2, \ldots$ be the realization of the Burnside chain of Theorem 1. There is enough symmetry around that $|X_0| = |x|, |X_1|, |X_2|, \ldots$ forms a Markov chain on $\{0, 1, 2, \ldots, n\}$. Let $Q(i, j)$ be the transition matrix for this chain. The explicit form of Q is derived in Proposition 4 below (originally it was written out in [15, (3.1)–(3.3)]). The Q chain turns out to be symmetric with a uniform stationary distribution $u(j) = \frac{1}{n+1}$, $0 \le j \le n$.

Let $T_j^n(x)$ be the discrete Chebyshev polynomials on $\{0, \ldots, n\}$—the orthogonal polynomials for the uniform distribution ([10]).

Theorem 2 *For the Markov chain $Q(i, j)$ on $\{0, \ldots, n\}$, the non-zero eigenvalues are given by 1 and $\{\frac{\binom{2k}{k}^2}{2^{4k}}\}_{k=1}^{\lfloor \frac{n}{2} \rfloor}$. Moreover, the eigenfunctions corresponding to the zero eigenvalues are the discrete Chebyshev polynomials on $\{0, 1, \ldots, n\}$ with odd degree. The eigenfunction corresponding to the eigenvalue $\frac{\binom{2k}{k}^2}{2^{4k}}$ is the discrete Chebyshev polynomial on $\{0, 1, \ldots, n\}$ with degree $2k$, $1 \le k \le \frac{n}{2}$.*

Theorem 2 and classical analysis yield Theorem 1. The proof appears in Sect. 3.

1.2 Hahn polynomials

The discrete Chebyshev polynomials sit naturally in the family of Hahn polynomials: for $\alpha, \beta > 0$,

$$Q_j(x) = {}_3F_2(-j, j+\alpha+\beta-1, -x; \alpha, -n|1) \tag{1.3}$$

(Set $\alpha = \beta = 1$ for discrete Chebyshev). Q_j is a polynomial of degree j for $0 \le j \le n$. The usual notation is indicated with α, β, and n sub and superscripts suppressed here. These polynomials are orthogonal with respect to the 'beta-binomial distribution' defined and discussed in Sect. 2.3 below.

It is natural to ask, is there some variant of the Burnside process having Q_j as eigenfunctions? This is a mathematicians question, as per the first paragraph of this introduction. To answer it, we found a useful *abstract* generalization of the general Burnside process which we hope has a life of its own. Specialized to C_2^n and S_n it gives a Markov chain on $\{0, 1, \ldots, n\}$ with a 'beta-binomial distribution' having Hahn polynomial eigenfunctions and allowing a parallel to Theorem 1.

Section 2 of this paper sets out the needed background: on Markov chains, the Burnside process, and orthogonal polynomials (it also gives a brief overview of the many other Markov chains with polynomial eigenfunctions—from Askey–Wilson to Macdonald!). Theorems 1 and 2 are proved in Sect. 3. Section 4 explains our 'twisted Burnside process' from first principles. The analogs of Theorems 1 and 2 are stated in Sect. 5. In Sect. 6, following a suggestion of a referee, we pass to the limit as n tends to infinity to find a continuous analog of the Burnside process with Jacobi polynomial eigenfunctions. The limit is usefully related to the finite case.

2 Background

This section contains background material on the Burnside process and related 'auxiliary variables' algorithms (Sect. 2.1). The needed theory from the analytic-geometric theory of Markov chains is in Sect. 2.2. Background on orthogonal polynomials and connections to Markov chains (Cannings method) are in Sect. 2.3. Since our readership has diverse backgrounds we attempt a friendly tutorial.

2.1 The Burnside process and auxiliary variables

The Burnside process was introduced by Mark Jerrum [27] as a contribution to 'computational Pólya theory.' Pólya theory is about enumeration under symmetry: "how many different ways can ten dice be colored with red, white, blue where the order of the dice and the symmetry of a die are neglected?" This is a standard part of combinatorics. The book by Pólya and Read [37] gives a classical account. Jerrum, working with Goldberg [22,23,28], managed to show that for carefully selected problems, these

computations are #P-complete (this means being polynomially equivalent to dozens of other problems that are believed to require exponential time). Our work shows that for many natural problems, the Burnside process is highly efficient. This opens up a research area—when is it good? Applications of the Burnside process to practical problems in computer science [25] suggest it is worthy of study.

A different motivation comes because the Burnside process is a special case of a wonderful 'unifying algorithm' called variously 'auxiliary variables,' 'data augmentation,' or 'hit and run.' *Briefly* let $\mathcal{X}$ be a finite or countable set, $\pi(x) > 0$ a probability on $\mathcal{X}$. The job is to invent a Markov chain to sample from $\pi(x)$. Auxiliary variables create 'non-local' Markov chains, able to move far away in a single step. They seem to mix much more rapidly than standard local algorithms. Introduce a set I of auxiliary variables. For each $i \in I$ and $x \in \mathcal{X}$, choose a proposal kernel $w_x(i) \geq 0$ (a way of moving from x to i) such that $\sum_i w_x(i) = 1$ and for each i, there is at least one x with $w_x(i) > 0$. These ingredients define a joint probability $f(x, i) = \pi(x)w_x(i)$ on $\mathcal{X} \times I$. To proceed, for each i, a Markov chain $K_i(x, y)$ with stationary distribution $f(x|i)$ must be specified. The auxiliary variables algorithm is

- From x, choose i with probability $w_x(i)$
- From i, choose y with probability $K_i(x, y)$

The chain moves from x to y in one step. The transition matrix is

$$K(x, y) = \sum_i w_x(i) K_i(x, y).$$

A direct calculation shows that $\pi(x)$ is a stationary distribution:

$$\begin{aligned}\sum_x \pi(x) K(x, y) &= \sum_x \pi(x) \sum_i w_x(i) K_i(x, y) = \sum_i \sum_x w_x(i)\pi(x) K_i(x, y) \\ &= \sum_i m(i) \sum_x f(x|i) K_i(x, y) = \sum_i m(i) f(y|i) = \pi(y)\end{aligned}$$

($m(i) = \sum_x f(x, i)$ is the marginal distribution). Of course connectedness and aperiodicity of K must be checked so that the Perron–Frobenius theorem is in force.

A comprehensive review of this class of algorithms is in [2] which gives dozens of examples and special cases including the celebrated Swendsen–Wang algorithm for sampling from the Ising model.

The point for now is that essentially *none* of the algorithms have sharp running time analysis. The Burnside process falls into this class: take $I = G$, $w_x(i)$ the uniform distribution on G_x and $K_g(x, y)$ the uniform distribution on $\mathcal{X}_g$. The hope is, because of the extra group structure available, analysis will be easier for this special case. That was our initial motivation.

There has been some previous effort to study the Burnside process on $\mathcal{X} = [k]^n$, with $G = S_n$—here $[k] = \{1, 2, \ldots, k\}$. See [15] and Example 1 in Sect. 4 for the description of this process.

Theorems 1 and 2 study $k = 2$. Jerrum [27] showed that for $k = 2$ the total variation mixing time is at most of order $\sqrt{n}$; sharpening this, Diaconis [15] proved that, for

$k = 2$ and $x_0 = (1, \ldots, 1)$, $\|K^l_{x_0} - \pi\| \leq (1-c)^l$ with $c = \frac{1}{\pi}$ (with $\pi = 3.14159\ldots$). Theorem 1 sharpens this to get the exact rate and makes the connection to orthogonal polynomials (versus the complex, difficult argument of [15]). Aldous [1] used a clever coupling argument to show that, for general k, $\|K^l_x - \pi\| \leq n(1 - \frac{1}{k})^l$. This implies l of order $\log n$ steps suffice for $k = 2$. Further results are in [9].

2.2 Analysis of Markov chain convergence

For readers unfamiliar with Markov chains, the best introduction is the book by Levin and Peres [35]. This contains all the bounds below and much more. The *analytic-geometric* theory of Markov chains is well developed in Saloff-Coste's [38].

Let $\mathcal{X}$ be a finite set, $\pi(x) > 0$, $\sum_x \pi(x) = 1$ a probability distribution on $\mathcal{X}$. A Markov chain on $\mathcal{X}$ is specified by a matrix $K(x, y) \geq 0$ with $\sum_y K(x, y) = 1$. Suppose that π is a stationary distribution for K: $\sum_x \pi(x)K(x, y) = \pi(y)$ (so π is a left eigenvector for K with eigenvalue 1). Call π, K *reversible* if the detailed balance condition $\pi(x)K(x, y) = \pi(y)K(y, x)$ is satisfied for all $x, y \in \mathcal{X}$. All the Markov chains below satisfy detailed balance. Let $L^2(\pi) = \{f : \mathcal{X} \to \mathbb{R} : \sum f(x)^2\pi(x) < \infty\}$. K acts on $L^2(\pi)$ by $Kf(x) = \sum_y K(x, y)f(y)$. Detailed balance is equivalent to saying that $\langle Kf|g\rangle = \langle f|Kg\rangle$ so K is a bounded, self-adjoint operator on $L^2(\pi)$. The spectral theorem is in force: there exist *eigenvalues* β_i ($1 = \beta_0 \geq \beta_1 \geq \cdots \geq \beta_{|\mathcal{X}|-1} \geq -1$) and *eigenfunctions* $\psi_i(x)$ (so $K\psi_i(x) = \beta_i\psi_i(x)$).

Throughout this section assume K is connected (for every $x, y \in \mathcal{X}$ there is $l \geq 1$ with $K^l(x, y) > 0$) and aperiodic ($\beta_{|\mathcal{X}|-1} > -1$). It is useful to introduce two distances

- Total variation distance—$\|K^l_x - \pi\| = \frac{1}{2}\sum_y |K^l(x, y) - \pi(y)| = \max_{\|f\|_\infty \leq 1} K^l(f) - \pi(f)$. Here $\|f\|_\infty = \max_x |f(x)|$, $\pi(f) = \sum \pi(x)f(x)$. This equality is easy to prove (the maximizing f is 1 at y if $K^l(x, y) \geq \pi(y)$ and -1 otherwise).
- Chi-square distance—$\chi^2_x(l) = \sum_y \frac{(K^l(x,y)-\pi(y))^2}{\pi(y)}$.

The Cauchy–Schwarz inequality implies

$$4\|K^l_x - \pi\|^2 \leq \chi^2_x(l). \tag{2.1}$$

This is a useful route to getting bounds on convergence: bound L^1 by L^2 and use eigenvalues to bound L^2. Eigenvalues and eigenfunctions come in by

$$\chi^2_x(l) = \sum_{i=1}^{|\mathcal{X}|-1} \beta_i^{2l}\psi_i^2(x). \tag{2.2}$$

The following lower bound on total variation and chi-square distance is proved in [19, Lemma 2.1]. It is applied in Sect. 3.

Proposition 1 *Suppose that ψ is an eigenfunction of K with eigenvalue β such that $\pi(\psi) = 0$. Assume $\|\psi\|^2 = 1$. Then*

$$\chi^2_x(l) \geq |\psi(x)|^2|\beta|^{2l}.$$

For ψ with $\|\psi\|_\infty < \infty$,

$$\|K_x^l - \pi\| \geq \frac{|\psi(x)||\beta|^l}{2\|\psi\|_\infty}.$$

2.3 Orthogonal polynomials and Markov chains

Background on orthogonal polynomials is wonderfully presented in the introductory text by Chihara [10]. The recent encyclopedia by Ismail [26] brings in much further material.

These are the orthogonal polynomials on $\{0, 1, \ldots, n\}$ with respect to the beta-binomial distribution. For references and variations of the beta-binomial distribution see [41]. Fix $\alpha, \beta > 0$. Define

$$m(j) = \binom{n}{j} \frac{(\alpha)_j (\beta)_{n-j}}{(\alpha+\beta)_n}, (\alpha)_x = \alpha(\alpha+1)\cdots(\alpha+x-1), \quad (\alpha)_0 = 1. \quad (2.3)$$

This is a probability distribution on $\{0, 1, \ldots, n\}$ generated as a beta mixture of binomial distributions

$$m(j) = \int_0^1 \binom{n}{j} x^j (1-x)^{n-j} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx.$$

When $\alpha = \beta = 1$, $m(j) = \frac{1}{n+1}$—the uniform distribution.

The orthogonal polynomials for $m(j)$ are called *Hahn polynomials*. Detailed development is in [29]. They have the explicit form

$$Q_j(x) = {}_3F_2(-j, j+\alpha+\beta-1, -x; \alpha, -n|1)$$

normalized so $Q_j(0) = 1$, $Q_j(n) = \frac{(-\beta-j)_j}{(\alpha+1)_j}$. The standard form

$${}_rF_s(a_1, \ldots, a_r; b_1, \ldots, b_s|x) = \sum_{l=0}^{\infty} \frac{(a_1 \ldots a_r)_l x^l}{(b_1 \ldots b_s)_l l!}$$

with $(a_1 \ldots a_r)_l = \prod_{i=1}^r (a_i)_l$, is used. When $\alpha = \beta = 1$ these are the discrete Chebyshev polynomials (which we denote by $T_j(x)$)

$$T_0(x) = 1, \quad T_1(x) = \frac{n-2x}{n}, \quad T_2(x) = \frac{6x^2 - 6nx + n(n-1)}{n(n-1)}.$$

The discrete Chebyshev polynomials satisfy the following recurrence relation ([26, (6.2.8)])

$$(j+1)(n-j)T_{j+1}(x) = (2j+1)(n-2x)T_j(x) - j(j+n+1)T_{j-1}(x), \quad (2.4)$$

where $j = 0, 1, \ldots, n-1$ and $T_{-1}(x) = 0$.

In Sect. 3 the proof of Theorem 2 uses 'Cannings argument' ([8]), a frequently used tool for proving that a Markov chain has orthogonal polynomial eigenfunctions:

'If the operator K sends polynomials of degree j to polynomials of degree j for all $0 \le j \le n$, then the operator has the orthogonal polynomials for the stationary distribution as its eigenfunctions.'

The following lemma is a formal statement and proof of Cannings lemma. A very similar argument works for multivariate polynomials.

Lemma 1 *Consider a sequence of reversible Markov chains with state space $\mathcal{X}_n = \{0, 1, \ldots, n\}$, transition matrix K_n, and stationary distribution π_n (with $\pi_n(x) > 0$ for any $x \in \mathcal{X}_n$) for $n = 1, 2, \ldots$. For any $n \in \{1, 2, \ldots\}$, suppose that for any polynomial f on $\mathcal{X}_n$ of degree $l \le n$, $K_n f$ is a polynomial on $\mathcal{X}_n$ of degree less than or equal to l. Then we conclude that the eigenfunctions of K_n are given by orthogonal polynomials for π_n of degree less than or equal to n.*

Moreover, for any $l \in \mathbb{N}$, if we further assume that the degree l monomial in the expansion of $K_n[x^l]$ in terms of $\{1, x, \ldots, x^n\}$ (note that this is well-defined for $n \ge l$, as the monomials $\{1, x, \ldots, x^n\}$ are linearly independent on $\mathcal{X}_n$) does not depend on n for $n \ge l$, then the eigenvalue of K_n that corresponds to the eigenfunction of degree l does not depend on n for $n \ge l$.

Proof We prove by induction that for any $0 \le l \le n$, the orthogonal polynomial for π_n of degree l is an eigenfunction of K_n. We denote the lth orthogonal polynomial for π_n by $\phi_{n,l}$. For $l = 0$, we have $K_n \mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the constant function taking value 1. Now for l such that $1 \le l \le n$, by the induction hypothesis we suppose that for any $0 \le i \le l-1$, $K_n \phi_{n,i} = \lambda_{n,i} \phi_{n,i}$. Now $\phi_{n,l}$ is a polynomial of degree l, hence $K_n \phi_{n,l}$ is a polynomial of degree $\le l$. Expanding $K_n \phi_{n,l}$ in terms of $\{\phi_{n,i}\}_{i=0}^{l}$, we get

$$K_n \phi_{n,l} = \lambda_{n,l} \phi_{n,l} + \sum_{i=0}^{l-1} a_{n,l,i} \phi_{n,i} \tag{2.5}$$

for some $\lambda_{n,l}$, $\{a_{n,l,i}\}_{i=0}^{l-1}$. Now consider the inner product for π_n defined as

$$\langle f, g \rangle_{\pi_n} := \sum_{k=0}^{n} f(k) g(k) \pi_n(k).$$

K_n is self-adjoint with respect to this inner product, hence we have for any $0 \le i \le l-1$,

$$\langle K_n \phi_{n,l}, \phi_{n,i} \rangle_{\pi_n} = \langle \phi_{n,l}, K_n \phi_{n,i} \rangle_{\pi_n},$$

which leads to

$$a_{n,l,i} = \frac{\lambda_{n,i} \langle \phi_{n,l}, \phi_{n,i} \rangle_{\pi_n}}{\langle \phi_{n,i}, \phi_{n,i} \rangle_{\pi_n}} = 0.$$

Therefore, we conclude that $K_n\phi_{n,l} = \lambda_{n,l}\phi_{n,l}$.

For the second part, it suffices to consider the case when $l \geq 1$. We assume that the degree l monomial in the expansion of $K_n[x^l]$ in terms of $\{1, x, \ldots, x^n\}$ does not depend on n for $n \geq l$. We also denote by $\lambda_{n,i}$ the eigenvalue of K_n that corresponds to the eigenfunction $\phi_{n,i}$ for any n, i with $n \geq i$. For any $n \geq l$, we expand x^l in terms of $\{\phi_{n,i}\}_{i=0}^{l}$, and get

$$x^l = \sum_{i=0}^{l} \beta_{n,l,i}\phi_{n,i} \tag{2.6}$$

for some $\{\beta_{n,l,i}\}_{i=0}^{l}$. Then we have

$$\begin{aligned} K_n[x^l] &= \sum_{i=0}^{l} \beta_{n,l,i} K_n[\phi_{n,i}] = \sum_{i=0}^{l} \beta_{n,l,i}\lambda_{n,i}\phi_{n,i} \\ &= \lambda_{n,l}x^l + \sum_{i=0}^{l-1} \beta_{n,l,i}(\lambda_{n,i} - \lambda_{n,l})\phi_{n,i}. \end{aligned} \tag{2.7}$$

As the term $\sum_{i=0}^{l-1} \beta_{n,l,i}(\lambda_{n,i} - \lambda_{n,l})\phi_{n,i}$ is of degree $\leq l-1$, we conclude that the degree l monomial in the expansion of $K_n[x^l]$ in terms of $\{1, x, \ldots, x^n\}$ is given by $\lambda_{n,l}x^l$. Hence by our assumption, $\lambda_{n,l}$ does not depend on n for $n \geq l$. Note that the eigenvalue of K_n that corresponds to the eigenfunction $\phi_{n,l}$ is $\lambda_{n,l}$, which does not depend on n for $n \geq l$. □

The connections between orthogonal polynomials and Markov chains are long-standing. Indeed, perhaps the first Markov chain—the Bernoulli–Laplace urn—is diagonalized by Hahn polynomials. To briefly recall, Bernoulli and Laplace considered two urns, the left containing n red balls, the right containing n black balls. Each time, a ball is picked uniformly at random and the two balls are switched. The number of red balls in the left urn evolves as a Markov chain on $\{0, 1, \ldots, n\}$. In [17] this chain is diagonalized and shown to have Hahn polynomial eigenfunctions. That same paper treats the Ehrenfest's urn (with Krawtchouk polynomial eigenfunctions) and several q-deformations of these two examples. These urn models, and the birth and death chains discussed below, are examples of 'local Markov chains.' Most appearances of orthogonal polynomials have occurred for such local chains (analogs of the Laplacian). The present examples are much more vigorous.

Jacobi polynomials arise as eigenfunctions of a Markov chain constructed from the 'Gibbs sampler' [19]. Extensions of this construction need the full power of the Askey–Wilson polynomials [7].

An extensive connection between orthogonal polynomials and birth–death chains follows from the Karlin–McGregor theory. A textbook account with full details is in [3]. Multivariate orthogonal polynomials as in [21] arise in natural genetics problems. See [16,31,42] for multivariate Hahn and Krawtchouk polynomials and [20] for Macdonald polynomials. This is just a small sample, drawn from our work with colleagues. The list goes on and on.

3 Proof of Theorems 1 and 2

The proof consists of three main parts. First we show that the operator Q of Sect. 1.1 sends polynomials to polynomials. Thus Cannings lemma (Lemma 1) shows that the discrete Chebyshev polynomials are the eigenfunctions of the Burnside process (lumped to orbits). Next, the eigenvalues are computed, proving Theorem 2. Finally, the eigen structure and analytic tools of Sect. 2.2 are used to prove Theorem 1.

Throughout this section $\alpha = \beta = 1$ and the stationary distribution $u(j) = \frac{1}{n+1}$ for $0 \le j \le n$.

3.1 Proof of Theorem 2

In the following, we denote by $T_j^n(x)$ the discrete Chebyshev polynomial on $\{0, 1, \ldots, n\}$ with degree j (to emphasize the dependence on n). We also assume that n is even below (the proof for odd n is similar).

Before the proof of Theorem 2, we present some preparatory lemmas.

The following lemma follows from Pólya's cycle index theorem ([40], [18, Sect. 5]). We recall the notation. For σ in the symmetric group S_n, write $a_i(\sigma)$ for the number of i-cycles when σ is written in cycle notation. So $a_1(\sigma)$ is the number of fixed points, $a_2(\sigma)$ is the number of transpositions,$\cdots$ Thus $0 \le a_i(\sigma) \le n$ and $\sum_{i=1}^n i a_i(\sigma) = n$. Write the cycle index of S_n as

$$C_n(x_1, \ldots, x_n) = \frac{1}{n!} \sum_{\sigma \in S_n} \prod_{i=1}^n x_i^{a_i(\sigma)}. \tag{3.1}$$

The generating function of these C_n is $C(t) = \sum_{n=0}^\infty C_n t^n$. Pólya showed

$$C(t) = e^{\sum_{i=1}^\infty x_i \frac{t^i}{i}}. \tag{3.2}$$

Repeatedly differentiating in the x_i, setting $x_i = 1$ and comparing coefficient gives

Lemma 2 *For $n = 1, 2, \ldots$, any $1 \le k_1 < \cdots < k_r \le n$ and $l_1, \ldots, l_r \ge 1$, if $\sum_{i=1}^r k_i l_i \le n$, then for a random permutation σ uniformly chosen in S_n,*

$$\left(\prod_{i=1}^r (l_i)! \mathbb{E} \left[\prod_{i=1}^r \binom{a_{k_i}(\sigma)}{l_i} \right] = \prod_{i=1}^r \frac{1}{k_i^{l_i}} \right); \tag{3.3}$$

while if $\sum_{i=1}^r k_i l_i > n$, then

$$\mathbb{E} \left[\prod_{i=1}^r \binom{a_{k_i}(\sigma)}{l_i} \right] = 0. \tag{3.4}$$

Below, by "even-order terms" we mean monomials of the form $\alpha X_{i_1}^{l_1} \ldots X_{i_k}^{l_k}$ with $i_1 < \cdots < i_k$, $\alpha \in \mathbb{R}$ and $l_1, \ldots, l_k$ even.

Lemma 3 *Suppose that $N, a \geq 1$. The "even-order terms" in the expansion of $(\sum_{i=1}^{N} X_i)^{2a}$ can be expressed as a linear combination of the following form:*

$$\prod_{q=1}^{k} \left(\sum_{i=1}^{N} X_i^{2l_q} \right) \tag{3.5}$$

for some $k \geq 1$, $l_1, \ldots, l_k \geq 1$ and $\sum_{q=1}^{k} l_q = a$. Moreover, the coefficients in the linear combination only depend on a.

Proof By adding finitely many indeterminates and taking them to be 0 in the end, we can assume without loss of generality that $N \geq 2a$. We note that the "even-order terms" of $(\sum_{i=1}^{N} X_i)^{2a}$ is a symmetric polynomial in the indeterminates $X_1^2, \ldots, X_N^2$. By the proof of [6, Proposition 2.9], it can be expressed as linear combination of the form (3.5), and the coefficients only depend on a. □

Proof of Theorem 2, eigenfunction part Note that $T_0^n(x) = 1$ and $T_1^n(x) = \frac{n-2x}{n}$. From this and the recurrence relation (2.4) it can be derived that for $0 \leq k \leq \frac{n}{2}$, T_{2k}^n only has even-order terms in $(x - \frac{n}{2})$, and T_{2k-1}^n (for $k \geq 1$) only has odd-order terms in $(x - \frac{n}{2})$.

Now as $Q(i, j) = Q(i, n - j)$, we conclude that $Q[(x - \frac{n}{2})^{2k-1}] = 0$ for any $1 \leq k \leq \frac{n}{2}$. By the previous conclusion, $QT_{2k-1}^n = 0$. Hence $\{T_{2k-1}^n\}_{k=1}^{\frac{n}{2}}$ are eigenfunctions of K corresponding to the zero eigenvalues.

In view of Lemma 1, in order to show that the rest of the eigenfunctions are also given by the discrete Chebyshev polynomials, it suffices to show that for any a such that $1 \leq a \leq \frac{n}{2}$, $Q[(x - \frac{n}{2})^{2a}](j)$ is a polynomial in j of degree $\leq 2a$ (where $j \in \{0, 1, \ldots, n\}$). Below we verify this fact.

Suppose that we start from the orbit j. For the first step in the Burnside process, $\sigma_1 \in S_j$ and $\sigma_2 \in S_{n-j}$ are drawn uniformly. Let $\{a_i\}_{i=1}^n$, $\{b_i\}_{i=1}^n$ denote the number of cycles of length i in σ_1, σ_2, respectively. For the second step, we label the entries of each cycle by 0/1 independently. Let $Z_i \sim Binomial(a_i + b_i, \frac{1}{2})$. The outcome of one iteration of the Burnside process (number of coordinates labeled by 1) can be represented by $X = \sum_{i=1}^{n} i Z_i$. For $1 \leq a \leq \frac{n}{2}$, we denote by

$$W_{2a} := \mathbb{E}\left[\left(X - \frac{1}{2}n \right)^{2a} \right] = \mathbb{E}\left[\left(X - \sum_{i=1}^{n} \frac{1}{2} i (a_i + b_i) \right)^{2a} \right]. \tag{3.6}$$

We note that

$$W_{2a} = \mathbb{E}\left[\mathbb{E}\left[\left(X - \sum_{i=1}^{n} \frac{1}{2} i (a_i + b_i) \right)^{2a} | \sigma_1, \sigma_2 \right] \right]. \tag{3.7}$$

Below we denote by $\tilde{W}_{2a}(\sigma_1, \sigma_2)$ the conditional expectation inside the above expression.

By the multinomial theorem and vanishing of odd moments, we obtain that

$$\begin{aligned}\tilde{W}_{2a}(\sigma_1,\sigma_2) &= \mathbb{E}\left[\left(\sum_{i=1}^{n} i\left(Z_i - \frac{1}{2}(a_i+b_i)\right)\right)^{2a} |\sigma_1,\sigma_2\right] \\ &= \sum_{\substack{j_1+\cdots+j_n=a,\\ j_1,\ldots,j_n\geq 0}} \binom{2a}{2j_1,\ldots,2j_n} \\ &\quad \times \prod_{i=1}^{n} i^{2j_i}\mathbb{E}\left[\left(Z_i - \frac{1}{2}(a_i+b_i)\right)^{2j_i} |\sigma_1,\sigma_2\right]. \end{aligned} \tag{3.8}$$

By splitting Z_i into Bernoulli random variables and using the multinomial theorem again we obtain

$$\mathbb{E}\left[\left(Z_i - \frac{1}{2}(a_i+b_i)\right)^{2j_i} |\sigma_1,\sigma_2\right] = \frac{1}{2^{2j_i}} \sum_{\substack{p_{i,1}+\cdots+p_{i,a_i+b_i}=j_i\\ p_{i,1},\ldots,p_{i,a_i+b_i}\geq 0}} \binom{2j_i}{2p_{i,1},\ldots,2p_{i,a_i+b_i}}. \tag{3.9}$$

Plugging the above expression into (3.8), we get

$$\begin{aligned}\tilde{W}_{2a}(\sigma_1,\sigma_2) &= \frac{1}{2^{2a}} \sum_{\substack{\sum_{i=1}^{n}\sum_{q=1}^{a_i+b_i} p_{i,q}=a\\ p_{i,q}\geq 0}} \prod_{i=1}^{n} i^{2(p_{i,1}+\cdots+p_{i,a_i+b_i})} \\ &\quad \times \binom{2a}{2p_{1,1},\ldots,2p_{1,a_1+b_1},\ldots,2p_{n,1},\ldots,2p_{n,a_n+b_n}}. \end{aligned} \tag{3.10}$$

We note that by (3.10), $2^{2a}\tilde{W}_{2a}$ can be obtained from "even-order terms" in

$$\left(\sum_{i=1}^{n}\sum_{q=1}^{a_i+b_i} X_{i,q}\right)^{2a} \tag{3.11}$$

by substituting $X_{i,q} = i$ for $1 \leq q \leq a_i + b_i$. Now by lemma 3, the "even-order terms" of $(\sum_{i=1}^{n}\sum_{q=1}^{a_i+b_i} X_{i,q})^{2a}$ can be expressed in terms of the following forms:

$$\prod_{r=1}^{k}\left(\sum_{i=1}^{n}\sum_{q=1}^{a_i+b_i} X_{i,q}^{2l_r}\right), \tag{3.12}$$

with $k \geq 1, l_1,\ldots,l_k \geq 1$ and $\sum_{r=1}^{k} l_r = a$. Moreover, the coefficients in the linear combination only depend on a. By substituting $X_{i,q} = i$ and taking expectation in

(3.12), we get

$$\mathbb{E}\left[\prod_{r=1}^{k}\left(\sum_{i=1}^{n} i^{2l_r}(a_i+b_i)\right)\right]. \tag{3.13}$$

In order to make use of Lemma 2, we introduce the following subspaces. We take the base field to be $\mathbb{R}$. For every monomial $\alpha X_{i_1}^{d_1}\cdots X_{i_k}^{d_k}$ ($i_1<\cdots<i_k, \alpha\in\mathbb{R}$), we define

$$B(\alpha X_{i_1}^{d_1}\cdots X_{i_k}^{d_k}) := \alpha(d_1)!\cdots(d_k)!\binom{X_{i_1}}{d_1}\cdots\binom{X_{i_k}}{d_k}, \tag{3.14}$$

and extend linearly. We also denote by

$$X(l_1,\ldots,l_k)=\prod_{r=1}^{k}\left(\sum_{i=1}^{n} i^{2l_r}X_i\right), \tag{3.15}$$

and $B(l_1,\ldots,l_r) := B(X(l_1,\ldots,l_r))$. For any $d\geq 1$, we define $\mathcal{X}(d,a)$ to be the set of $X(l_1,\ldots,l_k)$ with $1\leq k\leq d$ and $l_1+\cdots+l_k=a$, and $\mathcal{B}(d,a)$ to be the set of $B(l_1,\ldots,l_k)$ with $1\leq k\leq d$ and $l_1+\cdots+l_k=a$. We also define $\tilde{\mathcal{X}}(d,a)$ to be the linear span of $\mathcal{X}(d,a)$ and $\tilde{\mathcal{B}}(d,a)$ to be the linear span of $\mathcal{B}(d,a)$.

Now we prove that for $l_1+\cdots+l_k=a$, $\tilde{\mathcal{X}}(k,a)=\tilde{\mathcal{B}}(k,a)$, and that it is possible to express $X(l_1,\ldots,l_k)$ in terms of elements of $\mathcal{B}(k,a)$ so that the coefficients do not depend on n. The method is to prove the following stronger claim by induction on k: for any $l_1+\cdots+l_k=a$, $\tilde{\mathcal{X}}(k,a)=\tilde{\mathcal{B}}(k,a)$, $X(l_1,\ldots,l_k)-B(l_1,\ldots,l_k)\in\tilde{\mathcal{X}}(k-1,a)$, and it is possible to express $X(l_1,\ldots,l_k)$ in terms of elements of $\mathcal{B}(k,a)$, $B(l_1,\ldots,l_k)$ in terms of elements of $\mathcal{X}(k,a)$ and $X(l_1,\ldots,l_k)-B(l_1,\ldots,l_k)$ in terms of elements of $\mathcal{X}(k-1,a)$ so that the coefficients do not depend on n. When $k=1$, $X(l_1)=B(l_1)$ for any l_1, and the result holds. For the induction step, we assume that the claim holds for $\leq k$, and consider the $k+1$ case. We will make use of the following identity:

$$\begin{aligned}
&X(l_1,\ldots,l_{k+1})-B(l_1,\ldots,l_{k+1})\\
&\quad=\frac{1}{k+1}\sum_{j=1}^{k+1}(X(l_1,\ldots,\hat{l_j},\ldots,l_{k+1})-B(l_1,\ldots,\hat{l_j},\ldots,l_{k+1}))\\
&\qquad\times\left(\sum_{i=1}^{n} i^{2l_j}X_i\right)\\
&\qquad+\frac{1}{k+1}\sum_{j=1}^{k+1}\sum_{j'\neq j}B(l_1,\ldots,l_{j'}+l_j,\ldots,\hat{l_j},\ldots,l_{k+1}).
\end{aligned}$$

The identity can be proved by matching individual terms on both sides. Now by the induction hypothesis, when $l_1+\cdots+l_{k+1}=a$, $X(l_1,\ldots,\hat{l_j},\ldots,l_{k+1})-$

$B(l_1, \ldots, \hat{l_j}, \ldots, l_{k+1})$ is in $\tilde{\mathcal{X}}(k-1, a-l_j)$. Therefore, the first term of the right-hand side is in $\tilde{\mathcal{X}}(k, a)$. The second term of the right-hand side is in $\tilde{\mathcal{B}}(k, a)$, hence in $\tilde{\mathcal{X}}(k, a)$. Thus the left-hand side is in $\tilde{\mathcal{X}}(k, a) = \tilde{\mathcal{B}}(k, a)$. Hence $\tilde{\mathcal{X}}(k+1, a) = \tilde{\mathcal{B}}(k+1, a)$. The last conclusion can also be checked from the identity.

Now $\mathbb{E}\left[\prod_{r=1}^{k}(\sum_{i=1}^{n} i^{2l_r}(a_i + b_i))\right]$ can be expanded in terms of the forms (coefficients do not depend on n, j)

$$\mathbb{E}\left[\prod_{r\in\Gamma}\left(\sum_{i=1}^{n} i^{2l_r} a_i\right)\right]\mathbb{E}\left[\prod_{r\in[k]\setminus\Gamma}\left(\sum_{i=1}^{n} i^{2l_r} b_i\right)\right] \tag{3.16}$$

where $\Gamma \subseteq [k]$. By the above result, the first factor can be expressed in terms of elements in $\mathcal{B}(|\Gamma|, \sum_{r\in\Gamma} l_r)$, and similarly for the second factor.

Now we denote by $B'(l_1, \ldots, l_k)$ the result obtained by substituting $X_i = a_i$ in $B(l_1, \ldots, l_k)$, and $B''(l_1, \ldots, l_k)$ similar with $X_i = b_i$. Note that by Lemma 2,

$$\mathbb{E}[B'(l_1, \ldots, l_k)] = \sum_{i_1+\cdots+i_k\le j} i_1^{2l_1-1} \ldots i_k^{2l_k-1}. \tag{3.17}$$

It can be shown by induction on k that the above expression is a polynomial in j with degree $2(l_1+\cdots+l_k)$. Similarly, it can be shown that $\mathbb{E}[B''(l_1, \ldots, l_k)]$ is a polynomial in $(n-j)$ of degree $2(l_1+\cdots+l_k)$, hence it is also a polynomial of the same degree in j with coefficients depending on n (but the coefficient of highest degree does not depend on n). Putting these together, we conclude that $\mathbb{E}\left[\prod_{r=1}^{k}(\sum_{i=1}^{n} i^{2l_r}(a_i + b_i))\right]$ is a polynomial of degree $\le 2a$ in j, and the coefficient of its degree $2a$ term does not depend on n.

We conclude that W_{2a} is a polynomial of degree $\le 2a$ in j, and its coefficient of degree $2a$ does not depend on n for $n \ge 2a$. This implies that the coefficient of degree $2a$ of $\mathbb{E}[X^{2a}]$ does not depend on n for $n \ge 2a$. By Lemma 1, the eigenfunctions of K are given by the discrete Chebyshev polynomials, and the eigenvalue corresponding to T_{2k}^{n} does not depend on n as long as $n \ge 2k$. □

Now we present the proof for the eigenvalues. We make use of three lemmas in the proof, which we also present below. The first gives needed inequalities for the gamma function. For many further references, see [24].

Lemma 4 (Explicit Stirling approximation) *For any* $x > 0$,

$$\begin{aligned}&\sqrt{\pi}\left(\frac{x}{e}\right)^x\left(8x^3+4x^2+x+\frac{1}{100}\right)^{\frac{1}{6}}\\ &\quad<\Gamma(1+x)<\sqrt{\pi}\left(\frac{x}{e}\right)^x\left(8x^3+4x^2+x+\frac{1}{30}\right)^{\frac{1}{6}}.\end{aligned} \tag{3.18}$$

From this it can be easily derived that for any $x \geq 1$,

$$\sqrt{2\pi}\left(\frac{x}{e}\right)^x \sqrt{x} < \Gamma(1+x) < \sqrt{2\pi}\left(\frac{x}{e}\right)^x \sqrt{x}\left(1+\frac{1}{x}\right)^{\frac{1}{6}}\Bigg). \tag{3.19}$$

Lemma 5 (Clausen [11], see also [36])

$${}_3F_2\left(2a, 2b, a+b; a+b+\frac{1}{2}, 2a+2b|x\right) = \left({}_2F_1\left(a, b; a+b+\frac{1}{2}|x\right)\right)^2. \tag{3.20}$$

Lemma 6 (Gauss's hypergeometric theorem, see Page 2 of [5]) *For* $a, b, c \in \mathbb{R}$ *such that* $a + b < c$, *we have*

$${}_2F_1(a, b; c|1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}. \tag{3.21}$$

Proof of Theorem 2, eigenvalue part We use the notation as in the statement of Theorem 2. Let $\lambda_{k,n}$ be the eigenvalue corresponding to the eigenfunction T_{2k}^n for the Markov chain $Q(i, j)$ on $\{0, 1, \ldots, n\}$. Note that from the determination of eigenfunctions, $\lambda_{k,n}$ is constant for all $n \geq 2k$, and we denote it by λ_k. The strategy of the proof is to analyze the expression for λ_k in a system of size n and let $n \to \infty$ to get the desired result.

The eigenfunction that corresponds to λ_k has been proven previously to be the discrete Chebyshev polynomials T_{2k}^n. Thus by examining the first row of the eigenvalue–eigenfunction equation, we obtain that (see Proposition 4 and [15, (3.1)–(3.3)])

$$\lambda_{k,n} = \frac{\sum_{j=0}^n \alpha_j^n T_{2k}^n(j)}{T_{2k}^n(0)} = \sum_{j=0}^n \alpha_j^n T_{2k}^n(j), \tag{3.22}$$

where $\alpha_j^n = \frac{\binom{2j}{j}\binom{2(n-j)}{n-j}}{2^{2n}}$.

Plugging the expression

$$T_{2k}^n(j) = \sum_{l=0}^{2k} \frac{(-2k)_l (2k+1)_l (-j)_l}{(l!)^2 (-n)_l} \tag{3.23}$$

into the expression above, we obtain that

$$\lambda_k = \sum_{l=0}^{2k} \frac{(-2k)_l (2k+1)_l}{(l!)^2} \sum_{j=0}^n \frac{(-j)_l \alpha_j^n}{(-n)_l}. \tag{3.24}$$

Below we prove that

$$\lim_{n\to\infty}\sum_{j=0}^{n}\frac{(-j)_l\alpha_j^n}{(-n)_l}=\frac{(\frac{1}{2})_l}{l!}. \tag{3.25}$$

We fix a small $c>0$, and assume that n is sufficiently large below. Note that by Lemma 4, we can derive that for n sufficiently large,

$$\left|\sum_{j\le cn}\frac{(-j)_l\alpha_j^n}{(-n)_l}\right|\le\sum_{j\le cn}\frac{\binom{2j}{j}\binom{2(n-j)}{n-j}}{2^{2n}}\le\frac{2}{\sqrt{n}}+\sum_{1\le j\le cn}\frac{4}{\sqrt{j}\sqrt{n-j}}\le\frac{10\sqrt{c}}{\sqrt{1-c}}. \tag{3.26}$$

Similarly,

$$\left|\sum_{(1-c)n\le j\le n}\frac{(-j)_l\alpha_j^n}{(-n)_l}\right|\le\frac{10\sqrt{c}}{\sqrt{1-c}}. \tag{3.27}$$

Now note that for n sufficiently large, when $cn<j<(1-c)n$, we have $(\frac{j}{n})^l(1-\frac{l+1}{cn})^l\le\frac{(-j)_l}{(-n)_l}\le(\frac{j}{n})^l$. Moreover, using Lemma 4, we obtain that $\frac{1}{\pi}\frac{1}{\sqrt{j}\sqrt{n-j}}(1+\frac{6}{cn})^{-\frac{2}{3}}\le\alpha_j^n\le\frac{1}{\pi}\frac{1}{\sqrt{j}\sqrt{n-j}}(1+\frac{6}{cn})^{\frac{1}{3}}$ for $cn<j<(1-c)n$. Therefore, we have

$$\sum_{cn<j<(1-c)n}\frac{(-j)_l\alpha_j^n}{(-n)_l}\le\left(1+\frac{6}{cn}\right)^{\frac{1}{3}}\frac{1}{\pi}\sum_{cn<j<(1-c)n}\frac{1}{\sqrt{\frac{j}{n}}\sqrt{1-\frac{j}{n}}}\left(\frac{j}{n}\right)^l\frac{1}{n}, \tag{3.28}$$

$$\sum_{cn<j<(1-c)n}\frac{(-j)_l\alpha_j^n}{(-n)_l}\ge\left(1+\frac{6}{cn}\right)^{-\frac{2}{3}}(1-\frac{l+1}{cn})^l\frac{1}{\pi}\sum_{cn<j<(1-c)n}\frac{1}{\sqrt{\frac{j}{n}}\sqrt{1-\frac{j}{n}}}\left(\frac{j}{n}\right)^l\frac{1}{n}. \tag{3.29}$$

Now note that

$$\lim_{n\to\infty}\frac{1}{\pi}\sum_{cn<j<(1-c)n}\frac{1}{\sqrt{\frac{j}{n}}\sqrt{1-\frac{j}{n}}}\left(\frac{j}{n}\right)^l\frac{1}{n}=\frac{1}{\pi}\int_c^{1-c}\frac{x^l}{\sqrt{x}\sqrt{1-x}}. \tag{3.30}$$

Hence we have

$$\frac{1}{\pi}\int_c^{1-c}\frac{x^l}{\sqrt{x}\sqrt{1-x}}-\frac{20\sqrt{c}}{\sqrt{1-c}}\le\liminf_{n\to\infty}\sum_{j=0}^{n}\frac{(-j)_l\alpha_j^n}{(-n)_l}$$

$$\leq \limsup_{n\to\infty} \sum_{j=0}^{n} \frac{(-j)_l \alpha_j^n}{(-n)_l} \leq \frac{1}{\pi} \int_c^{1-c} \frac{x^l}{\sqrt{x}\sqrt{1-x}} + \frac{20\sqrt{c}}{\sqrt{1-c}}.$$

Sending $c \to 0$ gives

$$\lim_{n\to\infty} \sum_{j=0}^{n} \frac{(-j)_l \alpha_l^n}{(-n)_l} = \frac{1}{\pi} \int_0^1 \frac{x^l}{\sqrt{x}\sqrt{1-x}} = \frac{(\frac{1}{2})_l}{l!}. \tag{3.31}$$

Now in (3.24) we take $n \to \infty$ and get

$$\lambda_k = \sum_{l=0}^{2k} \frac{(-2k)_l (2k+1)_l (\frac{1}{2})_l}{(l!)^3} = {}_3F_2\left(\frac{1}{2}, -2k, 2k+1; 1, 1|1\right). \tag{3.32}$$

Take $a = -k$ and $b = k + \frac{1}{2}$ in Lemma 5, we obtain that

$$\lambda_k = ({}_2F_1(-k, k+\frac{1}{2}; 1|1))^2. \tag{3.33}$$

Finally, by Lemma 6,

$${}_2F_1\left(-k, k+\frac{1}{2}; 1|1\right) = \frac{\Gamma(1)\Gamma(\frac{1}{2})}{\Gamma(k+1)\Gamma(\frac{1}{2}-k)} = \frac{(-1)^k 1 \cdot 3 \cdot \cdots \cdot (2k-1)}{2 \cdot 4 \cdot \cdots \cdot (2k)}. \tag{3.34}$$

Hence

$$\lambda_k = \frac{\binom{2k}{k}^2}{2^{4k}} \tag{3.35}$$

□

3.2 Proof of Theorem 1

In this part, we prove Theorem 1 based on Theorem 2.

Proof of Theorem 1 First note that for the starting state $x_0 = (1, 1, \ldots, 1)$, by S_n-invariance of $K^l(x, y)$ and $\pi(x)$ for any x, y, l, we have

$$\|K_{x_0}^l - \pi\| = \frac{1}{2} \sum_{y\in\mathcal{X}} |K_{x_0}^l(y) - \pi(y)| = \frac{1}{2} \sum_{j=0}^{n} |Q_n^l(j) - u(j)| = \|Q_n^l - u\|,$$

where $u(j) = \frac{1}{n+1}, 0 \leq j \leq n$.

Without loss of generality suppose n is even. The analysis for odd n is similar. By Theorem 2,

$$\lambda_k = \frac{\binom{2k}{k}^2}{2^{4k}}. \tag{3.36}$$

By Lemma 4, we obtain that

$$\lambda_k \leq \frac{1}{\pi k}\left(1+\frac{1}{2k}\right)^{\frac{1}{3}}. \tag{3.37}$$

First we show the lower bound. By Proposition 1, taking $\beta = \lambda_1 = \frac{1}{4}$ and $\psi = T_2^n$, we obtain

$$\|Q_n^l - u\| \geq \frac{1}{4}\left(\frac{1}{4}\right)^l. \tag{3.38}$$

Now we show the upper bound. For $l = 1$ the upper bound follows by the fact that total variation distance is always upper bounded by 1. For $l \geq 2$,

$$4\|Q_n^l - u\|^2 \leq \chi_n^2(l) = \sum_{k=1}^{\frac{n}{2}} \lambda_k^{2l}\beta_{2k}(4k+1), \tag{3.39}$$

where $\beta_{2k} \leq \frac{n}{n+2} \leq 1$ (see [19, Sect. 2.5] for the estimates related to Hahn polynomials that are used above). Therefore, we obtain that

$$\begin{aligned}\chi_n^2(l) &\leq 5\sum_{k=1}^{\frac{n}{2}} k\lambda_k^{2l} \leq 15\left(\frac{1}{16}\right)^l + \sum_{k=3}^{\frac{n}{2}} k\left(\frac{2}{\pi k}\right)^{2l} \\ &\leq 15\left(\frac{1}{16}\right)^l + 27\left(\frac{2}{3\pi}\right)^{2l}\sum_{k=1}^{\infty}\frac{1}{k^2} \leq 60\left(\frac{1}{16}\right)^l.\end{aligned}$$

Hence we conclude that

$$\|Q_n^l - u\| \leq 4\left(\frac{1}{4}\right)^l \tag{3.40}$$

for any $l \geq 2$.

Therefore, for any $l \in \{1, 2, \ldots\}$,

$$\frac{1}{4}\left(\frac{1}{4}\right)^l \leq \|K_{x_0}^l - \pi\| \leq 4\left(\frac{1}{4}\right)^l. \tag{3.41}$$

□

4 The twisted Burnside process

In this section, we consider a generalization of the original Burnside process, which we will call "twisted Burnside process." This will allow deformation of the examples above to give a natural process with a larger family of Hahn polynomials as eigenfunctions. Several further examples are presented. We believe that it gives many new examples of easy to run, rapidly converging Markov chains with tunable stationary distributions.

The setting is the same: we have a finite group G acting on a finite set $\mathcal{X}$; for $x \in \mathcal{X}$, let $G_x = \{g \in G : x^g = x\}$; for $g \in G$, let $\mathcal{X}_g = \{x \in \mathcal{X} : x^g = x\}$.

We choose a positive weight w on the group G and let $W(x)$ be the sum of $w(g)$ for $g \in G_x$. We also choose a positive weight v on the set $\mathcal{X}$ and let $V(g)$ be the sum of $v(x)$ for $x \in \mathcal{X}_g$. The new Markov chain is as follows: from x, choose $g \in G_x$ with probability $\frac{w(g)}{W(x)}$; given g, choose $y \in \mathcal{X}_g$ with probability $\frac{v(y)}{V(g)}$; the chain goes from x to y.

Proposition 2 below gives the stationary distribution of the twisted Burnside process.

Proposition 2 *The twisted Burnside process as discussed in the preceding is a reversible Markov chain with stationary distribution*

$$\pi(x) \propto W(x)v(x) \tag{4.1}$$

for $x \in \mathcal{X}$.

Moreover, if w is constant on each conjugacy class of G and v is constant on each orbit of $\mathcal{X}$ (under the action of G), then the chain can be lumped onto orbits of $\mathcal{X}$. For $x \in \mathcal{X}$, let O_x denote the orbit containing x. The lumped chain is a reversible Markov chain with stationary distribution

$$\tilde{\pi}(O_x) \propto \frac{W(x)v(x)}{|G_x|}. \tag{4.2}$$

Proof For any $x, y \in \mathcal{X}$, the transition probability from x to y of the twisted Burnside process is given by

$$K_{x,y} = \frac{v(y)}{W(x)} \sum_{g \in G_x \cap G_y} \frac{w(g)}{V(g)}. \tag{4.3}$$

Note that the factor $\sum_{g \in G_x \cap G_y} \frac{w(g)}{V(g)}$ is symmetric in x, y. Hence if $\pi(x) = \frac{W(x)v(x)}{Z}$ for $x \in \mathcal{X}$ (where Z is a normalizing constant), then we have

$$\pi(x)K_{x,y} = \frac{v(x)v(y)}{Z} \sum_{g \in G_x \cap G_y} \frac{w(g)}{V(g)} = \pi(y)K_{y,x}. \tag{4.4}$$

This shows that the twisted Burnside process is reversible with stationary distribution proportional to $W(x)v(x)$.

For the second part, we assume that w is constant on each conjugacy class of G and v is constant on each orbit of $\mathcal{X}$. Below we show by Dynkin's criterion (see [30, p. 133]) that in this case the twisted Burnside process can be lumped onto orbits. Suppose that $x, y \in \mathcal{X}$ are in the same orbit. Then there exists $h \in G$ such that $y = x^h$. Hence

$$G_y = h^{-1} G_x h. \tag{4.5}$$

As w is constant on each conjugacy class of G, by (4.5) we have

$$W(y) = \sum_{g \in G_y} w(g) = \sum_{g \in G_x} w(h^{-1}gh) = \sum_{g \in G_x} w(g) = W(x). \tag{4.6}$$

For any $z \in \mathcal{X}$, we have

$$\begin{aligned} &\sum_{q \in O_z} v(q) \sum_{g \in G_y \cap G_q} \frac{w(g)}{V(g)} \\ &= \sum_{q \in O_z} v(q^h) \sum_{g \in G_y \cap G_{q^h}} \frac{w(g)}{V(g)} \\ &= \sum_{q \in O_z} v(q) \sum_{g \in h^{-1}(G_x \cap G_q)h} \frac{w(g)}{V(g)} \\ &= \sum_{q \in O_z} v(q) \sum_{g \in G_x \cap G_q} \frac{w(h^{-1}gh)}{V(h^{-1}gh)}, \end{aligned} \tag{4.7}$$

where the second equality follows from the fact that v is constant on each orbit of $\mathcal{X}$. Now note that

$$\mathcal{X}_{h^{-1}gh} = \{x^h : x \in \mathcal{X}_g\}. \tag{4.8}$$

By (4.8) and the fact that v is constant on each orbit of $\mathcal{X}$, we have

$$V(h^{-1}gh) = \sum_{x \in \mathcal{X}_{h^{-1}gh}} v(x) = \sum_{x \in \mathcal{X}_g} v(x^h) = \sum_{x \in \mathcal{X}_g} v(x) = V(g). \tag{4.9}$$

By (4.7), (4.9), and the fact that w is constant on each conjugacy class of G, we have

$$\sum_{q \in O_z} v(q) \sum_{g \in G_y \cap G_q} \frac{w(g)}{V(g)} = \sum_{q \in O_z} v(q) \sum_{g \in G_x \cap G_q} \frac{w(g)}{V(g)}. \tag{4.10}$$

Therefore, by (4.3), (4.6), and (4.10), we obtain that

$$\sum_{q \in O_z} K_{x,q} = \sum_{q \in O_z} K_{y,q}. \tag{4.11}$$

By Dynkin's criterion, the chain can be lumped onto orbits of $\mathcal{X}$. Note that by (4.6) and the fact that $v(x)$ is constant on each orbit, we have $W(x)v(x)$ is constant on every orbit. Thus the lumped chain is reversible with stationary distribution

$$\tilde{\pi}(O_x) \propto |O_x| W(x) v(x). \tag{4.12}$$

By the orbit–stabilizer theorem, we have

$$\tilde{\pi}(O_x) \propto \frac{W(x)v(x)}{|G_x|}. \tag{4.13}$$

□

An example of the twisted version of the Burnside process considered in Theorem 1 is presented below. Specializing $k = 2$ and $\gamma_2 = 1$ in the example gives the chain with beta-binomial stationary distribution considered in Sect. 5 below.

Example 1 Consider $\mathcal{X} = [k]^n$ and $G = S_n$ with G acting on $\mathcal{X}$ by permuting coordinates. Fix k positive parameters $\theta, \gamma_2, \ldots, \gamma_k$. We take

$$w(\sigma) = \theta^{c(\sigma)}, \tag{4.14}$$

for every $\sigma \in S_n$, where $c(\sigma)$ is the number of cycles of σ. For any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}$ and $j \in [k]$, we define

$$S(\mathbf{x}, j) := \#\{i \in [n] : x_i = j\}. \tag{4.15}$$

We further take

$$v(\mathbf{x}) = \prod_{j=2}^{k} \gamma_j^{S(\mathbf{x}, j)} \tag{4.16}$$

for every $\mathbf{x} \in \mathcal{X}$. Below we let $\gamma_1 := 1$ to simplify notation.

Now we discuss the twisted Markov chain. From $\mathbf{x} \in \mathcal{X}$, we choose $\sigma \in S_n$ fixing $\mathbf{x}$ with probability $\frac{w(\sigma)}{W(\mathbf{x})}$. This can be realized as follows: find the set of indices $I_j := \{l \in [n] : x_l = j\}$ for each $j \in [k]$; for each I_j, sample a permutation $\sigma_j \in S_{I_j}$ from the Ewens distribution with parameter θ (see Sect. 5 for details of the Ewens distribution); σ is the product of σ_j for $j \in [k]$.

Given σ, we choose $\mathbf{y} \in \mathcal{X}$ fixed by σ with probability $\frac{v(\mathbf{y})}{V(\sigma)}$. This can be done as below: break σ into cycles $C_1, \cdots, C_m$, and denote by c_d the length of the cycle C_d for every $d \in [m]$; for every $d \in [m]$, pick an integer $r_d \in [k]$ with probability $\frac{\gamma_{r_d}^{c_d}}{\sum_{l=1}^{k} \gamma_l^{c_d}}$, and take $y_i = r_d$ for every $i \in C_d$. Note that for any $\mathbf{y} \in \mathcal{X}_\sigma$, y_i for $i \in C_d$ takes the same value (assuming that it's r_d). Hence the probability of generating $\mathbf{y}$

(where $\mathbf{y} \in \mathcal{X}_\sigma$) through this procedure is

$$\prod_{d=1}^{m} \frac{\gamma_{r_d}^{c_d}}{\sum_{l=1}^{k} \gamma_l^{c_d}} \propto \prod_{d=1}^{m} \gamma_{r_d}^{c_d} = \prod_{j=2}^{k} \gamma_j^{S(\mathbf{y},j)}. \tag{4.17}$$

Note that w is constant on each conjugacy class of G. Moreover, $v(\mathbf{x})$ only depends on $S(\mathbf{x}, j)$ for $j \in [k]$, hence v is constant on each orbit of $\mathcal{X}$. By Proposition 2, the twisted Markov chain can be lumped onto orbits of $\mathcal{X}$. For $t_1, \ldots, t_k \in \mathbb{N}$ such that $\sum_{l=1}^{k} t_l = n$, let $\mathbf{t} := (t_1, \ldots, t_k)$ denote the orbit of $\mathcal{X}$ consisting of $\mathbf{x} = (x_1, \ldots, x_n)$ such that $S(\mathbf{x}, j) = t_j$ for every $j \in [k]$. Note that for any $\mathbf{x}$ in the orbit $\mathbf{t}$, we have $|G_\mathbf{x}| = \prod_{j=1}^{k} (t_j)!$ and

$$W(\mathbf{x}) = \sum_{\sigma \in S_n : \sigma \text{ fixes } \mathbf{x}} \theta^{c(\sigma)} = \prod_{j=1}^{k} (\theta \cdots (\theta + t_j - 1)) \propto \prod_{j=1}^{k} \Gamma(t_j + \theta). \tag{4.18}$$

Thus by Proposition 2, the stationary distribution of the lumped chain is

$$\tilde{\pi}(\mathbf{t}) \propto \prod_{j=1}^{k} \frac{\Gamma(t_j + \theta)}{(t_j)!} \prod_{j=2}^{k} \gamma_j^{t_j}. \tag{4.19}$$

Note that the term $\prod_{j=1}^{k} \frac{\Gamma(t_j+\theta)}{(t_j)!}$ is proportional to the probability corresponding to the symmetric Dirichlet-multinomial distribution of parameter $(\theta, \ldots, \theta)$. Thus the twisted Burnside process offers a k-dimensional deformation of this classical distribution. Specializing to the case of $k = 2$ and choosing parameters so that the base beta-binomial is uniform on $\{0, 1, \ldots, n\}$, the deformation is a discrete exponential distribution truncated to this interval.

We close with a final remark. There are many probability measures on S_n that are constant on conjugacy classes. One way to construct these is to define $P(\sigma)$ as proportional to $\theta^{d(\sigma)}$ where $d(\sigma) = d(id, \sigma)$ for d a bi-invariant metric on S_n. In turn, such bi-invariant metrics can be constructed as follows: Let $\rho : S_n \to GL(V)$ be a faithful unitary representation of S_n. Let $\|.\|$ be a unitarily invariant norm on V. Then $d(\sigma, \tau) = \|\rho(\sigma) - \rho(\tau)\|$ is a bi-invariant metric on S_n. In particular, the Cayley distance $d(\sigma, \tau) = n - C(\sigma\tau^{-1}) = \min \#\text{transpositions required to bring } \sigma \text{ to } \tau$ is bi-invariant, giving the example used above. Similarly, the Hamming distance $\#\{i \text{ with } \sigma(i) \text{ different from } \tau(i)\}$ is bi-invariant. A host of other examples appear in [14, Chap. 6C]. This includes von Neumann's useful characterization of unitarily invariant matrix norms.

5 The twisted Burnside process and Hahn polynomials

In this section, we add a parameter to the Burnside process on C_2^n with the group S_n so that more general Hahn polynomials appear as eigenfunctions. The idea is simple:

replace the uniform distribution on S_n, used in step one of the algorithm, by the Ewens distribution

$$\mathbb{P}_\theta(\sigma) = \frac{1}{(\theta)_n}\theta^{c(\sigma)}, \quad (\theta)_n = \theta(\theta+1)\cdots(\theta+n-1) \tag{5.1}$$

where $c(\sigma)$ is the number of cycles in σ and $0 < \theta < \infty$ is a parameter. This familiar distribution is studied in genetics and combinatorics [12]. It may be seen as 'the Mallows model through the Cayley metric' as discussed at the end of Sect. 4. This chain was discovered via the twisted Burnside construction of Sect. 4. In hindsight, the following simplified description is available.

On C_2^n, from $x \in C_2^n$ with $|x| = \#$ ones in x

- Identify G_x with $S_{|x|} \times S_{n-|x|}$
- Pick $\sigma \in S_{|x|} \times S_{n-|x|}$ choosing the two components independently from the Ewens measure (5.1)
- Break σ into cycles and label the cycles $0/1$ with probability $\frac{1}{2}$. Put this $0/1$ string into $y \in C_2^n$.

The argument of Sect. 4 shows

Proposition 3 *The twisted Burnside process given above, lumped to orbits, is a Markov chain on* $\{0, 1, \ldots, n\}$ *with a beta-binomial distribution having parameters* $\alpha = \beta = \theta$ *(see (2.3)).*

The transition matrix of this Markov chain, call it $p_{ij}^{n,\theta}$, can be written explicitly.

Proposition 4 *Consider the twisted Burnside process given above, lumped to orbits. The transition matrix is given by*

$$p_{0j}^{n,\theta} = \binom{n}{j}\frac{\frac{\theta}{2}\cdots(\frac{\theta}{2}+j-1)\frac{\theta}{2}\cdots(\frac{\theta}{2}+n-j-1)}{\theta(\theta+1)\cdots(\theta+n-1)}, \tag{5.2}$$

$$p_{jk}^{n,\theta} = \sum_{\max\{0,j+k-n\}\le l\le \min\{j,k\}} p_{0l}^{j,\theta}\, p_{0,k-l}^{n-j,\theta}. \tag{5.3}$$

Proof We prove this using Pólya's cycle index theorem (see (3.1),(3.2)). Note that we have

$$p_{0j}^{n,\theta} = \frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)}\frac{1}{n!}\sum_{g\in S_n,\lambda\vdash j}\prod_{i=1}^{j}\binom{a_i(g)}{b_i(\lambda)}(\frac{\theta}{2})^{a_1(g)+\cdots+a_n(g)}. \tag{5.4}$$

Using Pólya's cycle index theorem by taking derivatives and multiplying, we get

$$p_{0j}^{n,\theta} = \binom{n}{j}\frac{\frac{\theta}{2}\cdots(\frac{\theta}{2}+j-1)\frac{\theta}{2}\cdots(\frac{\theta}{2}+n-j-1)}{\theta(\theta+1)\cdots(\theta+n-1)}. \tag{5.5}$$

Moreover, by the definition of the twisted Burnside process given above, we have

$$p_{jk}^{n,\theta} = \sum_{\max\{0,j+k-n\}\le l\le \min\{j,k\}} p_{0l}^{j,\theta}\, p_{0,k-l}^{n-j,\theta}. \tag{5.6}$$

□

Theorems 1 and 2 above deform in the following form.

Theorem 3 *Consider the twisted Burnside process on C_2^n given above, lumped to orbits. The non-zero eigenvalues of the Markov chain are given by* 1 *and*

$$\lambda_k = {}_3F_2\left(-2k, 2k+2\theta-1, \frac{\theta}{2}; \theta, \theta|1\right) \tag{5.7}$$

for $1 \le k \le \frac{n}{2}$. Moreover, the eigenfunctions corresponding to the zero eigenvalues are the Hahn polynomials on $\{0, 1, \ldots, n\}$ with parameters $\alpha = \beta = \theta$ of odd degree. The eigenfunction corresponding to the eigenvalue λ_k is the Hahn polynomial on $\{0, 1, \ldots, n\}$ with parameters $\alpha = \beta = \theta$ of degree $2k$, $1 \le k \le \frac{n}{2}$.

Theorem 4 *Consider the twisted Burnside process on C_2^n given above with $\theta \ge 1$. Denote by π the stationary distribution of the chain (see Proposition 3), and denote by $K_{x_0}^l$ for $x_0 = (1, 1, \ldots, 1)$ the distribution after l steps starting from n ones. Then there exist positive constants $c(\theta), C(\theta)$ which only depend on θ, such that for all $n \ge 2$ and all $l \ge 1$,*

$$c(\theta)\left(\frac{1}{2(1+\theta)}\right)^l \le \|K_{x_0}^l - \pi\| \le C(\theta)\left(\frac{1}{2(1+\theta)}\right)^l. \tag{5.8}$$

The proofs of Theorems 3 and 4 are similar but quite a bit more involved, to the proofs of Theorems 1 and 2. The restriction that $\theta \ge 1$ in Theorem 4 is due to a technical step in our proof (for certain estimates of the eigenvalues). We refer the interested reader to [43,44]. This develops things for $k \ge 2$ and has other approaches to proof.

We have not (yet) succeeded in finding a two-parameter deformation of the Burnside process on C_2^n which gives the full set of Hahn polynomials as eigenfunctions. Similarly, we have not succeeded in diagonalizing the Burnside process on $[k]^n$ for any $k \ge 3$.

The point of this paper was to show (a) that orthogonal polynomials 'pop up' everyplace (b) seeing orthogonal polynomials as belonging to families leads to useful extension of classical algorithms.

We are sorry not to be able to ask Dick Askey for further help.

6 A continuous limit of the Burnside process

A referee has made the welcome suggestion that we try to 'pass to the limit' going from our discrete version of the Burnside process analyzed above to a continuous process.

While we do not know a continuous version of the Cauchy–Frobenius Lemma or Pólya theory, we were able to pass to the limit and this proved informative.

As motivation, recall that the discrete process begins with a point $x \in C_2^n$. A permutation in $S_k \times S_{n-k}$ is chosen at random (there are k ones in x), split into cycles, and these are labeled 0/1 to give $y \in C_2^n$. As explained, only the number of ones enters, not their positions. So the process can be thought of as taking place on $\{0, 1, 2, \ldots, n\}$. We divide by n and form a process on $[0, 1]$. The analog of the cycles of a random permutation is replaced by a stick-breaking process on $[0, 1]$ familiar from the Chinese restaurant process and Dirichlet random measures [4,32,39]. Combining gives the following Markov chain on $[0, 1]$:

From $x \in [0, 1]$, break the interval $[0, x]$ into countably many pieces by a stick-breaking process. Namely, let $R_1, R_2, \ldots$ be independent $Beta(1, \theta)$ random variables, and define $Y_1 = xR_1$ and $Y_j = x(1 - R_1) \cdots (1 - R_{j-1})R_j$ for every $j = 2, 3, \ldots$; then we break $[0, x]$ into pieces of lengths $Y_1, Y_2, \ldots$. Break the interval from x to 1 in the same way. Label each interval 0/1 by flipping a fair coin. Let y be the total length of the pieces labeled 1. This gives a Markov chain on $[0, 1]$. It is a natural limiting version of our discrete Burnside process.

This Markov chain can be equivalently described as follows ([13, Sect. 3], [34,39]). From $x \in [0, 1]$, sample two independent $Beta(\frac{\theta}{2}, \frac{\theta}{2})$ random variables Z, Z'. Let $y = xZ + (1 - x)Z'$, and move to y.

Theorem 5 *The Markov chain above is reversible with $Beta(\theta, \theta)$ stationary distribution. The non-zero eigenvalues are given by* 1 *and*

$$\lambda_k = {}_3F_2(-2k, 2k + 2\theta - 1, \frac{\theta}{2}; \theta, \theta|1) \tag{6.1}$$

for $k = 1, 2, \ldots$. An alternative expression for λ_k is given by

$$\lambda_k = \mathbb{E}[(Z - Z')^{2k}], \tag{6.2}$$

where Z, Z' are two independent $Beta(\frac{\theta}{2}, \frac{\theta}{2})$ random variables.

Moreover, the eigenfunctions corresponding to the zero eigenvalues are the Jacobi polynomials associated to the stationary distribution of odd degree. For every $k = 1, 2, \ldots$, the eigenfunction corresponding to the eigenvalue λ_k is the Jacobi polynomial associated to the stationary distribution of degree $2k$.

Proof From the definition of the Markov chain, the transition density is given by

$$k(x, y) = \int_{\max\{0, x+y-1\}}^{\min\{x,y\}} \frac{z^{\frac{\theta}{2}-1}(x - z)^{\frac{\theta}{2}-1}}{B(\frac{\theta}{2}, \frac{\theta}{2})} \frac{(y - z)^{\frac{\theta}{2}-1}(1 - x - y + z)^{\frac{\theta}{2}-1}}{B(\frac{\theta}{2}, \frac{\theta}{2})} \mathrm{d}z \times x^{1-\theta}(1 - x)^{1-\theta} \tag{6.3}$$

for $x, y \in [0, 1]$. Let $\pi(x) = \frac{x^{\theta-1}(1-x)^{\theta-1}}{B(\theta,\theta)}, x \in [0, 1]$ be the probability density function of $Beta(\theta, \theta)$. We have

$$\pi(x)k(x, y) = \pi(y)k(y, x)$$

for any $x, y \in (0,1)$. Therefore, the Markov chain is reversible with $Beta(\theta,\theta)$ stationary distribution.

Suppose that we start from $x \in [0,1]$. Sample two independent $Beta(\frac{\theta}{2}, \frac{\theta}{2})$ random variables Z, Z', and let $y = xZ + (1-x)Z'$. This gives one iteration of the Markov chain. For every $l = 1, 2, \ldots$

$$\mathbb{E}[y^l|x] = \mathbb{E}[(Z-Z')^l]x^l + \sum_{j=0}^{l-1} \binom{l}{j} \mathbb{E}[(Z-Z')^j (Z')^{l-j}]x^j. \tag{6.4}$$

The right-hand side of (6.4) is a polynomial in x of degree $\leq l$ with leading coefficient given by $\mathbb{E}[(Z-Z')^l]$. When l is odd, as the distribution of $Z - Z'$ is symmetric around 0, we have $\mathbb{E}[(Z-Z')^l] = 0$.

By Cannings argument (the analog of Lemma 1), the non-zero eigenvalues of the Markov chain are given by 1 and $\lambda_k := \mathbb{E}[(Z-Z')^{2k}]$ for $k = 1, 2, \ldots$. Moreover, the eigenfunctions corresponding to the zero eigenvalues are the Jacobi polynomials associated to $Beta(\theta,\theta)$ of odd degree, and the eigenfunction corresponding to the eigenvalue λ_k is the Jacobi polynomial associated to $Beta(\theta,\theta)$ of degree $2k$ for every $k = 1, 2, \ldots$.

Finally, we show that

$$\lambda_k = {}_3F_2\left(-2k, 2k+2\theta-1, \frac{\theta}{2}; \theta, \theta|1\right).$$

We denote by

$$\phi_{2k}(x) = {}_2F_1(-2k, 2\theta+2k-1; \theta|x) = \sum_{l=0}^{2k} \frac{(-2k)_l (2\theta+2k-1)_l}{(\theta)_l} \frac{x^l}{l!} \tag{6.5}$$

the Jacobi polynomial associated to $Beta(\theta,\theta)$ of degree $2k$, normalized so that $\phi_{2k}(0) = 1$. Note that the one-step distribution starting from 0 follows the $Beta(\frac{\theta}{2}, \frac{\theta}{2})$ distribution. Thus letting $T \sim Beta(\frac{\theta}{2}, \frac{\theta}{2})$, we have

$$\mathbb{E}[\phi_{2k}(T)] = \lambda_k \phi_{2k}(0) = \lambda_k. \tag{6.6}$$

For every $l = 0, 1, 2, \ldots$

$$\mathbb{E}[T^l] = \frac{1}{B(\frac{\theta}{2}, \frac{\theta}{2})} \int_0^1 x^{l+\frac{\theta}{2}-1}(1-x)^{\frac{\theta}{2}-1} \mathrm{d}x = \frac{(\frac{\theta}{2})_l}{(\theta)_l}. \tag{6.7}$$

Hence

$$\lambda_k = \mathbb{E}[\phi_{2k}(T)] = \sum_{l=0}^{2k} \frac{(-2k)_l (2\theta+2k-1)_l}{(\theta)_l} \frac{\mathbb{E}[T^l]}{l!}$$

$$= {}_3F_2\left(-2k, 2k+2\theta-1, \frac{\theta}{2}; \theta, \theta|1\right).$$

□

Remark 1 Comparison with Theorems 2 and 3 shows that this limit captures the essential features we encountered. Note that the eigenvalue λ_k here matches that of the discrete chain in Theorem 3 as long as $n \geq 2k$.

Remark 2 As $Z, Z' \in [0, 1]$, we have $|Z - Z'| \leq 1$. Hence for any $k = 1, 2, \ldots$

$$\lambda_k = \mathbb{E}[|Z - Z'|^{2k}] \geq \mathbb{E}[|Z - Z'|^{2k+2}] = \lambda_{k+1}.$$

Therefore, the eigenvalues λ_k are monotone decreasing.

Remark 3 The two expressions (6.1) and (6.2) for λ_k lead to the following identity

$$\sum_{l=0}^{2k} \frac{(-2k)_l (\frac{\theta}{2})_l (\frac{\theta}{2})_{2k-l}}{(\theta)_l (\theta)_{2k-l} l!} = \sum_{l=0}^{2k} \frac{(-2k)_l (\frac{\theta}{2})_l (2k+2\theta-1)_l}{(\theta)_l (\theta)_l l!}. \tag{6.8}$$

We did not know this identity but Dennis Stanton observes that it is a special case of the following transformation

$$(a+A)_n \, {}_3F_2(a, c-b, -n; c, a+A|1) = (A)_n \, {}_3F_2(a, b, -n; c, 1-A-n|1) \tag{6.9}$$

when $n = 2k$, $a = \frac{\theta}{2}$, $c = \theta$, $b = 2\theta + 2k - 1$, $A = 1 - \theta - 2k$. One of his proofs of (6.9) proceeds by multiplying the Pfaff transformation ([33, p. 43])

$$(1-x)^{-a} \, {}_2F_1\left(a, c-b; c|\frac{x}{x-1}\right) = {}_2F_1(a, b; c|x)$$

by $(1-x)^{-A}$ and equating coefficients of x^n.

Acknowledgements Richard Askey's support and encouragement was crucial throughout our fledgling efforts to learn and apply the beautiful subject that he built. He took us in, patiently fielded 'stupid questions,' introduced us to his community, and published our papers. We thank Krishnaswami Alladi, Howard Cohl, and Dennis Stanton for their help. We also thank the anonymous referees for their useful comments.

References

1. Aldous, D., Fill, J.: Reversible Markov chains and random walks on graphs. https://www.stat.berkeley.edu/~aldous/RWG/book.pdf (2002)
2. Andersen, H.C., Diaconis, P.: Hit and run as a unifying device. J. Soc. Fr. Stat. Rev. Stat. Appl. **148**(4), 5–28 (2007)
3. Anderson, W.J.: Continuous-Time Markov Chains. Springer Series in Statistics: Probability and Its Applications. An Applications-Oriented Approach. Springer, New York (1991)

4. Arratia, R., Barbour, A.D., Tavaré, S.: Logarithmic Combinatorial Structures: A Probabilistic Approach. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich (2003)
5. Bailey, W.N.: Generalized Hypergeometric Series. Cambridge Tracts in Mathematics and Mathematical Physics. Stechert-Hafner Inc., New York (1964)
6. Borodin, A., Olshanski, G.: Representations of the Infinite Symmetric Group, Cambridge Studies in Advanced Mathematics, vol. 160. Cambridge University Press, Cambridge (2017)
7. Bryc, W., Wesołowski, J.: Askey-Wilson polynomials, quadratic harnesses and martingales. Ann. Probab. **38**(3), 1221–1262 (2010)
8. Cannings, C.: The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. Adv. Appl. Probab. **6**, 260–290 (1974)
9. Chen, W.K.: Mixing times for Burnside processes. Master's thesis, National Chiao Tung University (2006)
10. Chihara, T.S.: An Introduction to Orthogonal Polynomials. Mathematics and Its Applications, vol. 13. Gordon and Breach Science Publishers, New York (1978)
11. Clausen, T.: Ueber die Fälle, wenn die Reihe von der Form...ein Quadrat von der Form...hat. J. Reine Angew. Math. **3**, 89–91 (1828)
12. Crane, H.: The ubiquitous Ewens sampling formula. Stat. Sci. **31**(1), 1–19 (2016)
13. Diaconis, P.: Group representations in probability and statistics, Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 11. Institute of Mathematical Statistics, Hayward (1988)
14. Diaconis, P.: Analysis of a Bose-Einstein Markov chain. Ann. Inst. H. Poincaré Probab. Statist. **41**(3), 409–418 (2005)
15. Diaconis, P., Griffiths, R.: An introduction to multivariate Krawtchouk polynomials and their applications. J. Stat. Plan. Inference **154**, 39–53 (2014)
16. Diaconis, P., Kemperman, J.: Some new tools for Dirichlet priors. In: Bayesian Statistics, 5 (Alicante, 1994), pp. 97–106. Oxford University Press, New York (1996)
17. Diaconis, P., Ram, A.: A probabilistic interpretation of the Macdonald polynomials. Ann. Probab. **40**(5), 1861–1896 (2012)
18. Diaconis, P., Shahshahani, M.: Time to reach stationarity in the Bernoulli–Laplace diffusion model. SIAM J. Math. Anal. **18**(1), 208–218 (1987)
19. Diaconis, P., Shahshahani, M.: On the eigenvalues of random matrices. J. Appl. Probab. **31A**, 49–62 (1994). Studies in applied probability
20. Diaconis, P., Khare, K., Saloff-Coste, L.: Gibbs sampling, exponential families and orthogonal polynomials. Stat. Sci. **23**(2), 151–178 (2008). With comments and a rejoinder by the authors
21. Dunkl, C.F., Xu, Y.: Orthogonal Polynomials of Several Variables, Encyclopedia of Mathematics and Its Applications, vol. 155, 2nd edn. Cambridge University Press, Cambridge (2014)
22. Goldberg, L.A.: Automating Pólya theory: the computational complexity of the cycle index polynomial. Inf. Comput. **105**(2), 268–288 (1993)
23. Goldberg, L.A.: Computation in permutation groups: counting and randomly sampling orbits. In: Surveys in Combinatorics, 2001 (Sussex), London Mathematical Society Lecture Note Series, vol. 288, pp. 109–143. Cambridge University Press, Cambridge (2001)
24. Gordon, L.: A stochastic approach to the gamma function. Am. Math. Mon. **101**(9), 858–865 (1994)
25. Holtzen, S., Millstein, T., Van den Broeck, G.: Generating and sampling orbits for lifted probabilistic inference. In: Uncertainty in Artificial Intelligence, pp. 985–994. PMLR (2020)
26. Ismail, M.E.H.: Classical and quantum orthogonal polynomials in one variable, Encyclopedia of Mathematics and Its Applications, vol. 98. Cambridge University Press, Cambridge (2005). With two chapters by Walter Van Assche, With a foreword by Richard A. Askey
27. Jerrum, M.: Computational Pólya theory. In: Surveys in Combinatorics, 1995 (Stirling), London Mathematical Society Lecture Note Series, vol. 218, pp. 103–118. Cambridge University Press, Cambridge (1995)
28. Jerrum, M.: Uniform sampling modulo a group of symmetries using Markov chain simulation. In: Expanding Graphs (Princeton, NJ, 1992), DIMACS Series Discrete Mathematical Theoretical Computing Science, vol. 10, pp. 37–47. American Mathematical Society, Providence (1993)
29. Karlin, S., McGregor, J.L.: The Hahn polynomials, formulas and an application. Scripta Math. **26**, 33–46 (1961)
30. Kemeny, J.G., Snell, J.L.: Finite Markov Chains. Springer, New York (1976). Reprinting of the 1960 original, Undergraduate Texts in Mathematics

31. Khare, K., Zhou, H.: Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. Ann. Appl. Probab. **19**(2), 737–777 (2009)
32. Kingman, J.F.C.: Random discrete distribution. J. R. Stat. Soc. Ser. B. Stat. Methodol. **37**, 1–22 (1975)
33. Koepf, W.: Hypergeometric Summation, 2nd edn. Universitext. Springer, London (2014). An algorithmic approach to summation and special function identities
34. Letac, G.: Donkey walk and Dirichlet distributions. Stat. Probab. Lett. **57**(1), 17–22 (2002)
35. Levin, D.A., Peres, Y.: Markov Chains and Mixing Times. American Mathematical Society, Providence (2017). With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson
36. Milla, L.: A detailed proof of the Chudnovsky formula with means of basic complex analysis–Ein ausführlicher Beweis der Chudnovsky-Formel mit elementarer Funktionentheorie. arXiv preprint arXiv:1809.00533 (2018)
37. Pólya, G., Read, R.C.: Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds. Springer, New York (1987). Pólya's contribution translated from the German by Dorothee Aeppli
38. Saloff-Coste, L.: Lectures on finite Markov chains. In: Lectures on Probability Theory and Statistics (Saint-Flour, 1996), Lecture Notes in Mathematics, vol. 1665, pp. 301–413. Springer, Berlin (1997)
39. Sethuraman, J.: A constructive definition of Dirichlet priors. Stat. Sinica **4**(2), 639–650 (1994)
40. Shepp, L.A., Lloyd, S.P.: Ordered cycle lengths in a random permutation. Trans. Am. Math. Soc. **121**, 340–357 (1966)
41. Wilcox, R.R.: A review of the beta-binomial model and its extensions. J. Educ. Stat. **6**(1), 3–32 (1981)
42. Xu, Y.: Tight frame with Hahn and Krawtchouk polynomials of several variables. SIGMA symmetry integrability. Geom. Methods Appl. **10**, 19 (2014). Paper 019
43. Zhong, C.: A Ewens deformation of a Bose-Einstein Markov chain (in prep)
44. Zhong, C.: PhD thesis, Stanford University (in progress)