Random doubly stochastic tridiagonal matrices

Persi Diaconis
Departments of Statistics and Mathematics
Stanford University
450 Serra Mall
Stanford, CA 94305
diaconis@math.stanford.edu

Philip Matchett Wood
Department of Mathematics
University of Wisconsin
480 Lincoln Dr.
Madison, WI 53706-1388
pmwood@math.wisc.edu

December 7, 2011

Abstract

Let $\mathcal{T}_n$ be the compact convex set of tridiagonal doubly stochastic matrices. These arise naturally in probability problems as birth and death chains with a uniform stationary distribution. We study 'typical' matrices $T \in \mathcal{T}_n$ chosen uniformly at random in the set $\mathcal{T}_n$. A simple algorithm is presented to allow direct sampling from the uniform distribution on $\mathcal{T}_n$. Using this algorithm, the elements above the diagonal in T are shown to form a Markov chain. For large n, the limiting Markov chain is reversible and explicitly diagonalizable with transformed Jacobi polynomials as eigenfunctions. These results are used to study the limiting behavior of such typical birth and death chains, including their eigenvalues and mixing times. The results on a uniform random tridiagonal doubly stochastic matrices are related to the distribution of alternating permutations chosen uniformly at random.

Keywords: Markov chain, birth and death chain, cutoff phenomenon, random matrix

1 Introduction

Let $\mathcal{T}_n$ be the set of $(n+1)\times(n+1)$ tridiagonal doubly stochastic matrices, each element of which has the form:

$$
\begin{pmatrix}
1-c_1 & c_1 & & & & & 0 \\
c_1 & 1-c_1-c_2 & c_2 & & & & \\
 & c_2 & 1-c_2-c_3 & & c_3 & & \\
 & & \ddots & & \ddots & & \ddots \\
 & & & c_{n-1} & & 1-c_{n-1}-c_n & c_n \\
0 & & & & & c_n & 1-c_n
\end{pmatrix}, \quad (1.1)
$$

where all entries not on the main diagonal, superdiagonal, or subdiagonal are zero. Such matrices are completely determined by the numbers $c_1, c_2, \ldots, c_n$ above the diagonal, and so we may view $\mathcal{T}_n$ as a subset of $\mathbb{R}^n$. As a polytope, $\mathcal{T}_n$ has interesting combinatorial properties. For example, the number of extreme points of $\mathcal{T}_n$ is F_{n+1}, the $(n + 1)$-st Fibonacci number (where $F_0 = F_1 = 1$ and $F_i = F_{i-1} + F_{i-2}$). Clearly $\mathcal{T}_n$ is a compact convex subset of $\mathbb{R}^n$, and using Lebesgue measure in $\mathbb{R}^n$ the volume of $\mathcal{T}_n$ is $E_n/n!$, where E_n is the number of alternating (up/down) permutations in the symmetric group,

namely, those permutations $\sigma \in \mathfrak{S}_n$ such that $\sigma(1) < \sigma(2) > \sigma(3) < \sigma(4) > \cdots$. Using these properties, we give a simple direct way to sample uniform random elements of $\mathcal{T}_n$. These results are presented in Section 2.

Section 3 presents experimental results on the distribution of eigenvalues and mixing times of the associated birth and death chains. These results show that typical elements of $\mathcal{T}_n$ mix in order $n^2 \log n$ steps and do not have a 'cutoff' in their approach to stationarity (by convention we use log to denote the natural logarithm $\log_e$).

The question of proving whether a random element of $\mathcal{T}_n$ exhibits a cutoff is at the heart of this paper. We will discuss our approach below, demonstrating that $c_1, c_2, \ldots, c_n$ form a Markov chain and proving new results that characterize the limiting chain as $n \to \infty$. The limiting-chain results, in turn, provide an important tool for analyzing both the mixing time and the eigenvalue gap, leading to an answer to the cutoff question, discussed further at the end of this introduction.

In Section 4, the joint distribution of $c_1, c_2, \ldots, c_n$ is shown to be a Markov chain with a simple large n limit. Section 6 studies the limiting chain as $n \to \infty$, showing the following:

- $\Pr(c_1 \leq y) = \sin(\frac{\pi}{2} y)$ for $0 \leq y \leq 1$.

- $\Pr(c_i \leq y | c_{i-1} = x) = \sin(\frac{\pi}{2} \min\{y, 1 - x\}) / \sin(\frac{\pi}{2}(1 - x))$, for $0 \leq y \leq 1$ and $0 \leq x \leq 1$.

- The Markov chain c_i is reversible with stationary density $\pi(y) = 2\cos^2(\frac{\pi}{2} y)$ for $0 \leq y \leq 1$.

- The eigenvalues are $1, -\frac{1}{3}, \frac{1}{5}, -\frac{1}{7}, \ldots, \frac{(-1)^m}{2m+1}, \ldots$, and the eigenfunctions are transformed Jacobi polynomials.

- The total variation distance is bounded by $\|\mathcal{L}(c_\ell) - \pi\|_{\mathrm{TV}} \leq \sum_{i=1}^{\infty} i \left(\frac{1}{2i+1}\right)^{2\ell}$.

These results are used to study the distribution of eigenvalues and mixing times in Section 7, where it is proved that, for the limiting distribution, the spectral gap is of (stochastic) order $1/(n^2 \log n)$ and the mixing time is at most of order $n^2 \log n$. In Section 5, it is shown that a similar Markov chain governs the entries of a randomly chosen length n alternating permutation in the limit as $n \to \infty$. In particular, we prove in Theorem 5.1 that for any fixed positive integer ℓ, the joint distribution of the first ℓ entries of a randomly chosen alternating permutation is the same as the joint distribution of the first ℓ superdiagonal entries of a randomly chosen tridiagonal doubly stochastic matrix in the large n limit.

Our study of the matrices in this paper arose from the study of the *cutoff phenomena* in convergence of Markov chains to their stationary distributions. Let $d(\bullet, \bullet)$ denote a distance between two probability measures such as total variation. Briefly, a sequence $K_n(x, y)$ of Markov chains on finite state spaces $\mathcal{X}_n$ with stationary distribution π_n shows a cutoff at l_n if for every $\epsilon > 0$,

$$d\left(K_n^{l_n(1+\epsilon)}, \pi_n\right) \to 0 \qquad \text{and} \qquad d\left(K_n^{l_n(1-\epsilon)}, \pi_n\right) \to 1, \qquad (1.2)$$

where the chain K_n is started at state x_n.

As an example, the random walk on the hypercube C_2^n which changes a randomly chosen coordinate (or holds) with probability $1/(n+1)$ has a cutoff at $\frac{1}{4} n \log n$ [25]. The

random transposition chain on the symmetric group S_n has a cutoff at $\frac{1}{2}n \log n$ [24] and the Gilbert–Shannon–Reeds riffle shuffling chain has a cutoff at $\frac{3}{2} \log_2 n$ [7]. A survey of many examples is in [20].

The cutoff phenomena was named and studied by Aldous and Diaconis [1]. The fact that it was discovered very early in the quantitative study of rates of convergence suggests that it is endemic. Do most Markov chains show a cutoff? It took a while to find chains without a cutoff; simple random walk on a path of length n and walks on finite parts of lattices in fixed dimension do not show cutoffs. These questions motivated the present study.

Yuval Peres [22, Conjecture 1 on page 2] noticed that for all of the available examples two simple features of the Markov chain determine if there is a cutoff. The spectral gap, gap_n, is the difference between one and the (absolute) second-largest eigenvalue. The *mixing time* is the smallest number of steps r_n such that the distance to stationarity is smaller than $1/e$. Peres observed that, in many examples, there is a cutoff if and only if $\mathrm{gap}_n \times r_n \to \infty$. For example, the walk on the hypercube has $\mathrm{gap}_n = 2/(n+1)$ and $r_n = n \log n$ so $\mathrm{gap}_n \times r_n$ tends to infinity. For riffle shuffling, $\mathrm{gap}_n = \frac{1}{2}$ while $r_n = \log n$. For random walk on a path, $\mathrm{gap}_n = c/n^2$ while $r_n = c'n^2$. Isolated counter-examples have been found by Aldous and Pak but the finding largely holds. Furthermore, [27, Lemma 2.1] proves for any reversible Markov chain that if there is a cutoff, then $\mathrm{gap}_n \times r_n \to \infty$.

Simple random walk on a path of length n, where the probability of moving left or right is $1/2$ and the probability of holding at an endpoint is also $1/2$, is a notable example. Combining [27, Lemma 2.1] and the bounds in the previous paragraph gives a proof that simple random walk does not have a cutoff. It is natural to ask: should the fact that simple random walk does not have a cutoff suggest that a general birth-and-death process with uniform stationary distribution also has no cutoff, or is simple random walk a special case? On one hand, simple random walk seems very natural; on the other hand, simple random walk is also extreme in certain ways. For example Boyde, Diaconis, Sun, and Xiao [9] show that simple random walk has the largest eigenvalue gap of birth-and-death processes with uniform stationary distribution. Furthermore, Jonas Kahn and James Fill [31] have recently shown that for any $t \geq 0$, simple random walk is closer (measured in total variation distance or in L^2 distance) to uniform after t steps than any other birth-and-death chain with uniform stationary distribution if both chains are started from an endpoint.

To address the question of whether the lack of a cutoff for simple random walk is part of a general phenomenon, we return to Peres's observation that cutoffs typically are present when $\mathrm{gap}_n \times r_n \to \infty$. In fact, Diaconis and Saloff-Coste [23] proved Peres observation is true for all birth-and-death chains. In their version, the chains started from one endpoint of their interval of definition and the distance used was separation; the analysis was carried out in continuous time. Ding, Lubetzky and Peres [27] proved that the observation held without these caveats as well (in discrete time, from any start, and in total variation) so long as the chain is lazy, namely the probability of holding at any given point is at least $\delta > 0$, where δ is a constant. Further developments on birth/death cutoffs are seen in Barrera, Bertoncini, and Fernández [4] and Diehl [26]. Another step forward: Chen and Saloff-Coste [12, 13, 14] have proved that the Peres observation is true in l_p distances, $p > 1$, for *any* sequence of reversible Markov chains.

All of this work points to the question, "Well, which is it?" Does the cutoff phenomena usually hold or not? The Peres observation reduces this to a study of eigenvalues

and mixing times, but it does not help with the details.

Since so much is known about birth-and-death chains, this seems like a good place to start. What do the eigenvalues of a typical birth-and-death chain look like? To focus further, we fixed the stationary distribution as uniform and thus ask, What is the distribution of the eigenvalues and the mixing time of a random, symmetric, tridiagonal, doubly stochastic matrix? Our results above show that most birth-and-death chains with uniform stationary distributions mix in order $n^2 \log n$ steps and do not show a cutoff.

2   Polytope combinatorics and random generation

From Equation (1.1), it is clear that the polytope $\mathcal{T}_n$ is n dimensional and determined by

$$c_i \geq 0 \text{ and } c_i + c_{i+1} \leq 1, \text{ for all } 0 \leq i \leq n \quad (\text{we let } c_0 = c_{n+1} = 0). \qquad (2.1)$$

The extreme points are determined by setting c_i to be 0 or 1. Of course, Display (2.1) prevents two consecutive entries c_i from both being equal to 1. The binary sequences of length n with no two consecutive ones are in bijection with the Fibonacci numbers, for example $|\{000, 001, 010, 100, 101\}| = 5$. Thus, $\mathcal{T}_n$ has F_{n+1} extreme points, where $F_0 = F_1 = 1$ and $F_i = F_{i-1} + F_{i-2}$. Explicitly, the extreme points are $n+1$ by $n+1$ tridiagonal permutation matrices. See [17] for more on these Fibonacci permutations, including a study of the graph formed by the vertices and edges of the polytope $\mathcal{T}_n$. Chebikin and Ehrenborg [10] give a nice but somewhat complicated expression for the generating function for the f-vector of $\mathcal{T}_n$. See [16] for a combinatorial description of the faces of the polytope $\mathcal{T}_n$, including counting the number of vertices on each face, and see [15] for enumeration of the vertices, edges, and cells in terms of formulas using Fibonacci numbers.

We note here that random tridiagonal matrices are also studied from another viewpoint, as tridiagonalizations of the standard Gaussian orthogonal ensembles (GOE) or Gaussian unitary ensembles (GUE). First studied by Trotter [51] in 1984, such matrix ensembles gained considerable interest in 2002 with the introduction by Dumitriu and Edelman [28] of a continuous family of tridiagonal matrix ensembles parametrized by β, where $\beta = 1$ corresponds to GOE and $\beta = 2$ corresponds to GUE; for example, see [32, 52].

The volume of $\mathcal{T}_n$ was determined in [46] (see also [47]) as

$$\mathrm{vol}(\mathcal{T}_n) = \frac{E_n}{n!}, \qquad (2.2)$$

where E_n is the number of alternating (up/down) permutations on n letters (which is also equal to the number of reverse alternating permutations on n letters). Recall that a permutation σ is *alternating* if $\sigma(1) < \sigma(2) > \sigma(3) < \sigma(4) > \cdots$; and σ is *reverse alternating* if the reverse inequalities all hold. (Note that some papers use a different convention, calling down/up permutations alternating and up/down permutations reverse alternating.) For example, $E_4 = 5$ corresponds to the permutations 3412, 2413, 1423, 2314, 1324. A classical result of Desiré André in 1879 [3] gives an elegant way to compute E_n.

Theorem 2.1. [3]
$$\sum_{n \geq 0} E_n \frac{x^n}{n!} = \sec(x) + \tan(x).$$

The survey of alternating permutations by Richard Stanley [49] includes an elementary proof of Theorem 2.1, and has other connections with work in the current paper. For example [49] discusses enumeration of alternating permutations and Euler numbers and refinements such as the Entringer numbers, topics that we will return to in Subsection 5.1. Stanley's survey [49] also gives connections between the theory of alternating permutations and Euler numbers and the more general theory of permutations with a given descent set.

Alternating permutations may be identified with a special case of standard Young tableaux taking a particular shape. In [5], Baryshnikov and Romik derive combinatorial identities that generalize André's identity (Theorem 2.1) for a wider class of standard Young tableaux. In particular, an alternating permutation corresponds to something called a width-2 diagonal strip standard Young tableau, and [5] considers width-m diagonal strip standard Young tableau, for any $m \geq 2$. One tool in [5] is an extension of a transfer operator approach developed by Elkies [30] for alternating permutations (see [5, Section 2]).

In [47], Richard Stanley gives a decomposition of the polytope $\mathcal{T}_n$ into equal volume unit simplices, indexed by the set of alternating permutations. This gives a nice way to prove Equation (2.2), and we will use the decomposition to give a simple algorithm to choose an element of $\mathcal{T}_n$ uniformly at random.

2.1 Algorithm for randomly generating tridiagonal doubly stochastic matrices, with respect to Lebesgue measure

1. Choose an alternating permutation σ uniformly at random (see below).

2. Choose n points uniformly in $(0, 1)$ and order them from smallest to largest, calling them $0 < x_1 < x_2 < \cdots < x_n < 1$.

3. Define the c_i as follows:
$$c_i := \begin{cases} x_{\sigma_i} & \text{if } i \text{ is odd, and} \\ 1 - x_{\sigma_i} & \text{if } i \text{ is even.} \end{cases}$$

This step uses the map given by Richard Stanley in [47, Theorem 2.3].

The point of the above algorithm is that it generates an element of $\mathcal{T}_n$ uniformly with respect to Lebesgue measure, which we formalize below.

Proposition 2.2. *Let $(x_1, x_2, \ldots, x_n)$ be an element of $\mathcal{T}_n$ generated by the above algorithm, and let B be a ball in $\mathbb{R}^n$. Then*
$$\Pr((x_1, x_2, \ldots, x_n) \in B) = \frac{\text{Vol}(B \cap \mathcal{T}_n)}{\text{Vol}(\mathcal{T}_n)},$$
where the volume $\text{Vol}(\bullet)$ is Lebesgue measure in $\mathbb{R}^n$.

Proof. In [47, Theorem 2.3], Richard Stanley shows that the polytope $\mathcal{T}_n = \{(x_1, x_2, \ldots, x_n) : 0 \le x_i \le 1$ for $1 \le i \le n$ and $x_i + x_{i+1} \le 1$, for $1 \le i \le n-1\}$ is affinely equivalent to the polytope $\mathcal{P}_n := \{(y_1, y_2, \ldots, y_n) : 0 \le y_i \le 1$ for $1 \le i \le n$ and $y_1 \le y_2 \ge y_3 \le y_4 \ge \cdots y_n\}$. In particular, the map $\phi(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ defined by

$$y_i = \begin{cases} x_i & \text{if } i \text{ is odd} \\ 1 - x_i & \text{if } i \text{ is even.} \end{cases}$$

is a volume-preserving bijection between the $\mathcal{T}_n$ and $\mathcal{P}_n$. Note that it is shown in [36] that $\mathcal{T}_n$ and $\mathcal{P}_n$ are in fact 0/1 equivalent, and hence also congruent.

By step 3 of the above algorithm, $\Pr((x_1, x_2, \ldots, x_n) \in B) = \Pr(\phi(y_1, y_2, \ldots, y_n) \in B)$, where $(y_1, y_2, \ldots, y_n)$ is a uniformly chosen element of $\mathcal{P}_n$ with respect to Lebesgue measure by steps 1 and 2. Furthermore, $\Pr(\phi(y_1, y_2, \ldots, y_n) \in B) = \Pr((y_1, y_2, \ldots, y_n) \in \phi(B))$, since ϕ is volume-preserving and an involution. Finally $\Pr((y_1, y_2, \ldots, y_n) \in \phi(B)) = \mathrm{Vol}(\mathcal{P}_n \cap \phi(B))/\mathrm{Vol}(\mathcal{P}_n) = \mathrm{Vol}(\mathcal{T}_n \cap B)/\mathrm{Vol}(\mathcal{T}_n)$, since $(y_1, y_2, \ldots, y_n)$ was chosen uniformly in $\mathcal{P}_n$ and ϕ is volume-preserving. $\qquad\square$

There is a more complicated map given in [47] that can be used in place of the final step in the above algorithm. Namely, one can define the c_i as follows:

$$c_i := \begin{cases} x_{\sigma_i} & \text{if } i \text{ is odd, and} \\ \min\left\{x_{\sigma_i} - x_{\sigma_{i-1}}, x_{\sigma_i} - x_{\sigma_{i+1}}\right\} & \text{if } i \text{ is even.} \end{cases}$$

This more complicated map is useful in [47] for dealing with polytopes derived from arbitrary posets (the above is specialized to the case where the only relations are $a_1 < a_2 > a_3 < a_4 > \cdots$, which is the poset associated to an alternating permutation). However, in the case of posets with length at most 1 (which includes the alternating poset), the simpler map given in the algorithm is sufficient.

Example: Say that $n = 7$ and we have the alternating permutation 4627153. Let the uniformly chosen points in the interval $[0, 1]$ be $0 \le x_1 \le x_2 \le \cdots \le x_n \le 1$. To help remember which of the x_i cover other elements (which is determined by the alternating permutation), we write:

$$x_4 < x_6 > x_2 < x_7 > x_1 < x_5 > x_3.$$

Finally, we define the c_i as follows:

$$c_1 = x_4$$
$$c_2 = 1 - x_6$$
$$c_3 = x_2$$
$$c_4 = 1 - x_7$$
$$c_5 = x_1$$
$$c_6 = 1 - x_5$$
$$c_7 = x_3$$

Choosing the alternating permutation. Richard Stanley [48] has given the following procedure for choosing an alternating permutation σ uniformly at random based on the recurrence $E_n = \sum_{k \text{ even}} \binom{n-1}{k-1} E_{k-1} E_{n-k}$, where E_n is the number of alternating

permutations (which also equals the number of reverse alternating permutations). Note that Nijenhuis and Wilf [43] discuss in general how any recursive formula can be turned into a recursive random sampling method, and their procedure matches the one we use here.

1. Choose even k between 1 and n with probability $p_k := \binom{n-1}{k-1} E_{k-1} E_{n-k}/E_n$. Insert n into position k.

2. Choose a $k-1$ element subset S of $\{1, 2, \ldots, n-1\}$.

3. Choose an alternating permutation U of S (recursively).

4. Choose a reverse alternating permutation V of $\{1, 2, \ldots, n-1\} \setminus S$ (by a similar recursive algorithm).

5. Let $\sigma = UnV$.

The fact that $\sum_{n \geq 0} E_n x^n/n! = \sec(x) + \tan(x)$ enables us to compute the numbers E_n quickly using Taylor series. A different way of generating random alternating permutations we have found efficient is to run a Markov chain by making random transpositions (accepting to move only if the resulting permutation is still alternating). It is straightforward to show that this walk is connected, and experiments indicate that it mixes rapidly. In addition to being efficient in practice, this second method also has the advantage that one does not need to compute the numbers E_n.

2.2 Fast, approximately uniform random generation

One very fast way to generate a random tridiagonal doubly stochastic matrix with respect to Lebesgue measure, or at least a very close approximation of Lebesgue measure, is to use Gibbs sampling. One Gibbs sampling algorithm that we use extensively in Section 3 may be described using the following subroutine that operates on a matrix with superdiagonal entries $(c_1, c_2, \ldots, c_n)$:

Subroutine(i):
1. Choose x uniformly in the interval $[0, 1]$.
2. If $c_{i-1} + x \leq 1$ and $c_{i+1} + x \leq 1$, then set $c_i = x$. By convention $c_0 = c_{n+1} = 0$.
3. Adjust entries (i, i), $(i+1, i+1)$, and $(i+1, i)$ in the matrix so that the new matrix with i-th superdiagonal entry x is doubly stochastic and tridiagonal.

To sample in $\mathcal{T}_n$ close to uniformly, one can successively apply Subroutine(i) many times so that each superdiagonal entry c_i has had a number of chances to be updated. In particular, the Gibbs sampling algorithm we use in Section 3 starts with an $n+1$ by $n+1$ identity matrix (so $c_i = 0$ for all i) and does the following simple procedure $10 \log n$ times: apply Subroutine(i) for $i = 1, 2, \ldots, n$ in order. Thus in total, each superdiagonal entry has $10 \log n$ opportunities to be changed. The resulting algorithm is a very fast way to generate a tridiagonal doubly stochastic matrix, and empirically the resulting distribution on tridiagonal doubly stochastic matrices is very close to Lebesgue measure.

3 Experiments and conjectures

This section collects experimental results using Gibbs sampling to produce a random element of $\mathcal{T}_{n-1}$ (so there are $n-1$ superdiagonal entries, and the matrices are each n by
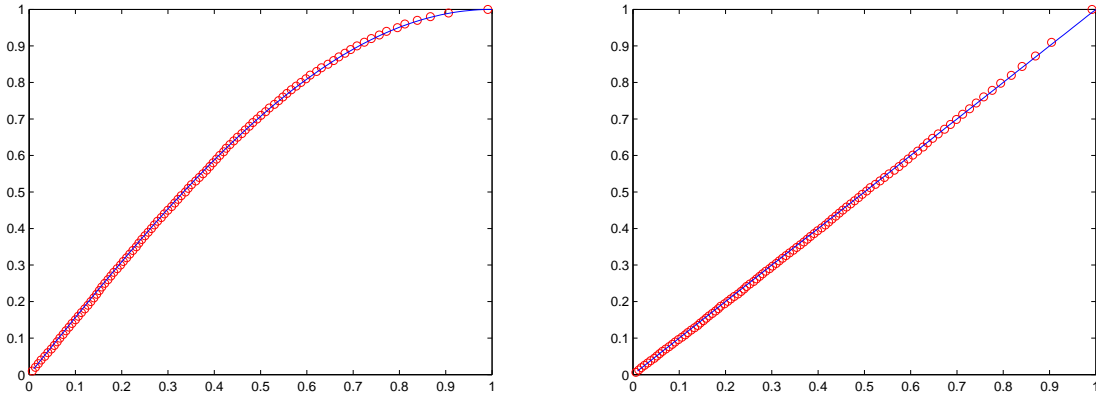
Figure 1: Above left: the distribution function for the first superdiagonal entry of a 10 by 10 tridiagonal doubly stochastic matrix. The circles represent the 100-quantiles from data of 10,000 random trials using Gibbs sampling, and the curve is $\sin(\pi x/2)$. Above right: the corresponding Q-Q plot, which shows that the fit is good.

n). We have compared Gibbs sampling to the (much slower) exact sampling algorithm in many examples and see no difference.

Figures 1, 2, and 3 give experimental verification for Corollary 4.4, which proves that the distribution in the limit as $n \to \infty$ of first superdiagonal entry is $\sin(x\pi/2)$ and also describes the marginal distribution for the k-th entry given the $(k-1)$-st entry. Here, with $n = 50$ or even $n = 10$, the experimental distributions are extremely close to the limiting distribution as $n \to \infty$.

Figure 4 demonstrates that the distribution of the superdiagonal entries rapidly become close to the stationary distribution function $\sin(x\pi)/\pi + x$ as one moves away from the ends of the superdiagonal. In particular Figure 4 shows that, while the first superdiagonal entry has distribution $\sin(x\pi/2)$, the fourth superdiagonal already has a distribution that is almost indistinguishable from $\sin(x\pi)/\pi + x$, even for when $n = 9$. Data for larger n produces plots virtually identical to those in Figure 4.

In Figure 5, we experimentally compare the distribution of the limiting Markov chain formed by the superdiagonal entries in the limit as $n \to \infty$. Theorem 6.1 shows that the stationary distribution should have distribution function $\sin(\pi x)/\pi + x$, and the data shows that the average value of the superdiagonal entries away from the top and bottom of the matrix closely matches this distribution.

Figure 6 shows the growth rate of the eigenvalue gap gap_{n-1}, that is, the second smallest absolute difference between an eigenvalue and 1 (note that 1 is always an eigenvalue). The figures suggest that the growth rate of the random function gap_{n-1} satisfies

$$3.4 \le n^2 \log(n) \, \text{gap}_{n-1} \le 4.3,$$

with high probability for large n.

We can analogously study the mixing time r_n of a randomly selected tridiagonal doubly stochastic matrix. Figure 7 shows plots of r_{n-1}, each averaged over 100 trials, for values of n up to 2000. The plots suggest that

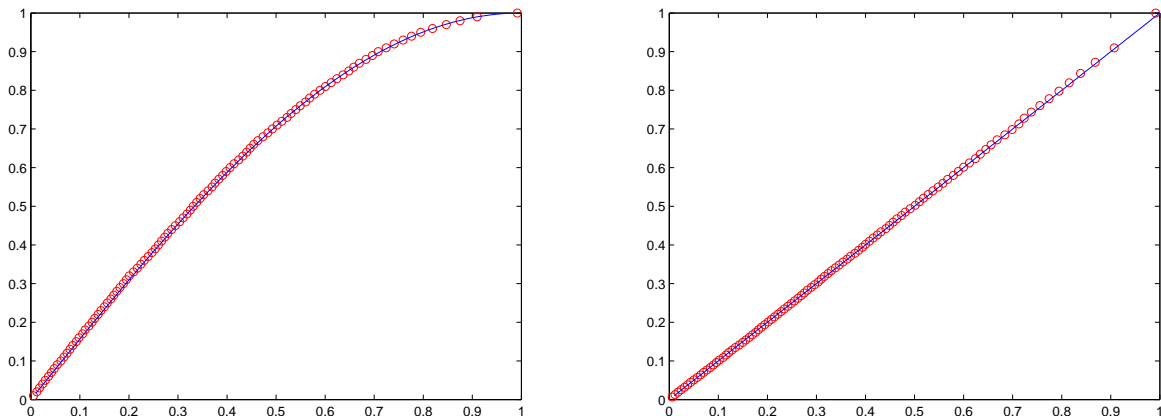$$\frac{2}{9} n^2 \log(n) \le r_{n-1} \le \frac{2}{5} n^2 \log(n)$$

8

Figure 2: Above left: the distribution function for the first superdiagonal entry of a 50 by 50 tridiagonal doubly stochastic matrix. The circles represent the 100-quantiles from data of $10,000$ random trials using Gibbs sampling, and the curve is $\sin(\pi x/2)$. Above right: the corresponding Q-Q plot, which shows that the fit is good.
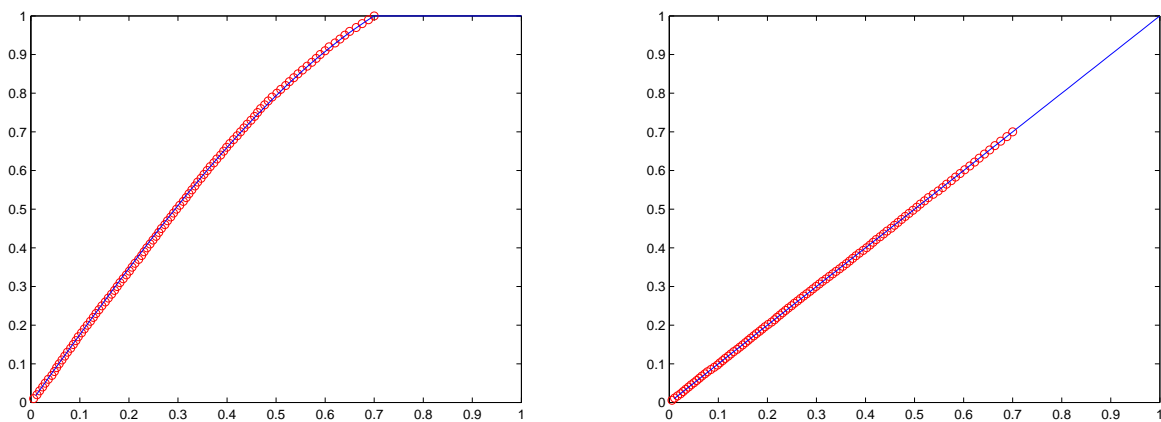

Figure 3: Above left: the distribution function for the seventh superdiagonal entry of a 50 by 50 tridiagonal doubly stochastic matrix, given that the sixth superdiagonal entry is 0.3. The circles represent the 100-quantiles from data of $10,000$ random trials using Gibbs sampling, and the curve is $\dfrac{\sin(\frac{\pi}{2}\min\{x, 0.7\})}{\sin(\pi(0.3)/2)}$. Above right: the corresponding Q-Q plot, which shows again that the fit is good.
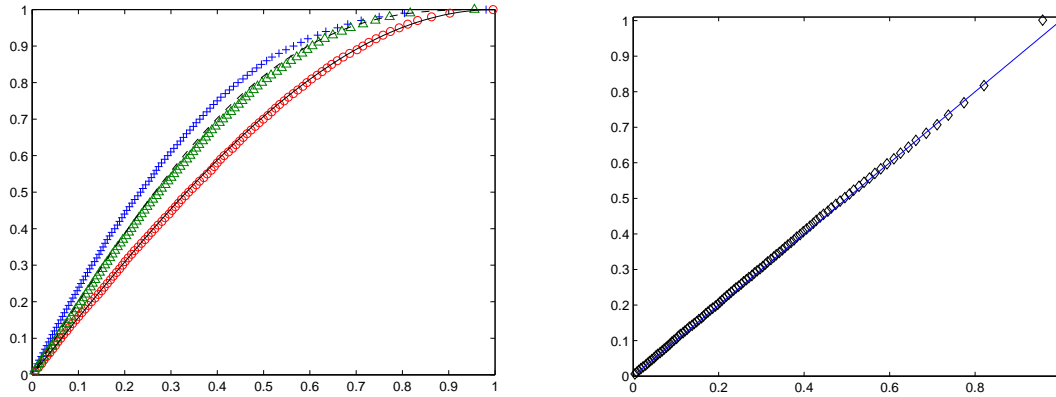
Figure 4: Above left: the distributions of the first, second, and third superdiagonal entries for an 10 by 10 tridiagonal doubly stochastic matrix, denoted by, respectively, circles, plus signs, and triangles. Notice that the distribution of the first closely matches the solid curve $\sin(x\pi/2)$, and that the third comes close to (though is slightly below) the dashed curve $\sin(x\pi)/\pi + x$, which is the stationary distribution. Above right: the Q-Q plot comparing the distribution of the fourth superdiagonal entry in a 10 by 10 tridiagonal doubly stochastic matrix to the stationary distribution $\sin(x\pi)/\pi + x$. The fit is remarkably good.
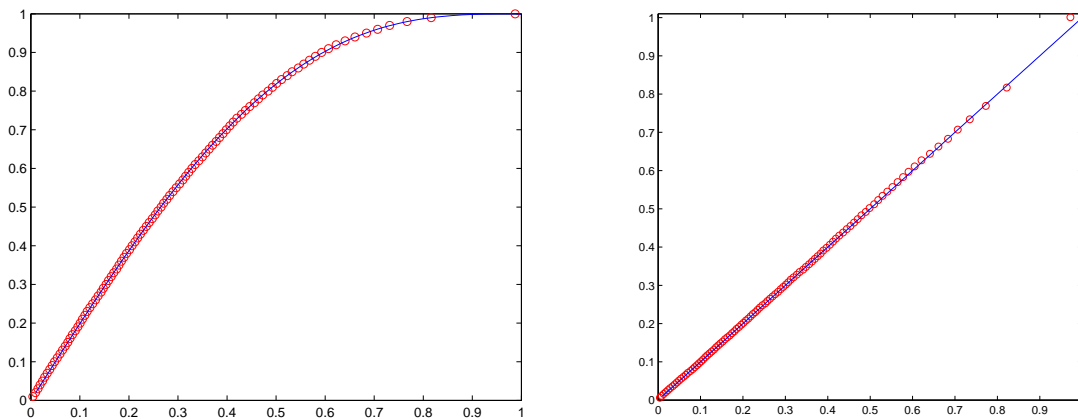

Figure 5: Above left: the curve is $\sin(\pi x)/\pi + x$, and the circles represent the 100-quantiles from the data from 100 trials using Gibbs sampling of all superdiagonal entries in rows 10 through 189 of a 200 by 200 tridiagonal doubly stochastic matrix. Above right: the corresponding Q-Q plot.
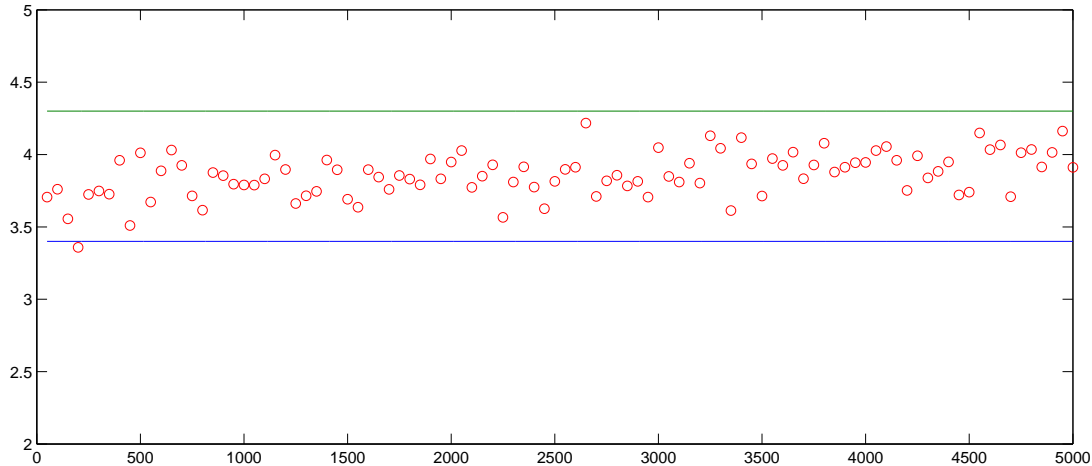
10

Figure 6: Above: an empirical plot of the function $n^2 \log(n) \operatorname{gap}_{n-1}$ using Gibbs sampling for $n = 50$ to $n = 5000$, with the results averaged over 100 trials. The plots suggest that $n^2 \log(n) \operatorname{gap}_{n-1}$ is bounded by a constant.

with high probability for large n. Taken together with the fact (see [27, Lemma 2.1] or also [23]) that there is a cutoff for a birth and death chain only if $\operatorname{gap}_{n-1} \times r_{n-1} \to \infty$, we see that data on the eigenvalue gap in Figure 6 and the data on the mixing time in Figure 7 suggest that, with high probability, a random element of $\mathcal{T}_{n-1}$ does not have a cutoff. In Section 7, we will prove, in fact, that with probability tending to 1, $\operatorname{gap}_{n-1} \times r_{n-1}$ is bounded as $n \to \infty$, thus proving that with high probability, a random element of $\mathcal{T}_{n-1}$ does not have a cutoff (see Theorems 7.1 and 7.2).

We have seen that the second largest (absolute) eigenvalue has an important effect on whether or not a birth and death chain has a cutoff, and one can consider the more general question of determining the distribution of the eigenvalues of a random element of $\mathcal{T}_{n-1}$. Figure 8 shows a histogram of the eigenvalues for $n = 100,000$. The pictured distribution seems empirically stable as n increases and does not seem to belong to one of the standard ensembles. It would be interesting to describe some of the persistent features of this distribution in the large n limit. Though it does not directly give any information on the limiting shape, note that using the tools in [33], it is possible to prove that, with probability 1, the histogram approaches some fixed, nonrandom shape as $n \to \infty$. Recently, Anderson and Zeitouni [2] used ideas from [39] to develop some tail estimates on the limiting eigenvalue distribution for a random element of $\mathcal{T}_{n-1}$. For example, [2] shows that for sufficiently small $\delta > 0$, the number of eigenvalues in the interval $[1 - \delta, 1]$ is at most $O(n/\log \delta)$ with high probability.

Another interesting question to consider is the behavior of the smallest superdiagonal entry of a random tridiagonal doubly stochastic matrix. Figure 9 provides some experimental evidence suggesting that the smallest superdiagonal entry may have roughly the distribution of the smallest of n independent uniform random samples from the interval $[0, 1/2]$, which would have distribution function $1 - (1 - 2x)^n$. However, the Q-Q plot shows that the match is not perfect when the smallest superdiagonal entry is in the larger part of its range. It would be interesting to describe the behavior of the
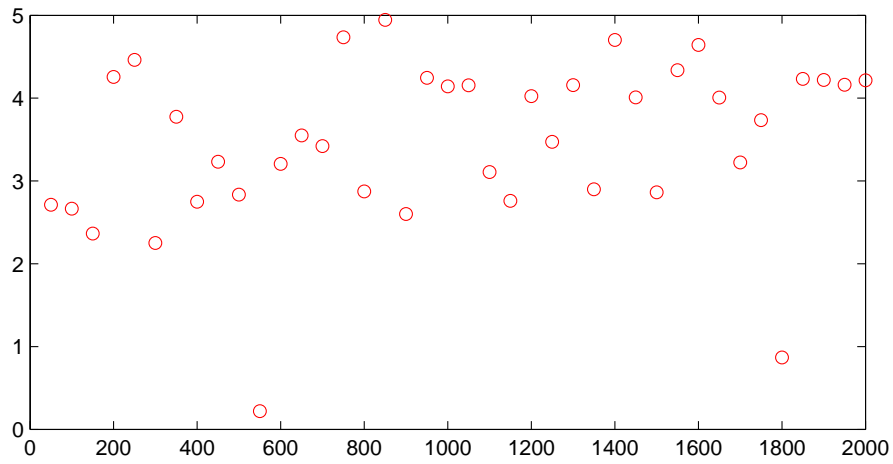
11

Figure 7: The random function r_n denotes the mixing time of a randomly chosen tridiagonal doubly stochastic matrix. Using Gibbs sampling, the plot above gives $n^2 \log(n)/r_{n-1}$ for values of n equal to multiples of 50 between 50 and 2000 averaged over 50 trials.
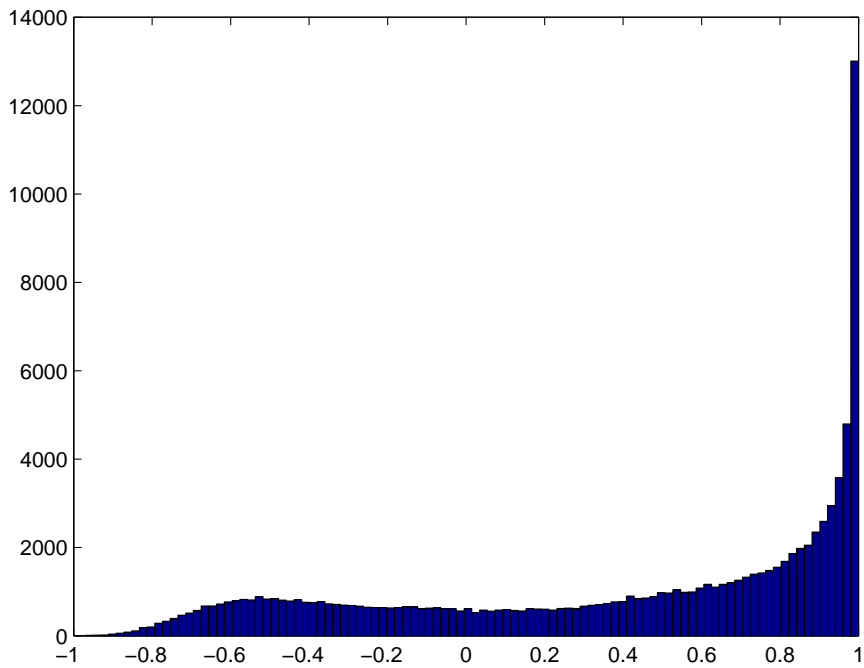

Figure 8: Above is a histogram of the eigenvalues from a single $100,000$ by $100,000$ randomly generated tridiagonal doubly stochastic matrices using Gibbs sampling.
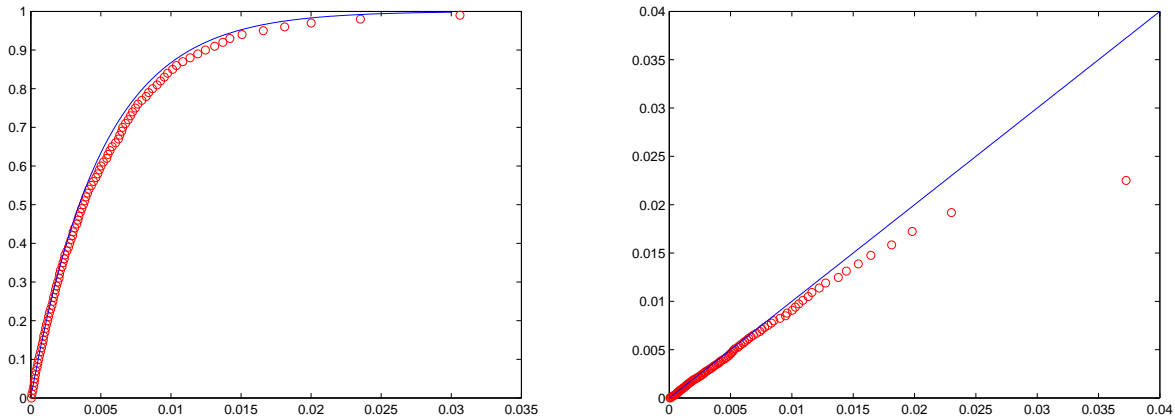
Figure 9: Above is are plots for the distribution of the smallest superdiagonal entry for 100 by 100 tridiagonal doubly stochastic matrices, where the data is taken from 1000 matrices generated with Gibbs sampling. On the left, the circles represent the 100-quantiles and the curve is a graph of $1 - (1 - 2x)^{99}$. On the right is the corresponding Q-Q plot, which shows that the fit becomes less good when the smallest superdiagonal entry is in the larger part of its range.

distribution of the smallest superdiagonal entry of a random element of $\mathcal{T}_n$.

Finally, it would also be interesting to determine the quantitative behavior of the smallest eigenvalue of a randomly chosen tridiagonal doubly stochastic matrix. In Figure 10, data is shown suggesting that the average smallest eigenvalue approaches a value less than -0.9. It would be interesting to determine whether this average approaches -1 as n goes to infinity.

4   Distribution of the superdiagonal

As noted in Section 2, the elements of an $n + 1$ by $n + 1$ tridiagonal doubly stochastic matrix are determined by the superdiagonal $c_1, c_2, \ldots, c_n$. For a uniformly chosen matrix, we determine the joint distribution of $\{c_i\}$. For both fixed n and in the large n limit, the c_i form a Markov chain. We compute the distribution of the $(1, 2)$ entry and the distribution of the $(i, i + 1)$ conditioned on the $(i - 1, i)$ entry. Section 6 studies the limiting Markov chain defined by letting n tend to infinity. We first state the results. Proofs are brought together at the end of this section.

Let $(c_1, c_2, \ldots, c_n)$ be the superdiagonal of an $n + 1$ by $n + 1$ tridiagonal doubly stochastic matrix chosen uniformly at random with respect to Lebesgue measure (for example, using the algorithm in Section 2). Write $c_i = c_i^{(n)}$ when it is useful to emphasize the dependence of c_i on n.

Theorem 4.1. *For any $1 \leq i \leq n - 1$, for any real constants $a_1, \ldots, a_i$ in the interval $[0, 1]$, and for any $0 \leq t \leq 1$,*

$$\Pr(c_{i+1} \leq t | c_1 = a_1, c_2 = a_2, \ldots, c_i = a_i) = \Pr(c_{i+1} \leq t | c_i = a_i). \qquad (4.1)$$
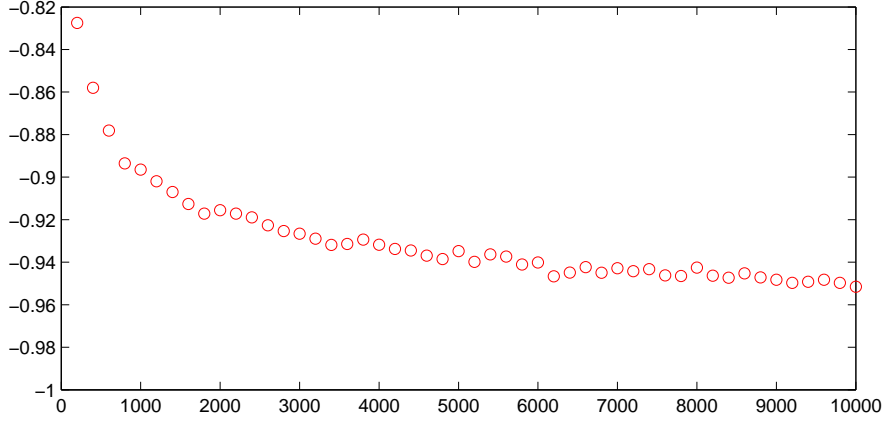
Figure 10: The data above was generated using Gibbs sampling to find the average value of the smallest eigenvalue out of 200 trials for n ranging over multiples of 200 between 200 and 10000.

Proof. The probabilities in Equation (4.1) can be computed via integration. In particular, define the function

$$f_{i,n}(x) := \int_{c_i=0}^{x} \int_{c_{i+1}=0}^{1-c_i} \int_{c_{i+2}=0}^{1-c_{i+1}} \int_{c_{i+3}=0}^{1-c_{i+2}} \cdots \int_{c_n=0}^{1-c_{n-1}} dc_n \, dc_{n-1} \cdots dc_i. \qquad (4.2)$$

The left-hand side of Equation (4.1) thus becomes $f_{i+1,n}(\min\{t, 1 - a_i\})/f_{i+1,n}(1 - a_i)$. The right-hand side of Equation (4.1) can be represented via

$$g_{i+1,n}(x, y) := \left(\int_{c_{i-1}=0}^{y} \int_{c_{i-2}=0}^{1-c_{i-1}} \cdots \int_{c_1=0}^{1-c_2} dc_1 \, dc_2 \ldots dc_{i-1} \right)$$
$$\times \left(\int_{c_{i+1}=0}^{x} \int_{c_{i+2}=0}^{1-c_{i+1}} \cdots \int_{c_n=0}^{1-c_{n-1}} dc_n \, dc_{n-1} \cdots dc_{i+1} \right)$$
$$= f_{i+1,n}(x) f_{1,i-1}(y).$$

The right-hand side of Equation (4.1) equals $g_{i+1,n}(\min\{t, 1 - a_i\}, 1 - a_i)/g_{i+1,n}(1 - a_i, 1 - a_i) = f_{i+1,n}(\min\{t, 1 - a_i\})/f_{i+1,n}(1 - a_i)$, thus proving Equation (4.1). $\qquad \square$

Remark 4.2. One interesting feature to note is that the distribution of c_i is the same as the distribution of c_{n-i+1} for each $1 \leq i \leq n$. This fact can be proven by demonstrating a volume preserving bijection between the following two n-dimensional polytopes:

$$\begin{aligned}
\mathcal{P}_i(t): \quad & 0 \leq c_1, c_2, \ldots, c_n \leq 1 \\
& c_j + c_{j+1} \leq 1 \qquad \text{for } 1 \leq j \leq n-1, \\
& c_i \leq t \\
\mathcal{P}_{n-i+1}(t): \quad & 0 \leq c_1, c_2, \ldots, c_n \leq 1 \\
& c_j + c_{j+1} \leq 1 \qquad \text{for } 1 \leq j \leq n-1, \\
& c_{n-i+1} \leq t.
\end{aligned}$$

14

One simple volume-preserving bijection is the map $\phi : c_j \mapsto c_{n-j+1}$ for all $1 \leq j \leq n$. It is clear that ϕ is a bijection, is volume preserving, and maps $\mathcal{P}_i$ to $\mathcal{P}_{n-i+1}$. Since the probability that c_i is at most t is exactly the volume of $\mathcal{P}_i$, and the probability that c_{n-i+1} is at most t is exactly the volume of $\mathcal{P}_{n-i+1}$, it is clear that c_i and c_{n-i+1} have the same distribution.

Probabilities from integration. Let $c_1^{(n)}, c_2^{(n)}, \ldots, c_n^{(n)}$ be the superdiagonal entries of an $n+1$ by $n+1$ tridiagonal doubly stochastic matrix chosen uniformly with respect to Lebesgue measure. From the definition of $f_{i,n}$ in Equation (4.2), it is clear that $f_{i,n}(x) = f_{1,n-i+1}(x)$. Furthermore,

$$\Pr(c_1^{(n)} \leq x) = \frac{f_{1,n}(x)}{f_{1,n}(1)}, \text{ and} \tag{4.3}$$

$$\Pr(c_{i+1}^{(n)} \leq x | c_i = a_i) = \frac{f_{1,n-i}(\min\{x, 1 - a_i\})}{f_{1,n-i}(1 - a_i)}, \text{ for } 1 \leq i \leq n - 1. \tag{4.4}$$

Thus, the distribution of $c_1^{(n)}$ and also the distribution of $c_{i+1}^{(n)}$ given $c_i^{(n)}$ may be computed from $f_{1,m}(x)$ for various values of m.

We will call $f_{1,n}(x)$ the *volume of $\mathcal{T}_n$ up to height x in the first dimension*. As we have seen above, $f_{1,m}(x)$ is useful in computing the marginal distributions of the entries of an element of $\mathcal{T}_n$ chosen uniformly at random. The main theorem for the current section is the following.

Theorem 4.3. *There is an exact polynomial formula for the volume of $\mathcal{T}_n$ up to height x in the first dimension, namely,*

$$f_{1,n}(x) = \frac{1}{n!} \left(\sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^k x^{2k+1} E_{n-(2k+1)} \binom{n}{2k+1} + (-1)^{n/2} x^n \delta_{n,\text{even}} \right), \tag{4.5}$$

where $\delta_{n,\text{even}}$ is 0 if n is odd and 1 if n is even. Here, as usual, E_n denotes the n-th Euler number, the number of alternating permutations on $[n]$.

Proof. Our original proof was an elementary but lengthy induction. We would like to thank Richard Stanley for pointing out the following elegant proof.

Stanley [49, page 13] proves the following generating function

$$\sum_{n \geq 0} f_{1,n}(x) t^n = \sec(t) \left(\cos((x-1)t) + \sin(xt) \right).$$

Expanding the right hand side as a Taylor series in x we have

$$\sum_{n \geq 0} f_{1,n}(x) t^n = \sum_{k \geq 0} (-1)^k \frac{(xt)^{2k}}{(2k)!} + (\sec(t) + \tan(t)) \sum_{k \geq 0} (-1)^k \frac{(xt)^{2k+1}}{(2k+1)!}.$$

Noting that $\sec(t) + \tan(t) = \sum_{\ell \geq 0} E_\ell \frac{t^\ell}{\ell!}$ (Theorem 2.1) and collecting terms on the right hand side by powers of t, the desired result follows. $\square$

Theorem 4.3 together with Equations (4.3) and (4.4) provide a way to compute the distribution of $c_1^{(n)}$ and the conditional distribution of $c_{i+1}^{(n)}$ given $c_i^{(n)}$. In particular, it is known that

$$\frac{E_n}{n!} = \frac{2^{n+2}}{\pi^{n+1}} + O\left(\left(\frac{2}{3\pi} \right)^n \right), \tag{4.6}$$

15

(see, for example, [49]). Plugging these asymptotics into Equation (4.5) gives the following corollary to Theorem 4.3 describing the limiting distributions of the superdiagonal entries; Figures 1, 2, 3, and 4 show that this is quite accurate when $n \geq 9$.

Corollary 4.4. *For any $0 \leq x \leq 1$, for any fixed integer $i \geq 1$, and for any $0 \leq a_i \leq n$, we have*

$$\lim_{n \to \infty} \Pr(c_1^{(n)} \leq x) = \sin(x\pi/2) \quad and$$

$$\lim_{n \to \infty} \Pr(c_{i+1}^{(n)} \leq x | c_i^{(n)} = a_i)) = \frac{\sin(\frac{\pi}{2} \min\{x, 1 - a_i\})}{\sin(\frac{\pi}{2}(1 - a_i))}.$$

Sketch of proof. Note that $n! f_{1,n}(1) = E_n$, which follows from Equation (2.2), since $f_{1,n}(1)$ is the volume of the polytope $\mathcal{T}_n$ of all $n+1$ by $n+1$ tridiagonal doubly stochastic matrices. To prove the corollary, it is sufficient to show that

$$\frac{n! f_{1,n}(x)}{E_n} \to \sin(\pi x/2)$$

as $n \to \infty$. We will leave out $\lfloor \bullet \rfloor$ and $\lceil \bullet \rceil$ notation below for readability.

First, we may use Equation (4.6), the fact that $x \leq 1$, a union bound, and Stirling's formula to show that

$$\left| \sum_{k=n/4}^{n/2} (-1)^k x^{2k+1} \frac{E_{n-(2k+1)}}{E_n} \binom{n}{2k+1} \right| \leq \frac{n}{(n/2)!} \frac{E_{n/2}}{(n/2)!} \frac{n!}{E_n} \leq \left(\frac{c}{n} \right)^{n/2}$$

for some constant c.

Second, note that the $n/4$ term Taylor series expansion for $\sin(\pi x/2)$ is a good approximation, namely:

$$\left| \sum_{k=0}^{n/4} (-1)^k \frac{(\pi x/2)^{2k+1}}{(2k+1)!} - \sin(\pi x/2) \right| \leq (c/n)^{n/4}.$$

Third, we may use Equation (4.6) and Stirling's formula to show that each term in the sum

$$\sum_{k=0}^{n/4} (-1)^k x^{2k+1} \frac{E_{n-(2k+1)}}{E_n} \binom{n}{2k+1} - (-1)^k \frac{(\pi x/2)^{2k+1}}{(2k+1)!} =$$

$$\sum_{k=0}^{n/4} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \left(\frac{E_{n-(2k+1)} n!}{(n - (2k+1))! E_n} - (\pi/2)^{2k+1} \right)$$

is at most $O\left((\pi^2/12)^{n/2}\right)$ in absolute value.

Combining the three steps above, we have that

$$\left| \frac{n! f_{1,n}(x)}{E_n} - \sin(\pi x/2) \right| \leq \left(\frac{\pi^2}{12} - o(1) \right)^{n/2},$$

where $o(1)$ is a small quantity that tends to zero as $n \to \infty$. $\qquad \square$

4.1 Interpreting the sine function in Corollary 4.4

The appearance of the sine function in Corollary 4.4 is notable, and it would be interesting to have a heuristic reason why the first superdiagonal entry has an asymptotic sine distribution. We are indebted to Neil O'Connell for pointing out that the limiting Markov chain described by Corollary 4.4 appears in work of Manon Defosseux [18] from a completely different direction. Defosseux [18] studies the eigenvalues of products of random reflections in dimension $n \geq 3$, and the main result in the dimension 3 case may be described as follows. Let u_k, for $k = 0, 1, 2, 3, \ldots$, be a sequence of iid uniform random points on the 2-dimensional real sphere $\{x_1^2 + x_2^2 + x_3^2 = 1\}$, represented as length-3 column vectors. Let R_k be the random reflection defined by $R_k := I - 2u_k u_k^{\mathrm{t}}$, where u_k^{t} of u_k. For $k \geq 0$, let $M_k = R_0 R_1 R_2 \ldots R_k$ be the product of the first $k + 1$ random reflection matrices defined in this way. Note that M_k has $(-1)^{k-1}$ as an eigenvalue, and the other two eigenvalues for M_k are a complex conjugate pair, say $e^{i\pi\alpha_k}$ and $e^{-i\pi\alpha_k}$ where $\alpha_k \in [0, 1]$. Then the following holds for the random sequence $(\alpha_k)_{k \geq 0}$:

Theorem 4.5. [18]

(i) *We have $\alpha_1 \geq \alpha_2 \leq \alpha_3 \geq \alpha_4 \leq \cdots$.*

(ii) *The α_k process is Markovian.*

(iii) *Finally,*

$$\Pr(\alpha_1 \leq x) = 1 - \cos(\pi x/2),$$

$$\Pr(\alpha_{2\ell} \leq x | \alpha_{2\ell-1} = b_{2\ell-1}) = \frac{\sin(\frac{\pi}{2} \min\{x, b_{2\ell-1}\})}{\sin(\frac{\pi}{2} b_{2\ell-1})}$$

$$\Pr(\alpha_{2\ell+1} \leq x | \alpha_{2\ell} = b_{2\ell}) = 1 - \frac{\cos(\frac{\pi}{2} \max\{x, b_{2\ell}\})}{\cos(\frac{\pi}{2} b_{2\ell})},$$

where $\ell \geq 1$ and $b_{2\ell}, b_{2\ell-1} \in [0, 1]$ are fixed.

The alternating sequence α_k may be used to define a new sequence β_k where $\beta_{2\ell} = \alpha_{2\ell}$ and $\beta_{2\ell-1} = 1 - \alpha_{2\ell-1}$ for all $\ell \geq 1$. One may check that the β_k sequence has exactly the distributions given in Corollary 4.4 for the limit as $n \to \infty$ of the c_i sequence. Aaron Abrams, Henry Landau, Zeph Landau, James Pommersheim, and Eric Zaslow (personal communication) provide an elementary proof of (iii) in the dimension 3 case. Peter Windridge [53, Section 4.2.1] derives the limiting density for the α_k, process which is a re-scaling of our result in Theorem 6.1(i).

5 Connections with random alternating permutations, Entringer numbers, and parking functions

Consider the following question: If an alternating permutation of length n is chosen uniformly at random and a_i denotes the number in coordinate i divided by n, what is the distribution of a_i for large n? For example, if $n = 3$, there are two alternating permutations, 132 and 231, and thus a_1 is 1/3 with probability 1/2 and is 2/3 with probability 1/2. The following result shows that the distribution of a_i as n goes to infinity (with i fixed) has a very close connection to the superdiagonal entries in a random tridiagonal doubly stochastic matrix.

Theorem 5.1. *Let K be a positive integer constant, let $(c_1^{(n)}, c_2^{(n)}, \ldots, c_n^{(n)})$ be an element of $\mathcal{T}_n$ chosen uniformly at random, let σ be a length n alternating permutation chosen uniformly at random, and let $a_i^{(n)} = \sigma(i)/n$ for each $1 \le i \le n$. For any real numbers $0 \le t_1, \ldots, t_K \le 1$, we have*

$$\lim_{n \to \infty} \Pr(\bigcap_{i=1}^K \{a_i^{(n)} \le t_i\}) = \lim_{n \to \infty} \Pr(\bigcap_{i=1}^K \{\tilde{c}_i^{(n)} \le t_i\}),$$

where

$$\tilde{c}_i^{(n)} = \begin{cases} c_i^{(n)} & \text{if } i \text{ is odd, and} \\ 1 - c_i^{(n)} & \text{if } i \text{ is even.} \end{cases}$$

In Section 4, we determine $\lim_{n \to \infty} \Pr(c_1^{(n)} \le t)$ exactly (see Corollary 4.4), which when combined with the above theorem thus also gives the limiting distribution of $a_1^{(n)}$.

The proof of Theorem 5.1 depends on three lemmas which we state and prove below. We will say that $\tilde{c}_i^{(n)}$ has *rank* k if $\tilde{c}_i^{(n)}$ is the k-th smallest among $\tilde{c}_1^{(n)}, \tilde{c}_2^{(n)}, \ldots, \tilde{c}_n^{(n)}$. We will omit the superscript (n) where the value of n is clear from context.

Lemma 5.2. *Let $(c_1, c_2, \ldots, c_n)$ be an element of $\mathcal{T}_n$ chosen uniformly at random, let*

$$\tilde{c}_i = \begin{cases} c_i & \text{if } i \text{ is odd, and} \\ 1 - c_i & \text{if } i \text{ is even,} \end{cases}$$

let τ be a length n alternating permutation chosen uniformly at random, and let $a_i = \tau(i)/n$. For any real numbers $0 \le t_1, t_2, \ldots, t_n \le 1$,

$$\Pr(\bigcap_{i=1}^n \tilde{c}_i \text{ has rank } \le \lfloor nt_i \rfloor) = \Pr(\bigcap_{i=1}^n a_i \le t_i).$$

Proof. Recall from Section 2 that a uniform random element of $\mathcal{T}_n$ may be chosen by picking real numbers $x_1 < x_2 < \cdots < x_n$ each independently and uniformly at random from $[0, 1]$ and choosing a length n alternating permutation σ uniformly from all length n alternating permutations and then setting $\tilde{c}_i = x_{\sigma(i)}$. The relative order of the $\tilde{c}_i$ is thus determined entirely by the alternating permutation σ; in particular,

$$\Pr(\bigcap_{i=1}^n \tilde{c}_i \text{ has rank } \le \lfloor nt_i \rfloor) = \Pr(\bigcap_{i=1}^n \sigma(i) \le \lfloor nt_i \rfloor) = \Pr(\bigcap_{i=1}^n \sigma(i) \le nt_i).$$

Since σ was chosen uniformly at random among length n alternating permutations, the proof is complete. $\square$

Lemma 5.3. *Let F be an arbitrary event, let $\tilde{c}_i$ be defined as in Lemma 5.2, and let $0 < t < 1$ be a constant. Then, for every $0 < \epsilon$ and for any $1 \le i \le n$ we have*

$$\Pr(\{\tilde{c}_i \le t_i - \epsilon\} \cap F) - \exp(-2n\epsilon^2) \le \Pr(\{\tilde{c}_i \text{has rank} \le \lfloor nt_i \rfloor\} \cap F)$$
$$\le \Pr(\{\tilde{c}_i \le t_i + \epsilon\} \cap F) + \exp(-2n\epsilon^2).$$

18

Proof. The idea of the proof is to show that the $\lfloor nt \rfloor$ smallest of the $\tilde{c}_i$ is typically very close to t.

Note that

$$\Pr(\{\tilde{c}_i \text{ has rank } \leq \lfloor nt \rfloor\} \cap F) \tag{5.1}$$
$$= \Pr(\{\tilde{c}_i \text{ has rank } \leq \lfloor nt \rfloor\} \cap \{\tilde{c}_i \leq t + \epsilon\} \cap F) +$$
$$\Pr(\{\tilde{c}_i \text{ has rank } \leq \lfloor nt \rfloor\} \cap \{\tilde{c}_i > t + \epsilon\} \cap F)$$
$$\leq \Pr(\{\tilde{c}_i \leq t + \epsilon\} \cap F) + \Pr(\{\tilde{c}_i \text{ has rank } \leq \lfloor nt \rfloor\} \cap \{\tilde{c}_i > t + \epsilon\}).$$
$$\leq \Pr(\{\tilde{c}_i \leq t + \epsilon\} \cap F) + \Pr(\text{ the } \lfloor nt \rfloor \text{ smallest of } n \text{ uniforms is } > t + \epsilon). \tag{5.2}$$

The last inequality uses the fact from the algorithm in Subsection 2.1 that the c_i, and hence the $\tilde{c}_i$, are uniquely determined by a set of n (distinct) elements of $[0, 1]$ chosen uniformly and independently at random along with an alternating permutation chosen uniformly at random.

We now note that

$$\Pr(\text{ the } \lfloor nt \rfloor \text{ smallest of } n \text{ uniforms is } > t + \epsilon) = \sum_{k=0}^{\lfloor nt \rfloor - 1} \binom{n}{k} (t + \epsilon)^k (1 - t - \epsilon)^{n-k}$$
$$\leq \exp\left(-2 \frac{(n(t + \epsilon) - nt)^2}{n}\right)$$
$$= \exp(-2n\epsilon^2),$$

where the inequality comes from using Hoeffding's inequality [34] to bound the tail of a binomial distribution.

On the other hand,

$$\Pr(\{\tilde{c}_i \text{ has rank } \leq \lfloor nt \rfloor\} \cap F)$$
$$\geq \Pr(\{\tilde{c}_i \text{ has rank } \leq \lfloor nt \rfloor\} \cap F \cap \{\tilde{c}_i \leq t - \epsilon\})$$
$$= \Pr(\{\tilde{c}_i \leq t - \epsilon\} \cap F) - \Pr(\{\tilde{c}_i \text{ has rank } > \lfloor nt \rfloor\} \cap F \cap \{\tilde{c}_i \leq t - \epsilon\}).$$
$$\geq \Pr(\{\tilde{c}_i \leq t - \epsilon\} \cap F) - \Pr(\text{ the } \lfloor nt \rfloor \text{ smallest of } n \text{ uniforms is } \leq t - \epsilon).$$

Using similar analysis to the above, we can show that

$$\Pr(\text{ the } \lceil nt \rceil \text{ smallest of } n \text{ uniforms is } \leq t - \epsilon) \leq \exp(-2n\epsilon^2),$$

and thus the proof is complete. $\qquad\square$

Lemma 5.4. *The function* $\lim_{n \to \infty} \Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \leq t_i)$ *is continuous in* $t_1, \ldots, t_K$.

Proof. Let $g_K(t_1, \ldots, t_K) := \lim_{n \to \infty} \Pr(\bigcap_{i=1}^{K} c_i^{(n)} \leq t_i)$, and note that the function $\lim_{n \to \infty} \Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \leq t_i)$ is continuous in $t_1, \ldots, t_K$ if and only if $g_K(t_1, \ldots, t_K)$ is continuous in $t_1, \ldots, t_K$.

We will use induction on K to prove that $g_K(t_1, \ldots, t_K)$ is continuous in $t_1, \ldots, t_K$. If $K = 1$, then $g_1(t_1) = \lim_{n \to \infty} \Pr(c_i^{(n)} \leq t_1) = \sin(t_1 \pi/2)$ by Corollary 4.4, and thus is continuous in t_1.

For the induction step, assume that $g_K(t_1, \ldots, t_K)$ is continuous in $t_1, \ldots, t_K$. We will show the corresponding statement for $K + 1$. Define

$$F_{\ell,n}(t_1, \ldots, t_\ell) := \int_0^{t_1} \int_0^{\min\{t_2, 1-x_1\}} \cdots \int_0^{\min\{t_\ell, 1-x_{\ell-1}\}} \int_0^{1-x_\ell} \cdots \int_0^{1-x_{n-1}} dx_n \cdots dx_1.$$

19

Thus, we have

$$g_{K+1}(t_1, \ldots, t_{K+1}) = \lim_{n \to \infty} \frac{F_{K+1,n}(t_1, \ldots, t_{K+1})}{E_n/n!},$$

since $F_{K+1,n}(1, \ldots, 1) = E_n/n!$ by Equation (2.2).

We may now write

$$F_{K+1,n}(t_1, \ldots, t_{K+1})$$
$$= \int_0^{t_1} \int_0^{\min\{t_2, 1-x_1\}} \cdots \int_0^{\min\{t_{K+1}, 1-x_K\}} \int_0^{1-x_{K+1}} \cdots \int_0^{1-x_{n-1}} dx_n \cdots dx_1,$$
$$= A_{K-1} \int_0^{\min\{t_K, 1-x_{K-1}\}} \int_0^{\min\{t_{K+1}, 1-x_K\}} B_{K+2,n} \, dx_n \cdots dx_1,$$

where for notational expedience, we define the symbols

$$A_{K-1} := \int_0^{t_1} \int_0^{\min\{t_2, 1-x_1\}} \cdots \int_0^{\min\{t_{K-1}, 1-x_{K-2}\}} \quad \text{and} \quad B_{K+2,n} := \int_0^{1-x_{K+1}} \cdots \int_0^{1-x_{n-1}}$$

to represent, respectively, the first $K - 1$ integrals over the variables $x_1, \ldots, x_{K-1}$ and the last $n - K - 1$ integrals over the variables $x_{K+2}, \ldots, x_n$. With this notation, we have

$$F_{K+1,n}(t_1, \ldots, t_{K+1})$$
$$= \quad A_{K-1} \int_0^{\min\{t_K, 1-x_{K-1}, 1-t_{K+1}\}} \int_0^{t_{K+1}} B_{K+2,n} \, dx_n \cdots dx_1$$
$$+ A_{K-1} \int_{\min\{t_K, 1-x_{K-1}, 1-t_{K+1}\}}^{\min\{t_{K+1}, 1-x_K\}} \int_0^{1-x_{K+1}} B_{K+2,n} \, dx_n \cdots dx_1$$
$$= \quad A_{K-1} \int_0^{\min\{t_K, 1-x_{K-1}, 1-t_{K+1}\}} dx_K \cdots dx_1 \int_0^{t_{K+1}} B_{K+2,n} \, dx_n \cdots dx_{K+1}$$
$$+ A_{K-1} \int_0^{\min\{t_{K+1}, 1-x_K\}} \int_0^{1-x_{K+1}} B_{K+2,n} \, dx_n \cdots dx_1$$
$$- A_{K-1} \int_0^{\min\{t_K, 1-x_{K-1}, 1-t_{K+1}\}} \int_0^{1-x_{K+1}} B_{K+2,n} \, dx_n \cdots dx_1$$
$$= \quad F_{K,K}(t_1, \ldots, t_{K-1}, \min\{t_K, 1 - t_{K+1}\}) \cdot F_{1,n-K}(t_{K+1})$$
$$+ F_{K,n}(t_1, \ldots, t_K) - F_{K,n}(t_1, \ldots, t_{K-1}, \min\{t_K, 1 - t_{K+1}\}).$$

Thus,

$$g_{K+1}(t_1, \ldots, t_{K+1})$$
$$= F_{K,K}(t_1, \ldots, t_{K-1}, \min\{t_K, 1 - t_{K+1}\}) \cdot \lim_{n \to \infty} \frac{E_{n-K}/(n-K)!}{E_n/n!} \frac{F_{1,n-K}(t_{K+1})}{E_{n-K}/(n-K)!}$$
$$+ g_K(t_1, \ldots, t_K) - g_K(t_1, \ldots, t_{K-1}, \min\{t_K, 1 - t_{K+1}\})$$
$$= F_{K,K}(t_1, \ldots, t_{K-1}, \min\{t_K, 1 - t_{K+1}\}) \cdot \left(\frac{\pi}{2}\right)^K \sin(\frac{\pi}{2} t_{K+1})$$
$$+ g_K(t_1, \ldots, t_K) - g_K(t_1, \ldots, t_{K-1}, \min\{t_K, 1 - t_{K+1}\}),$$

where the last equality follows from Equation (4.6) and from Corollary 4.4, using the fact that $F_{1,n-K}(t_{K+1}) = f_{1,n-K}(t_{K+1})$. It is not hard to show (by induction) that

20

$F_{K,K}(s_1, \ldots, s_K)$ is a composition of polynomials and the function $\min\{x, y\}$, and thus $F_{K,K}(t_1, \ldots, t_{K-1}, \min\{t_K, 1 - t_{K+1}\})$ is continuous in $t_1, \ldots, t_{K+1}$. Furthermore, the other functions that appear on the right-hand side of the last equation are all continuous by induction or by inspection. Thus, we have proven that $g_{K+1}(t_1, \ldots, t_{K+1})$ is continuous in $t_1, \ldots, t_{K+1}$. $\qquad\square$

We now return to the proof of the main theorem of this section.

Proof of Theorem 5.1. By Lemma 5.2 we have

$$\Pr(\bigcap_{i=1}^{K} a_i^{(n)} \leq t_i) = \Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \text{ has rank } \leq \lfloor nt_i \rfloor).$$

Iterating Lemma 5.3 K times, we have

$$\Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \leq t_i - \epsilon) - K\exp(-n\epsilon^2) \leq \Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \text{has rank} \leq \lfloor nt_k \rfloor) = \Pr(\bigcap_{i=1}^{K} a_i \leq t_i)$$

$$\leq \Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \leq t_i + \epsilon) + K\exp(-n\epsilon^2).$$

Taking the limit as n goes to infinity, we have

$$\lim_{n\to\infty} \Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \leq t_i - \epsilon) \leq \lim_{n\to\infty} \Pr(\bigcap_{i=1}^{K} a_i^{(n)} \leq t_i) \leq \lim_{n\to\infty} \Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \leq t_i + \epsilon).$$

By Lemma 5.4, the function $\lim_{n\to\infty} \Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \leq t_i)$ is continuous in $t_1, \ldots, t_K$, and thus, we can let ϵ tend to zero to prove that

$$\lim_{n\to\infty} \Pr(\bigcap_{i=1}^{K} a_i^{(n)} \leq t_i) = \lim_{n\to\infty} \Pr(\bigcap_{i=1}^{K} \tilde{c}_i^{(n)} \leq t_i).$$

$\qquad\square$

Remark 5.5. The results above suggest the following approximate picture of the coordinates (divided by n) of an alternating permutation chosen uniformly at random when n is large. We know from Theorem 5.1 and Corollary 4.4 that the first coordinate has distribution $\sin(a_1\pi/2)$. Given that the first coordinate takes the value a_1, Theorem 5.1 and Corollary 4.4 suggest that the second coordinate ought to have distribution function one minus $\cos(a_2\pi/2)/\cos(a_1\pi/2)$ conditioned on $a_1 \leq a_2 \leq 1$ (since the permutation is alternating up-down). Thus, the distribution function for the second coordinate should be

$$\max\left\{0, 1 - \frac{\cos(a_2\pi/2)}{\cos(a_1\pi/2)}\right\}.$$

Given that the second coordinate is a_2, the same heuristic suggests that third coordinate a_3 ought to be drawn from a sine distribution conditioned on $0 \leq a_3 \leq a_2$; in particular, it should be

$$\min\left\{1, \frac{\sin(a_3\pi/2)}{\sin(a_2\pi/2)}\right\}.$$

21

The distributions of the coordinates should continue in this way, with the distribution for odd i being determined by a sine distribution constrained by the fact that a_i must be larger than the previous coordinate, and with the distribution for even i being determined by a cosine distribution constrained by the fact that a_i must be smaller than the previous coordinate. Computer simulations give strong evidence for the claims above, and it would be interesting to prove them in detail and to study related questions, for example how the $i + 2$ coordinate is distributed given the value of the i-th coordinate.

5.1   Euler and Entringer numbers

The *Entringer number* $E(n, k)$ is defined to be the number of length $n + 1$ reverse alternating (down/up) permutations that start with $k + 1$. Thus, $E(n, n) = E_n$, the n-th Euler number, which is the number of alternating (up/down) permutations of length n. From the definition of Entringer numbers, we have for every non-negative real number t that

$$\lim_{n\to\infty} \Pr(a_1^{(n)} \le t) = \lim_{n\to\infty} \sum_{i=0}^{\lfloor nt \rfloor - 1} \frac{E(n, n - i)}{E_{n+1}}.$$

This fact lets us derive a local limit theorem, below, for the Entringer numbers, describing the growth of the Entringer numbers $E(n, k)$ as k increases compared to the total number of alternating permutations of length $n + 1$ in the large n limit. More information on Entringer numbers may be found in [38] and [49, Section 3].

Theorem 5.6. *Let $0 \le t \le 1$ be a constant. Then,*

$$\lim_{n\to\infty} \frac{nE(n, \lceil nt \rceil)}{E_{n+1}} = \frac{\pi}{2} \sin(t\frac{\pi}{2}).$$

Loosely put, $E(n, i)/E_{n+1} \sim \frac{\pi}{2n} \sin(\pi i/2n)$. Thus, $E(n, i)$ increases smoothly in i for large n. Before proving Theorem 5.6, we will remark on a nice heuristic reason that was pointed out by Eric Rains for the function $F(t) := \frac{2nE(n, \lceil nt \rceil)}{\pi E_{n+1}}$ to approach $\sin(t\frac{\pi}{2})$ in the limit as $n \to \infty$. One can prove (for example, by finding a bijection), that the second difference for the $E(n, j)$ sequence satisifes $E(n, j) - 2E(n, j - 1) + E(n, j - 2) = E(n - 2, j - 2)$. Multiplying both sides of this equation by $\frac{2n^3}{\pi E_{n+1}}$ and using the asymptotics of E_n (see Equation (4.6)) gives an analog of the differential equation $F''(t) = -\left(\frac{\pi}{2}\right)^2 F(t)$, and since $F(0) = 0$ and $F(1) \to 1$, we see that F ought to be asymptotically $\sin(t\frac{\pi}{2})$. We will now give a formal proof of this fact.

Proof of Theorem 5.6. The main idea is differentiating both sides of the equation in Theorem 5.1. We start with the equation

$$\sin(t\pi/2) = \lim_{n\to\infty} \Pr(c_1^{(n)} \le t) = \lim_{n\to\infty} \Pr(a_1^{(n)} \le t) = \lim_{n\to\infty} \sum_{i=0}^{\lfloor nt \rfloor - 1} \frac{E(n, n - i)}{E_{n+1}}. \quad (5.3)$$

All sides of Equation (5.3) are differentiable for $0 < t < 1$, are differentiable from the right at $t = 0$, and are differentiable from the left at $t = 1$. Because $\frac{\pi}{2} \cos(t\pi/2)$ is continuous for all $0 \le t \le 1$, it is sufficient to show that

$$\lim_{n\to\infty} \frac{nE(n, \lceil nt \rceil)}{E_{n+1}} = \frac{\pi}{2} \sin(t\frac{\pi}{2})$$

for all $0 < t < 1$, since the result at the endpoints follows from continuity in t. We will now proceed to bound the derivative of the right-hand-side of Equation (5.3) appropriately from above and from below.

Using the definition of the derivative and the fact that if a limit exists, it is equal to the right-hand limit, we have

$$
\begin{aligned}
\frac{\pi}{2}\cos(t\pi/2) &= \lim_{\Delta t \to 0^+}\frac{1}{\Delta t}\lim_{n\to\infty}\sum_{i=\lfloor nt \rfloor}^{\lfloor n(t+\Delta t) \rfloor - 1}\frac{E(n, n-i)}{E_{n+1}} \\
&\leq \lim_{\Delta t \to 0^+}\frac{1}{\Delta t}\lim_{n\to\infty}\left(\lfloor n(t+\Delta t)\rfloor - \lfloor nt\rfloor\right)\frac{E(n, n-\lfloor nt\rfloor)}{E_{n+1}} \\
&\leq \lim_{\Delta t \to 0^+}\frac{1}{\Delta t}\lim_{n\to\infty}\left(\Delta tn + 1\right)\frac{E(n, \lceil n(1-t)\rceil)}{E_{n+1}} \\
&= \lim_{n\to\infty}\frac{nE(n, \lceil n(1-t)\rceil)}{E_{n+1}},
\end{aligned}
$$

where the last equality follows from Lemma 5.7 below and the fact that $0 < t$ by assumption.

To provide a matching lower bound, we proceed in a similar fashion, using a left-hand limit instead of a right-hand limit.

$$
\begin{aligned}
\frac{\pi}{2}\cos(t\pi/2) &= \lim_{\Delta t \to 0^+}\frac{1}{-\Delta t}\lim_{n\to\infty}\sum_{i=\lfloor n(t-\Delta t)\rfloor}^{\lfloor nt\rfloor - 1}\frac{-E(n, n-i)}{E_{n+1}} \\
&\geq \lim_{\Delta t \to 0^+}\frac{1}{\Delta t}\lim_{n\to\infty}\left(\lfloor nt\rfloor - \lfloor n(t-\Delta t)\rfloor\right)\frac{E(n, n-\lfloor nt\rfloor + 1)}{E_{n+1}} \\
&\geq \lim_{\Delta t \to 0^+}\frac{1}{\Delta t}\lim_{n\to\infty}\left(\Delta tn - 1\right)\frac{E(n, \lceil n(1-t)\rceil)}{E_{n+1}} \\
&= \lim_{n\to\infty}\frac{nE(n, \lceil n(1-t)\rceil)}{E_{n+1}},
\end{aligned}
$$

where again the last equality holds due to Lemma 5.7. The upper and lower bounds are equal, and so the proof is complete. $\qquad\square$

Lemma 5.7. *If $0 < t \leq 1$, then*

$$
\frac{E(n, \lceil n(1-t)\rceil)}{E_{n+1}} \to 0
$$

as $n \to \infty$.

Proof. Note that $E(n, k)$ is increasing in k, since the set of length $n + 1$ reverse alternating permutations starting with k can be mapped injectively into the set of length $n + 1$ reverse alternating permutations starting with $k + 1$ by switching k and $k + 1$ in the permutation. Given $\delta > 0$, choose N_0 large enough that $n - \lceil n(1-t)\rceil > \lceil\frac{1}{\delta}\rceil$. Then, for every $n > N_0$, we have

$$
E_{n+1} = \sum_{i=0}^{n}E(n, n-i) > \sum_{i=0}^{\lceil 1/\delta\rceil}E(n, n-i) > \frac{1}{\delta}E(n, \lceil n(1-t)\rceil).
$$

$\qquad\square$

5.2 Chain polytopes and parking functions

In [11], Chebikin and Postnikov compute the volume of the chain polytope for any ribbon poset. In the special case where the ribbon poset has only the relations $x_1 < x_2 > x_3 < x_4 > \cdots$, the corresponding chain polytope is exactly $\mathcal{T}_n$, the polytope defined by the superdiagonal of a tridiagonal doubly stochastic matrix. Chebikin and Postnikov's main result [11, Theorem 3.1] can be used to evaluate $f_{1,n}(x)$ (see Equation (4.2)) in terms of a sum over parking functions of length n. We will state precisely below how this special case of [11, Theorem 3.1] relates to Theorem 4.3 and Corollary 4.4.

The sequence $(b_1, b_2, \ldots, b_n)$ is a *parking function* of length n if the reordered sequence $b'_1 \leq b'_2 \leq \cdots \leq b'_n$ satisfies $b'_i \leq i$ for each $1 \leq i \leq n$. For example, the parking functions of length 3 are 111, 112, 121, 211, 113, 131, 311, 122, 212, 221, 123, 132, 213, 231, 312, and 321. Let $\mathcal{P}_n$ be the set of all parking functions of length n. The sequence $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a *weak composition* of n if $0 \leq \alpha_i$ for each $1 \leq i \leq n$ and also $\sum_{i=1}^{n} \alpha_i = n$. Let K_n denote the set of weak compositions of n satisfying $\sum_{i=1}^{\ell} \alpha_i \geq \ell$ for all $1 \leq \ell \leq n$. Note that $(b_1, b_2, \ldots, b_n)$ is a parking function of length n if and only if it has content $\alpha \in K_n$, where the content of $(b_1, \ldots, b_n)$ is the list of non-negative integers $(c_1, \ldots, c_n)$ where c_j is the number of indices i such that $b_i = j$.

Theorem 5.8. [11] *For every $0 \leq x \leq 1$,*

$$n! f_{1,n}(x) = \left| \sum_{(b_1,\ldots,b_n) \in \mathcal{P}_n} \prod_{i=1}^{n} (-1)^{b_i} h(b_i) \right|$$

$$= \left| \sum_{\alpha \in K_n} \binom{n}{\alpha} (-1)^{\alpha_1 + \alpha_3 + \alpha_5 + \cdots} x^{\alpha_1} \right|,$$

where $h(b_i) = x$ if $b_i = 1$ and $h(b_i) = 1$ otherwise, and where $\binom{n}{\alpha} = \frac{n!}{\alpha_1! \alpha_2! \cdots \alpha_n!}$.

Note that the second equality above follows from grouping terms in the sum over parking functions. Combining Theorem 5.8 with Theorem 4.3 and Corollary 4.4 we have the following.

Corollary 5.9. *For $0 \leq x \leq 1$,*

$$\sin(x\pi/2) = \lim_{n \to \infty} \frac{1}{E_n} \left| \sum_{(b_1,\ldots,b_n) \in \mathcal{P}_n} \prod_{i=1}^{n} (-1)^{b_i} h(b_i) \right|$$

$$= \lim_{n \to \infty} \frac{1}{E_n} \left| \sum_{\alpha \in K_n} \binom{n}{\alpha} (-1)^{\alpha_1 + \alpha_3 + \alpha_5 + \cdots} x^{\alpha_1} \right|,$$

where $h(b_i) = x$ if $b_i = 1$ and $h(b_i) = 1$ otherwise, and where $\binom{n}{\alpha} = \frac{n!}{\alpha_1! \alpha_2! \cdots \alpha_n!}$.

Much more general chain polytopes of ribbon posets are considered in [11], and it would be interesting to see how much of the analysis of the current paper could be applied to the more general polytopes. The more general polytopes are unlikely to satisfy the Markov property analogous to Theorem 4.1; however, it seems like it may be possible to use similar analysis to study the distribution of a coordinate in a randomly chosen point in the polytope.

6 The limiting Markov chain

As shown in Section 4, in the large n limit, the entries above the diagonal in a uniformly chosen tridiagonal doubly stochastic matrix form a Markov chain with starting distribution

$$\Pr(c_1 \leq x) = \sin(x\pi/2) \qquad \text{and transition distribution} \qquad (6.1)$$

$$\Pr(c_{i+1} \leq y | c_i = x) = \frac{\sin\left(\frac{\pi}{2}\min\{y, 1-x\}\right)}{\sin(\frac{\pi}{2}(1-x))}, \qquad (6.2)$$

where $0 \leq x, y \leq 1$ (see Corollary 4.4).

In the development below, we determine the stationary distribution, eigenvalues, and eigenvectors, along with good rates of convergence for this chain. We summarize the main results:

Theorem 6.1. *For the Markov chain $K(x, dy)$ defined by Equations (6.1) and (6.2) on $[0, 1]$, we have the following:*

(i) *The stationary distribution has density $2\cos^2(\pi x/2) = \cos(\pi x) + 1 = \pi(x)$ with respect to Lebesgue measure on $[0, 1]$.*

(ii) *The Markov chain is reversible, with a compact, Hilbert-Schmidt kernel.*

(iii) *The eigenvalues are $\beta_0 = 1$, $\beta_1 = -1/3$, $\beta_2 = 1/5$, ..., $\beta_j = (-1)^j/(2j+1)$, ... (there is no other spectrum).*

(iv) *The eigenfunctions are transformed Jacobi polynomials.*

(v) *For any fixed starting state $x \in [0, 1]$ and all $\ell \geq 2$, we have*

$$4\left\|K_x^\ell - \pi\right\|_{TV}^2 \leq \sum_{i=1}^{\infty} i\left(\frac{1}{2i+1}\right)^{2\ell}.$$

The bound in (v) shows that the chain converges extremely rapidly (see, for example Table 1). Convergence from the true starting distribution may be even more rapid. The

ℓ	2	3	4	5	$\cdots$
$\left\|K_x^\ell - \pi\right\|_{TV}^2 \leq$	0.0185608791	0.0015383426	0.0001581840	0.0000171513	$\cdots$

Table 1: Upper bounds from Theorem 6.1(v) on the total variation distance of the limiting Markov chain K after ℓ steps.

Markov chain K defined by Equations (6.1) and (6.2) is a close relative of a collection of related chains, and some parts of the theorem hold in general. These are developed first.

Consider the following generalization. Let $F(x)$ be a distribution function on $[0, 1]$. We may form a Markov chain $\{Y_n\}$ on $[0, 1]$ with the following transitions:

$$\Pr(Y_{n+1} \leq y | Y_n = x) = \frac{F(\min\{y, (1-x)\})}{F(1-x)}. \qquad (6.3)$$

This has the following "stochastic meaning": From x, pick y from F, conditional on $y \in [0, 1-x]$. In the following, suppose that F is absolutely continuous with positive density $f(x)$ on $(0,1)$. Then, the chain defined by Equation (6.3) has a transition density:

$$k(x,y) = \begin{cases} f(y)/F(1-x) & \text{if } y \leq 1-x \\ 0 & \text{otherwise.} \end{cases} \quad (6.4)$$

Proposition 6.2. *The transition density $k(x,y)$ in Equation (6.4) is reversible with stationary density $\pi(x)$ (with respect to Lebesgue measure on $[0,1]$) where, up to normalization Z, we have*

$$\pi(x) = Z^{-1}f(x)F(1-x), \quad for\ 0 \leq x \leq 1, \quad and\ where \quad Z = \int_0^1 f(x)F(1-x)\,dx.$$

Proof. We must check that for all x, y that $\pi(x)k(x,y) = \pi(y)k(y,x)$. Both sides are zero unless $x + y \leq 1$. In this case,

$$\pi(x)k(x,y) = Z^{-1}f(x)F(1-x)\frac{f(y)}{F(1-x)} = Z^{-1}f(x)f(y) = \pi(y)k(y,x).$$

$\square$

Remark 6.3. Reversibility means the operator associated to k is self-adjoint on $L^2(\pi)$. This implies all the benefits of the spectral theorem—real spectrum (eigenvalues and eigenvectors if they exist). It seems a bit counterintuitive at first.

In our case, Proposition 6.2 gives an easy proof of Theorem 6.1(i):

Example 6.4. For $F(x) = \sin(\pi x/2)$, we have $f(x) = F'(x) = \frac{\pi}{2}\cos(\pi x/2)$ and $F(1-x) = \sin(\frac{\pi}{2}(1-x)) = \cos(\pi x/2)$, so

$$f(x)F(1-x) = \frac{\pi}{2}\cos^2(\pi x/2) = \frac{\pi}{4}\left(\cos(x\pi) + 1\right).$$

The normalizing constant comes from integration.

More generally, the following stochastic representation will be useful, and it puts us into the realm of iterated random functions [8], [21], [54].

Proposition 6.5. *The Markov chain generated by Equation (6.4) has the following stochastic representation:*

$$Y_{n+1} = F^{-1}(F(1-Y_n)U_{n+1}) \quad with\ \{U_i\}\ independent\ and\ uniform\ on\ [0,1]. \quad (6.5)$$

Proof. Note first that $F^{-1}(F(1-x)U) \leq 1-x$ if and only if $F(1-x)U \leq F(1-x)$, which always holds. Next, we compute

$$\Pr(F^{-1}(F(1-x)U) \leq y) = \Pr\left(U \leq \frac{F(\min\{y, 1-x\})}{F(1-x)}\right) = \frac{F(\min\{y, 1-x\})}{F(1-x)},$$

as required. $\square$

For strictly monotone F, we may make a one-to-one transformation in Equation (6.5), defining $W_n = F(Y_n)$. Then Equation (6.5) becomes $W_{n+1} = F(1-Y_n)U_{n+1}$. In our special case of $F(x) = \sin(\pi x/2)$, this becomes $W_{n+1} = \sin(\frac{\pi}{2}(1-Y_n))U_{n+1} = \cos(\frac{\pi}{2}Y_n)U_{n+1}$. Letting $V_n = \sin^2(\frac{\pi}{2}Y_n)$, we see by squaring that

$$V_{n+1} = (1-V_n)U_{n+1}^2. \quad (6.6)$$

Eigenfunction	Eigenvalue
$x - \frac{1}{4}$	$-\frac{1}{3}$
$x^2 - \frac{3}{4}x + \frac{1}{16}$	$\frac{1}{5}$
$x^3 - \frac{5}{4}x^2 + \frac{3}{8}x - \frac{1}{64}$	$-\frac{1}{7}$
$x^4 - \frac{7}{4}x^3 + \frac{15}{16}x^2 - \frac{5}{32}x + \frac{1}{256}$	$\frac{1}{9}$
$x^5 - \frac{9}{4}x^4 + \frac{7}{4}x^3 - \frac{35}{64}x^2 + \frac{15}{256}x - \frac{1}{1024}$	$-\frac{1}{11}$
$\vdots$	$\vdots$

Table 2: A few eigenfunctions of the V_n chain (see (6.6)) with corresponding eigenvalues.

Proof of Theorem 6.1. In outline, we analyze the V_n chain of Equation (6.6), finding the eigenvalues and a complete set of orthogonal polynomial eigenfunctions. Since $V_n = \sin^2(\pi Y_n/2)$, this gives the eigenvalues and eigenvectors of the Y_n chain from Theorem (6.1); If $P_n(x)$ is an eigenfunction of the V_n chain with eigenvalue λ, then $P_n(\sin^2(\pi x/2))$ is an eigenfunction of the Y_n chain with eigenvalue λ. The tools used here lean on the developments of [19], where many further details may be found.

To begin, note that the recurrence $V_{n+1} = (1 - V_n)U_{n+1}^2$ implies that the V_n chain has a full set of polynomial eigenvectors. To see this, consider first (changing notation slightly), $V_1 = (1 - V_0)U_1^2$ gives $\mathbb{E}(V_1|V_0 = v) = \mathbb{E}((1 - v)U_1^2) = (1 - v)\frac{1}{3}$. This implies that $\mathbb{E}((V_1 - \frac{1}{4})|V_0 = v) = -\frac{1}{3}(V_0 - \frac{1}{4})$, that is, $(V - \frac{1}{4})$ is an eigenfunction of the V-chain with eigenvalue $-1/3$. Similarly, $\mathbb{E}(V_1^2|V_0 = v) = (1 - v)^2\frac{1}{5}$ implies that the V-chain has a quadratic eigenfunction with eigenvalue $\frac{1}{5}$, and so on (see Table 2).

Since these are eigenfunctions for distinct eigenvalues for a self-adjoint operator on $L^2(\pi)$, they must be orthogonal polynomials for $\pi(x) = \beta(\frac{1}{2}, \frac{3}{2}; x)$, where $\pi(x)$ is the stationary distribution of the V_n chain. These are Jacobi polynomials $P_i^{-\frac{1}{2},\frac{1}{2}}$ (see [37, (1.8.2)]). These are classically given on $[-1, 1]$, and so we make the change of variables $x \mapsto \frac{1-x}{2}$ and write $p_i(x) = P_i^{-\frac{1}{2},\frac{1}{2}}(1 - 2x)$. Then

$$\int_0^1 p_j(x)p_k(x)\pi(x)\,dx = z_j^{-1}\delta_{j,k}, \tag{6.7}$$

where $z_j = \prod_{i=1}^{j}\left(1 - \frac{1}{2i}\right)^{-2}$.

Since the eigenvalues $(-1)^i/(2i + 1)$ are square summable, the operator is Hilbert-Schmidt. Since the Jacobi polynomials are a complete orthogonal system in $L^2(\pi)$, there is no further spectrum. This implies (ii), (iii), and (iv) of Theorem 6.1.

Finally, recall that the total variation distance $\left\|K_x^\ell - \pi\right\|_{\text{TV}} = \frac{1}{2}\int_0^1 \left|k^\ell(x,y) - \pi(y)\right|$ and the chi-square distance $\chi_x^2(\ell) = \int_0^1 \frac{|k^\ell(x,y) - \pi(y)|^2}{\pi(y)}\,dy$. Applying Cauchy-Schwartz, we have

$$4\left\|K_v^\ell - \pi\right\|_{\text{TV}}^2 \le \chi_v^2(\ell).$$

Using Mercer's theorem as in [19, Section 2.2.1] we see that

$$\chi_x^2(\ell) = \sum_{i=1}^{\infty}\frac{1}{(2i + 1)^\ell}p_i^2(x)z_i.$$

27

In [35, Lemma 4.2.1], it is shown that

$$\sup_{x \in [0,1]} |p_i| = \frac{(1/2)_i}{i!} = \frac{\frac{1}{2}(\frac{1}{2}+1)\cdots(\frac{1}{2}+i-1)}{i!} < 1.$$

The easy bound $z_i \leq i$ (in fact, $z_i \sim e^{\gamma/2}\sqrt{i}$) completes the proof of Theorem 6.1(v). $\square$

Returning to the generalization above, the same arguments work without essential change for the distribution function $F(x) = x^a$ on $[0,1]$, for any fixed $0 < a < \infty$. Then, the representation in Equation (6.5) gives the representation $Y_{n+1} = (1 - Y_n)U_{n+1}^{1/a}$. It follows that the chain has a $\beta(a, 1+a)$ stationary distribution and Jacobi polynomial eigenfunctions with eigenvalues $\frac{a}{i+a}$ for $0 \leq i < a$. Sharp rates of convergence as in Theorem 6.1(v) are straightforward to derive, as above. Further details are omitted.

7   The spectral gap and mixing time

Throughout this section, a random $(n+1) \times (n+1)$ tridiagonal doubly stochastic matrix M is chosen by choosing the above diagonal entries $c_1, c_2, \ldots, c_n$ from the limiting Markov chain defined by Equation (6.2) with c_1 chosen from the stationary distribution. As shown in Section 3 (see Figure 4), the stationary distribution gives a good approximation to the distribution of the superdiagonal entries of a random tridiagonal doubly stochastic matrix starting as early as the fourth superdiagonal entry. The approximation appears to be good even for small n—empirically, $n \geq 9$ is sufficient.

Since M is symmetric, it has real eigenvalues $\beta_0 = 1 \geq \beta_1 \geq \beta_2 \geq \cdots \geq \beta_n \geq -1$. Let $\mathrm{gap}_n(M) = 1 - \beta_1$ denote the spectral gap. The first result gives an upper bound on the gap.

Theorem 7.1. *For M of form (1.1) with $\{c_i\}_{i=1}^n$ chosen from the Markov chain defined by Equations (6.1),(6.2), if A_n tends to infinity as n tends to infinity, then with probability approaching one for all large n*

$$\mathrm{gap}_n(M)n^2 \log n < A_n.$$

This result is proved in Section 7.1. The simulations in Section 3 suggest that $\mathrm{gap}_n(M)n^2 \log n$ tends to a random variable. From the proof below, it is reasonable to conjecture that the limiting random variable has an asymmetric Cauchy distribution.

The second result gives a bound on the mixing time of the associated Markov chain. For simplicity, we work in continuous time, thus a rate one Poisson process directs the time of transitions from the matrix M. Let $K^t(x,y)$ be the associated Markov chain on $\{0, 1, \ldots, n\}$, $0 \leq t < \infty$. This chain has a uniform stationary distribution $\pi(j) = 1/(n+1)$. In Section 7.2 we prove

Theorem 7.2. *With notation as above, if A_n tends to infinity as n tends to infinity, with probability approaching one, if $t = A_n n^2 \log n$, then for all sufficiently large n*

$$\left\|K_0^t - \pi\right\|_{\mathrm{TV}} \leq 1/\sqrt{A_n}.$$

Note that by the results of Fill and Kahn [31], every birth-and-death chain with uniform stationary distribution has mixing time at least as large as the mixing time of

28

simple random walk on a path of length n, namely cn^2 where c is a constant (recall that simple random walk on a path of length n moves left or right with probability $1/2$ and holds at the endpoints with probability $1/2$). Thus, Theorem 7.2 shows that with probability approaching one, when sampling uniformly over all birth-and-death chains that have uniform stationary distribution, the mixing time is between cn^2 and a function very slightly larger than $n^2 \log n$.

These theorems show that, for typical M, the spectral gap times the mixing time is bounded. It follows from the results of [23, 27] that there is no cutoff in convergence to stationarity.

7.1 Bounding the spectral gap

Bounds on the spectral gap of the associated birth and death chain are obtained from a theorem of Miclo [42]. Let $m = \lfloor \frac{n}{2} \rfloor$ be the median of the stationary distribution π. Miclo shows that

$$\frac{1}{4B} \leq \text{gap}_n(M) \leq \frac{2}{B}$$

for $B = B_+(m) \vee B_-(m)$ with

$$B_+(m) = \max_{x>m} \left(\sum_{y=m+1}^{x} \frac{1}{\pi(y)c(y)} \right) \sum_{y=x}^{m+1} \pi(y) \quad \text{and} \quad B_-(m) = \max_{x<m} \left(\sum_{y=x}^{m-1} \frac{1}{\pi(y)c(y+1)} \right) \sum_{y=0}^{x} \pi(y).$$

In what follows, we want an upper bound on the spectral gap, and so a lower bound on B. Clearly, $B \geq B_- \geq B_* = \frac{n}{8} \sum_{y=m/4}^{m-1} \frac{1}{c(y)}$.

In outline, we bound the sum above by constructing the $c(i)$ chain via a coupling approach. This allows the sum above to be represented as a sum of independent blocks. Taking just the first term in each block gives a lower bound which is in the domain of attraction of a Cauchy distribution. Now, classical asymptotics shows that the sum is of size $C \cdot n \log n$, where C is a constant. Thus $B \geq C' n^2 \log n$ and $\text{gap}_n(M) \leq C''/(n^2 \log n)$.

To proceed, recall from Equations (6.1), (6.2) that the transition kernel has density (using $\sin\left(\frac{\pi}{2}(1-x)\right) = \cos\left(\frac{\pi}{2}x\right)$)

$$k(x,y) = \begin{cases} \left(\frac{\pi}{2}\right) \frac{\cos(\pi y/2)}{\cos(\pi x/2)} & \text{for } 0 \leq y \leq 1-x \\ 0 & \text{for } 1-x < y \leq 1. \end{cases}$$

For $0 < x \leq 1/2$, we may write this as a mixture density

$$k(x,y) = \epsilon 2\delta_{y \leq 1/2} + (1-\epsilon)q(x,y)$$

with

$$q(x,y) = (k(x,y) - \epsilon 2\delta_{y \leq 1/2})/(1-\epsilon)$$

for ϵ chosen so that $q(x,y) \geq 0$. Here $k(x,y)$ is monotone decreasing on $(0, 1-x)$. It takes value $c(x) = \left(\frac{\pi}{2}\right) \frac{\cos(\pi/4)}{\cos(\pi x/2)} \geq \left(\frac{\pi}{2}\right) \cos(\pi/4) = c$ and so $\epsilon = \frac{c}{2}$ works.

This allows the definition of a Markov chain $\{X_n, \delta_n\}$ on $[0,1] \times \{0,1\}$ with transitions

$$\Pr(\delta_{n+1} = 1 | X_n = x, \delta_n = \delta) = \begin{cases} 0 & \text{if } x > 1/2 \\ \epsilon & \text{if } x \le 1/2 \end{cases}$$

$$\Pr(\delta_{n+1} = 0 | X_n = x, \delta_n = \delta) = \begin{cases} 1 & \text{if } x > 1/2 \\ 1 - \epsilon & \text{if } x \le 1/2 \end{cases}$$

$$\Pr(X_{n+1} \in A | \delta_{n+1} = 0, X_n = x, \delta_n = \delta) = \begin{cases} k(x, A) & \text{if } x > 1/2 \\ q(x, A) & \text{if } x \le 1/2 \end{cases}$$

$$\Pr(X_{n+1} \in A | \delta_{n+1} = 1, X_n = x, \delta_n = \delta) = U_{1/2}(A).$$

This has a simple probabilistic interpretation: from x, if $x > 1/2$, set $\delta = 0$ and choose from $k(x, dy)$. If $x \le 1/2$, flip an ϵ coin. If heads, set $\delta = 1$ and choose from $U_{1/2}$ (the uniform distribution on $[0, 1/2]$). If tails, set $\delta = 0$ and choose from $q(x, dy)$. Thus, when $\delta = 1$, choices are made from $U_{1/2}$ independent of x. The marginal distribution of the X_n chain is a realization of the $k(x, dy)$ chain under study. To study a Markov chain with starting distribution π_0, choose $X_0 \sim \pi_0$ and set $\delta_0 = 0$. Clearly, the marginal distribution of $\{X_i\}_{0 \le i < \infty}$ is our underlying Markov chain.

Let ζ_i be the times i that $\delta_n = 1$ (so $\zeta_0 = 0$ and $\zeta_i = \inf\{n > \zeta_{i-1}, \delta_n = 1\}$). The sequence $\{\zeta_{i+1} - \zeta_i\}$ for $i \ge 1$ is independent and identically distributed. For any $i \gtrsim 1$, X_{ζ_i} is independent of $\{X_0, X_1, \ldots, X_{\zeta_i - 1}\}$. Therefore, the blocks are independent. We may bound

$$\sum_{i=1}^{N} \frac{1}{X_i} \ge \sum_{\zeta_i \le N} \frac{1}{X_{\zeta_i}} = \sum_{j=1}^{T} \frac{1}{v_j},$$

where v_j are independent identically distributed uniform on $[0, 1/2]$, and $T = \max\{i \le N : \delta_i = 1\}$. Now, the basic chain $k(x, dy)$ has the property that there is $a > 0$ such that $k(x, (0, 1/2]) \ge a$, uniformly in x. It follows that $\Pr(\delta_{i+2} = 1 | (X_0, \delta_0), (X_1, \delta_1), \ldots, (X_i, \delta_i)) \ge a\epsilon$ and so, for large N, we have $T \ge \left(\frac{\epsilon a}{10}\right) N$ with probability exponentially close to one. We have that

$$\frac{1}{m \log m} \sum_{i=1}^{m} \frac{1}{v_i} \to 2$$

in probability (see, for example, [29, page 41] on triangular arrays), which completes the proof of Theorem 7.1.

7.2 Bounding the mixing time

Consider a fixed tridiagonal doubly stochastic matrix M of the form in Display (1.1) with $c_1, c_2, \ldots, c_n$ above the diagonal. In this section, a continuous version of the associated birth and death chain is considered. Thus, a rate one Poisson process directs the moves which then take place according to the discrete rates. Alternatively, from state i the process remains at i for an exponential time with mean $1/(c_{i-1} + c_i)$. It then moves to $i - 1$ or $i + 1$ with respective probabilities $c_{i-1}/(c_{i-1} + c_i)$ and $c_i/(c_{i-1} + c_i)$ (when $i = 0$, the process always moves to 1; when $i = n$, the process always moves to $n - 1$). The advantage of working in continuous time is that two independent such processes cannot pass each other without meeting (with probability one).

Consider two such chains (X_t, Y_t) evolving independently with $X_0 = 0$ and Y_0 uniformly distributed on $\{0, 1, \ldots, n\}$. Let T be the first coupling time and T_0 the first time that Y_t hits 0, see for example [50],[40] for background. The standard coupling bound and elementary manipulations give

$$\left\|K_0^t - \pi\right\|_{\mathrm{TV}} \le \Pr(T > t) \le \Pr(Y_s \ne 0, 0 \le s \le t) \le \frac{1}{n+1} \sum_{i=1}^{n} \Pr(Y_s \ne 0, 0 < s \le t | Y_0 = i)$$

$$= \frac{1}{n+1} \sum_{i=1}^{n} \Pr_i(T_0 > t) \le \frac{1}{(n+1)t} \sum_{i=1}^{n} \mathbb{E}_i(T_0). \qquad (7.1)$$

It is shown below that

$$\mathbb{E}_i(T_0) = \frac{n}{c_1} + \frac{n-1}{c_2} + \cdots + \frac{n+1-i}{c_i}, \qquad \text{for } 1 \le i \le n. \qquad (7.2)$$

It follows that the right-hand side of Inequality (7.1) equals

$$\frac{1}{(n+1)t} \left(\frac{n^2}{c_1} + \frac{(n-1)^2}{c_2} + \cdots + \frac{1}{c_n} \right) \le \frac{n+1}{t} \left(\frac{1}{c_1} + \frac{1}{c_2} + \cdots + \frac{1}{c_n} \right). \qquad (7.3)$$

It will be further shown that $\left(\frac{1}{c_1} + \cdots + \frac{1}{c_n} \right)$ has order $n \log n$ when the c_i are chosen stochastically. These ingredients combine to prove Theorem 7.2.

To prove Equation (7.2), set $\mu_i = \mathbb{E}_i(T_0)$. Clearly, $\mu_n = \frac{1}{c_n} + \mu_{n-1}$. Similarly, $\mu_{n-1}(c_n + c_{n-1}) = 1 + c_n \left(\frac{1}{c_n} + \mu_{n-1} \right) + c_{n-1}\mu_{n-2}$. Equivalently, $\mu_{n-1}c_{n-1} = 2 + c_{n-1}\mu_{n-2}$ or $\mu_{n-1} = \frac{2}{c_{n-1}} + \mu_{n-2}$. Similarly, $\mu_{n-i} = \frac{i+1}{c_{n-i}} + \mu_{n-(i+1)}, \ldots, \mu_1 = \frac{n}{c_1}$. Working up from the bottom, $\mu_2 = \frac{n-1}{c_2} + \frac{n}{c_1}, \ldots, \mu_i = \frac{n-i+1}{c_i} + \frac{n-i+2}{c_{i-1}} + \cdots + \frac{n}{c_1}$. This proves Equation (7.2).

It remains to bound $1/c_1 + \cdots + 1/c_n$. We suppose that c_1 is chosen from the stationary distribution, so that $c_1, c_2, \ldots, c_n$ all have the same distribution with density $1 + \cos(\pi x)$ on $[0, 1]$. Set $Y_i = 1/c_i$ and $Y_i' = Y_i \delta_{Y_i \le n \log n}$ for $1 \le i \le n$. Note that $\Pr(Y_i > n \log n) = \int_0^{1/(n \log n)} (1 + \cos(\pi x)) \, dx \le 2/(n \log n)$. Thus, $\Pr \left(\bigcup_{i=1}^{n} \{Y_i > n \log n\} \right) \le 2/\log n \to 0$ as $n \to \infty$. So it is enough to study $\{Y_i'\}$. Now, $\mathbb{E}(Y_i') = \int_{1/(n \log n)}^{1} \frac{1}{t}(1 + \cos(\pi/t)) \, dt \le 2 \int_{1/(n \log n)}^{1} \frac{dx}{t} = 2 \log(n \log n) \le 3 \log n$. It follows that for all $B \ge 1$ we have $\Pr \left(\sum_{i=1}^{n} Y_i' \ge Bn \log n \right) \le \frac{3}{B}$. Combining bounds, we have

$$\Pr \left(\sum_{i=1}^{n} \frac{1}{c_i} > Bn \log n \right) \le \frac{3}{B} + \frac{2}{\log n}. \qquad (7.4)$$

Using Inequalities (7.1), (7.3), and (7.4) with $t = A(n+1)^2 \log n$, for any $A, B \ge 1$ and all n, we have

$$\left\|K_0^t - \pi\right\|_{\mathrm{TV}} \le \frac{B}{A} \qquad \text{with probability } 1 - \left(\frac{3}{B} + \frac{2}{\log n} \right). \qquad (7.5)$$

This proves Theorem 7.2 by taking $B_n = \sqrt{A_n}$. $\qquad \square$

We would like to thank one of the referees for pointing out that the above argument can be extended to bound the maximum mixing time for the chain started from any

point, rather than just for the chain started at the left endpoint, thus providing a bound on the so-called worst-case total variation (see [27]). In particular, one can show that if A_n tends to infinity as $n \to \infty$ and if $t = A_n n^2 \log n$, then

$$\max_{0 \le i \le n} \left\| K_i^t - \pi \right\|_{\text{TV}} \le 1/\sqrt{A_n}.$$

8   Follow up work

In follow up work to the present paper, our student Aaron Smith [45] has extended and refined our results in various ways. His main contribution extends results to birth and death chains with non-uniform stationary distributions. Smith still finds that in a uniform choice from the appropriate set of tridiagonal matrices, the entries above the diagonal determine the matrices, and furthermore, the same entries form a Markov chain. He finds that for some stationary distributions, e.g., a discrete exponential, most chains show a cutoff. Along the way, Smith [45] studies a slight variant on the Gibbs sampler in Section 2.2. In the variant, the coordinate i that will be updated is chosen at random, and the new value x is chosen uniformly from $[0, \min\{1 - c_{i-1}, 1 - c_{i+1}\}]$ so that c_i may always be replaced by x. With these modifications, Smith proves that the Gibbs sampler converges to the uniform distribution on $\mathcal{T}_n$ after at most $O(n \log^2 n)$ steps (a lower bound if $\Omega(n \log n)$ follows from the coupon collector problem). Smith [45] also considers a sped-up Gibbs sampler for tridiagonal matrices that updates a randomly chosen block of k consecutive entries on the superdiagonal, where k is a constant (in particular, $k = 56$ is analyzed). Here, Smith [45] proves that this Gibbs sampler converges to uniform in $\Theta(n \log n)$ steps.

Smith [45] also applies the main theorem of [6] to the process Y_t defined by Equations (6.1) and (6.2), which has transition density $k(x, y) = \frac{\pi \cos(\pi y/2)}{2 \cos(\pi x/2)}$. In particular, [45] shows that for $X_t := Y_t^{-1}$, the partial sum stochastic process

$$V_n(t) = \sum_{k=1}^{\lfloor nt \rfloor} \frac{X_k}{a_n} - \lfloor nt \rfloor \, \mathbb{E}\left(\frac{X_1}{a_n} \mathbf{1}_{\left\{ \frac{X_1}{a_n} \le 1 \right\}} \right), \qquad t \in [0, 1],$$

where a_n is a sequence of positive real numbers satisfying $n \Pr(|X_1| > a_n) \to 1$ as $n \to \infty$, converges in a certain sense to an α-stable Lévy process. The characteristic triple for the Lévy process is also computed in [45] using [6, Theorem 3.4].

Acknowledgments

References

[1] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8(1):69–97, 1987.

[2] Greg Anderson and Ofer Zeitouni. Some remarks on the Diaconis-Wood chain. *personal communication*, pages 1–2, November 06, 2011.

[3] Desiré André. Développement de sec x and tg x. *C.R. Math. Acad. Sci. Paris*, 88:965–979, 1879.

[4] Javiera Barrera M., Olivier Bertoncini, and Roberto Fernández. Abrupt convergence and escape behavior for birth and death chains. unpublished, 2009.

[5] Yuliy Baryshnikov and Dan Romik. Enumeration formulas for Young tableaux in a diagonal strip. *Israel J. Math.*, 178:157–186, 2010.

[6] Bojan Basrak, Danijel Krizmani´ c, and Johan Segers. A functional limit theorem for partial sums of dependent random variables with infinite variance. *preprint,* arXiv:1001.1345v1 [math.PR], pages 1–26, 8 Jan 2010.

[7] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2(2):294–313, 1992.

[8] Rabi Bhattacharya and Mukul Majumdar. *Random dynamical systems*. Cambridge University Press, Cambridge, 2007. Theory and applications.

[9] Stephen Boyd, Persi Diaconis, Jun Sun, and Lin Xiao. Fastest mixing Markov chain on a path. *Amer. Math. Monthly*, 113(1):70–74, 2006.

[10] Denis Chebikin and Richard Ehrenborg. The f-vector of the descent polytope. *preprint,* arXiv:0812.1249 [math.CO], pages 1–14, 6 Dec 2008.

[11] Denis Chebikin and Alexander Postnikov. Generalized parking functions, descent numbers, and chain polytopes of ribbon posets. *Adv. in Appl. Math.*, 44(2):145–154, 2010.

[12] Guan-Yu Chen. *The cutoff phenomenon for finite Markov chains.* PhD thesis, Cornell University, 2006. http://hdl.handle.net/1813/3047.

[13] Guan-Yu Chen and Laurent Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.*, 13:no. 3, 26–78, 2008.

[14] Guan-Yu Chen and Laurent Saloff-Coste. The cutoff phenomenon for randomized riffle shuffles. *Random Structures Algorithms*, 32(3):346–374, 2008.

[15] L. Costa, C.M. da Fonseca, and E.A. Martins. The number of faces of the tridiagonal birkhoff polytope. *Journal of Math. Sci.*, 161(6):867–877, 2009.

[16] C. M. da Fonseca and E. Marques de Sá. Fibonacci numbers, alternating parity sequences and faces of the tridiagonal Birkhoff polytope. *Discrete Math.*, 308(7):1308–1318, 2008.

[17] Geir Dahl. Tridiagonal doubly stochastic matrices. *Linear Algebra Appl.*, 390:197–208, 2004.

[18] Manon Defosseux. Eigenvalues of products of random reflections. *preprint*, pages 1–5, 2010.

[19] P. Diaconis, K. Khare, and L. Saloff-Coste. Gibbs sampling, exponential families and orthogonal polynomials, with discussion. *Statist. Sci.*, 23:151–200, 2008.

[20] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.*, 93(4):1659–1664, 1996.

[21] Persi Diaconis and David Freedman. Iterated random functions. *SIAM Rev.*, 41(1):45–76 (electronic), 1999.

[22] Persi Diaconis, Yuval Peres, Laurent Saloff-Coste, Ben Morris, David Aldous, David Revelle, Peter Winkler, David Wilson, Jim Fill, Dana Randal, Tom Hayes, Gérard Ben-Arous, Sharad Goel, Asaf Nachmias, Peter Ralph, and Devavrat Shah. Arcc workshop: Sharp thresholds for mixing times; speakers listed on order followed by scribes. *online at http://www.aimath.org/WWN/mixingtimes*, pages 1–10, 2004.

[23] Persi Diaconis and Laurent Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006.

[24] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

[25] Persi Diaconis and Mehrdad Shahshahani. Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.*, 18(1):208–218, 1987.

[26] Christoph Diehl. *Cut-off bei geburts-und todesprozessen*. PhD thesis, Westfälische Wilhelms-Universität Münster, 2009.

[27] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total variation cutoff in birth-and-death chains. Technical report, University of California, Berkeley, 2008.

[28] Ioana Dumitriu and Alan Edelman. Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830–5847, 2002.

[29] Richard Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.

[30] Noam D. Elkies. On the sums $\sum_{k=-\infty}^{\infty}(4k+1)^{-n}$. *Amer. Math. Monthly*, 110(7):561–573, 2003.

[31] James Allen Fill and Jonas Kahn. Fastest-mixing markov chains via comparison inequalities. *manuscript in preparation*, 2010.

[32] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

[33] Ilya Ya. Goldsheid and Boris A. Khoruzhenko. Eigenvalue curves of asymmetric tridiagonal random matrices. *Electron. J. Probab.*, 5:no. 16, 28 pp., 2000.

[34] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[35] Mourad E. H. Ismail. *Classical and quantum orthogonal polynomials in one variable*, volume 98 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005. With two chapters by Walter Van Assche, With a foreword by Richard A. Askey.

[36] Gil Kalai and G¨unter Ziegler. *Polytopes - Combinatorics and Computation*, volume 29 of *Oberwolfach Seminars*. Birkhäuser Verlag, Basel–Boston–Berlin, 2000.

[37] R. Koekoek and R.F. Swarttouw. *The Askey-scheme of hypergeometric orthogonal polynomials and its q-analog*. Delft University of Technology, online: `http://math.nist.gov/opsf/projects/koekoek.html`, 1994. Faculty of Technical Mathematics and Informatics, Report 94-05.

[38] A. G. Kuznetsov, I. M. Pak, and A. E. Postnikov. Increasing trees and alternating permutations. *Uspekhi Mat. Nauk*, 49(6(300)):79–110, 1994.

[39] Nathan Levy, Oren Somekh, Shlomo Shamai, and Ofer Zeitouni. On certain large random Hermitian Jacobi matrices with applications to wireless communications. *IEEE Trans. Inform. Theory*, 55(4):1534–1554, 2009.

[40] Torgny Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.

[41] Eyal Lubetzky and Allan Sly. Cutoff phenomena for random walks on random regular graphs, 2008. `http://www.citebase.org/abstract?id=oai:arXiv.org:0812.0060`.

[42] L. Miclo. An example of application of discrete Hardy's inequalities. *Markov Process. Related Fields*, 5(3):319–330, 1999.

[43] Albert Nijenhuis and Herbert S. Wilf. *Combinatorial algorithms*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, second edition, 1978. For computers and calculators, Computer Science and Applied Mathematics.

[44] Igor Pak and Van H. Vu. On mixing of certain random walks, cutoff phenomenon and sharp threshold of random matroid processes. *Discrete Appl. Math.*, 110(2-3):251–272, 2001.

[45] Aaron Smith. On gibbs samplers and random birth and death chains. Ph.D Dissertation, Department of Mathematics, Stanford University, 2011.

[46] Richard Stanley, I. G. Macdonald, and Roger B. Nelsen. Problems and Solutions: Solutions of Elementary Problems: E2701. *Amer. Math. Monthly*, 86(5):396, 1979.

[47] Richard P. Stanley. Two poset polytopes. *Discrete Comput. Geom.*, 1(1):9–23, 1986.

[48] Richard P. Stanley. *Enumerative combinatorics. Vol. 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997. With a foreword by Gian-Carlo Rota, Corrected reprint of the 1986 original.

[49] Richard P. Stanley. A survey of alternating permutations. *in preparation*, 2009.

[50] Hermann Thorisson. *Coupling, stationarity, and regeneration.* Probability and its Applications (New York). Springer-Verlag, New York, 2000.

[51] Hale F. Trotter. Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock, and Szegő. *Adv. in Math.*, 54(1):67–82, 1984.

[52] Benedek Valkó and Bálint Virág. Continuum limits of random matrices and the Brownian carousel. *Invent. Math.*, 177(3):463–508, 2009.

[53] Peter Windridge. Random matrix theory and the zeros of riemann's zeta function. *Master's Thesis* Department of Statistics, University of Warwick, pages 1–64, 2006.

[54] Wei Biao Wu and Michael Woodroofe. A central limit theorem for iterated random functions. *J. Appl. Probab.*, 37(3):748–755, 2000.