

PROPERTIES OF UNIFORM DOUBLY STOCHASTIC MATRICES

SOURAV CHATTERJEE, PERSI DIACONIS, AND ALLAN SLY

ABSTRACT. We investigate the properties of uniform doubly stochastic random matrices, that is non-negative matrices conditioned to have their rows and columns sum to 1. The rescaled marginal distributions are shown to converge to exponential distributions and indeed even large sub-matrices of side-length $o(n^{1/2-\epsilon})$ behave like independent exponentials. We determine the limiting empirical distribution of the singular values of the matrix. Finally the mixing time of the associated Markov chains is shown to be exactly 2 with high probability.

Random matrices have become a central area of focus for modern probability theory and numerous models have been intensely studied including Wigner, Wishart, GOE and GUE matrices [3]. In this paper we study a model for which much less is known, namely uniformly chosen entries of the set of doubly stochastic matrices (called Uniformly Distributed Stochastic Matrices). The Birkhoff polytope is an $(n-1)^2$ dimensional polytope in \mathbb{R}^{n^2} constituting the set of doubly stochastic matrices and is the convex hull of the permutation matrices (see e.g. [41]). While its extreme points are sparse matrices we shall see that typical entries chosen according to the uniform distribution are by contrast very dense. Little is known about the properties of uniformly distributed stochastic matrices as they fall outside the scope of techniques from the usual random matrix theory, however, important recent progress has been made by Barvinok and Hartigan.

We will let $X = (X_{ij})_{i,j=1,\dots,n}$ denote a uniform doubly stochastic matrix. By symmetry its rows and columns are exchangeable and all its entries have the same marginal distribution. It is natural then to ask what is the limiting distribution of nX_{11} , the first entry rescaled to have mean 1. In our first result we determine that the rescaled marginal distribution converges to an exponential random variable of mean 1.

Theorem 1. *With $X = (X_{ij})_{i,j=1,\dots,n}$ a uniformly chosen doubly stochastic matrix we have that,*

$$nX_{11} \xrightarrow{d} \exp(1)$$

as $n \rightarrow \infty$ where the convergence is in total variation distance. Further, for any $\epsilon > 0$,

$$d_{\text{tv}}(nX_{11}, \exp(1)) = O(n^{-1/2+\epsilon}).$$

A natural extension to this question is to ask about the joint distribution for a collection of several entries. It can be shown using the same approach that finite collections of random variables converge to independent exponentials with mean 1. This convergence holds not just in distribution but also in total-variation distance and its moments converge to the moments of independent exponentials (see Section 3.2). We believe that in many ways uniformly distributed stochastic

Key words and phrases. Doubly Stochastic Matrices, Birkhoff polytope.

Sourav Chatterjee's research was partially supported by NSF grant DMS-0707054 and a Sloan Research Fellowship.

matrices behave much like matrices of independent. For example the largest entry of the matrix is at most $(2 + o(1))\frac{1}{n} \log n$ with high probability,

Theorem 2. *For any $\epsilon > 0$,*

$$P \left(\max_{1 \leq i, j \leq n} nX_{ij} > (2 + \epsilon) \log n \right) \rightarrow 0,$$

as $n \rightarrow \infty$.

Another question one may ask is the limiting distribution of the singular values of $\bar{X} = n^{1/2}(X - EX)$. Denote these by $0 \leq \sigma_1(\bar{X}) \leq \dots \leq \sigma_n(\bar{X})$. Letting μ denote the measure on $[0, 2]$ with density

$$\frac{1}{\pi} \sqrt{4 - x^2}$$

we have the following result.

Theorem 3. *The limiting empirical singular value distribution of \bar{X} is given by*

$$\sum_{i=1}^n \delta_{\sigma_i(\bar{X})} \rightarrow \mu$$

where the convergence is in the weak topology, in probability as $n \rightarrow \infty$.

We conjecture that the empirical spectral distribution converges to the circular law.

One natural question is to ask how large a sub-matrix can one take so that the entries are still asymptotically independent. This problem was studied in the context of the random orthogonal matrix [31] where it was shown that an $k \times k$ sub-matrix is asymptotically distributed as independent normal random variables in total variation provided $k = o(n^{\frac{1}{2}})$ answering a question of the second author [24]. In [31] it is further shown that order $n/\log n$ entries simultaneously converge if weaker topologies are used. Here we show that for sub-matrices of uniformly distributed stochastic matrices of size almost $n^{1/2}$ the entries are asymptotically independent.

Theorem 4. *Let V denote the projection of a uniformly distributed stochastic matrix onto the $k \times k$ -sub-matrix of its first k rows and columns and let Δ be a $k \times k$ matrix of independent mean one exponential random variables. When $k = O(\frac{\sqrt{n}}{\log n})$ the rescaled law of V converges to Δ ,*

$$d_{\text{tv}}(nV, \Delta) \rightarrow 0$$

as $n \rightarrow \infty$ where d_{tv} denotes the total variation distance.

Unlike most other classes of random matrices, uniformly distributed stochastic matrices are of course stochastic which raises the question of the properties of the associated Markov chains. For any doubly stochastic Markov transition kernel the stationary distribution is the uniform distribution. For a uniform stochastic (but not necessarily doubly stochastic) matrix, that is a uniformly chosen Markov chain, the mixing time is two asymptotically almost surely [1]. We show that this holds also for uniformly chosen doubly stochastic random matrices.

Theorem 5. *The mixing time of the Markov chain given by a uniform double stochastic matrix is with high probability 2.*

In Section 1 we give background and history for the Birkhoff polytope. In Section 2 we give the proofs of Theorems 1 and 4. Then in Section 3 we begin by studying polytopes of matrices with non-constant row sums. By establishing that the volumes of the polytopes are maximized when the

row and column sums are equal, we get strong control over the distribution of a row in a uniformly distributed stochastic matrix through which we can bound the tails of the marginal distributions establishing convergence of the moments and Theorem 2. Finally, knowing that the entries are not too large allows us to show strong concentration for the entries of X^2 which guarantees that the mixing time is 2.

1. BACKGROUND

This section gives background and references for four topics that motivate our work: the Birkhoff polytope, prior distributions on Markov chains, limit theorems for entries of large random matrices in classical compact groups and contingency tables with fixed row and column sums

1.1. The Birkhoff Polytope. The set \mathcal{M}_n of $n \times n$ doubly stochastic matrices is known as the Birkhoff polytope, the bistochastic polytope and the assignment polytope. It is a basic object of study in operations research because of its appearance as the feasible set for the assignment problem. Given a cost matrix C_{ij} this asks for a permutation σ minimizing $\sum_i C_{i\sigma(i)}$. This is the same problem as minimizing $\sum_{ij} C_{ij} M_{ij}$ for $M \in \mathcal{M}_n$ because of Birkhoff's Theorem: the permutation matrices are the extreme points of \mathcal{M}_n . A thorough treatment of the assignment problem is in [33].

Because of this connection, the structure of \mathcal{M}_n has been intensively studied. Two permutations σ, ζ are adjacent on \mathcal{M}_n if and only if $\sigma\zeta^{-1}$ is a cycle (see [25] page 214). The diameter (the maximum distance between two vertices on the skeleton) of \mathcal{M}_n is two [25]. The face structure of \mathcal{M}_n is described in [11]. Finding a closed form expression for the volume of \mathcal{M}_n is a well known open problem. The volume is a rational number and is known for $n \leq 14$ (see [16] and references therein). The combinatorics suggest a simple probability problem: what is the mixing time of the nearest neighbor random walk on vertices of \mathcal{M}_n ? Pak [39] showed that it is two.

Birkhoff's characterisation of the extreme points is "equivalent" to other basic theorems in combinatorics such as Kontg's Lemma, Hall's Marriage Theorem and the Max-flow Min-Cut Theorem. A splendid account of these connections is in [33].

There are other polytopes with similarly nice descriptions. For example, the symmetric doubly stochastic matrices have extreme points $\frac{1}{2}(A_\sigma + A_\sigma^T)$ with A_σ the permutation matrix of σ [14, 40]. Perhaps the methods and results of our paper can be used to study the behavior of a randomly chosen point in these polytopes. The properties of the random tri-diagonal doubly stochastic matrices are thoroughly studied in [19].

1.2. Statistical Analysis of Markov Chains. Our original motivation for this work comes from the statistical analysis of a Markov chain on $\{1, 2, \dots, n\}$ with unknown transition matrix $(X_{ij}) \in Q_n$ (Q_n the set of stochastic matrices). One observes a run R_0, R_1, \dots, R_N and is required to estimate (X_{ij}) . A Bayesian approach to this problem starts with a prior distribution on Q_n . The classical Bayesian approach using, conjugate priors, sets each row to be an independent Dirichlet distribution. One natural choice has each Dirichlet distribution as uniform on the n -simplex. This gives the measure studied below. For background and references see [35, 22, 42].

Recent developments put priors on natural subclasses of Markov chains. For example [20, 4] develop and apply priors for reversible Markov chains and [5] develop priors for higher order Markov chains.

It is natural to consider priors on the space of Markov chains with a fixed (known) stationary distribution. This is again a connected convex set. Perhaps the most natural example is the uniform distribution on $\{1, 2, \dots, n\}$. Now the set of transition matrices is the Birkhoff polytope and the uniform distribution is a natural prior. Understanding the uniform distribution for large n leads to the topics in this paper.

Knowing about Birkhoff's Theorem it is also natural to study the prior measure on \mathcal{M}_n resulting from a uniform combination of extreme points. Thus if A_σ is the permutation matrix corresponding to σ and $\{X_\sigma\}$ is a uniform point of the $n!$ -simplex then $M = \sum_{\sigma \in S_n} A_\sigma X_\sigma$ is a uniform combination of extreme points. This distribution was proposed and studied in [37] as a way to put a prior on the parameters of an $n \times n$ -contingency table with known uniform margins. The following result suggests this is a strange distribution, sharply concentrated about the matrix with all entries $1/n$.

Proposition 1.1. *Let $M \in \mathcal{M}_n$ be a uniform convex combination of extreme points. Then*

$$E \sum_{ij} |M_{ij} - \frac{1}{n}| \leq n \sqrt{\frac{n-1}{n!+1}}$$

Proof. The distribution of M_{11} is given by Beta(a, b) distribution with $a = (n-1)!$ and $b = (n-1)(n-1)!$ which has mean $a/(a+b) = 1/n$ and variance $ab/(a+b)^2(a+b+1) = (n-1)/n^2(n!+1)$. Then by the symmetry of the entries

$$E \sum_{ij} |M_{ij} - \frac{1}{n}| = n^2 E |M_{11} - \frac{1}{n}| \leq n^2 \sqrt{\text{Var} M_{11}} \leq n \sqrt{\frac{n-1}{n!+1}}.$$

■

Of course, this prior is absolutely continuous with respect to the uniform distribution and a sufficiently large amount of data will swamp the prior (although this may be prohibitive large when n is large).

A variety of measures on the stochastic matrices were studied in the subject of “random random walks” [27]. This area was initiated with a theorem of Aldous and Diaconis [1]. If an $n \times n$ stochastic matrix is chosen by making the rows uniform on the n -simplex the expected time to stationarity is small, indeed two steps suffice (but one does not). This suggests that this models does not capture the essential features of real Markov chains which are usually “local”. Much of the work thus restricts attention to random walks on finite groups G (see [27] for more details).

Our discussion leaves many points untouched. To generate points from the uniform distribution on \mathcal{M}_n we use a basic “Gibbs sampling algorithm”: pick a pair of distinct rows and a pair of distinct columns at random. These intersect in a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. This is replaced

by $\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$ chosen uniformly on the set of matrices with the same row and column sums as A . This is easy to do choosing a' uniformly from the relevant range. We would like to understand the running time of this algorithm. A host of other algorithms for uniform choice in a compact set is in [2].

The posterior distribution on \mathcal{M}_n after observing the Markov chain of length N is proportional to $\prod_{i,j} x_{ij}^{N(i,j)}$ where $N(i, j)$ is the number of observed transitions from i to j in the run. How do such measures behave? Our work suggests a heuristic: the measures should behave like product Dirichlet distributions. The i th row having density proportional to $\prod_j x_{i,j}^{N(i,j)}$. The known properties of the

Dirichlet distribution now make basic questions accessible. For example, the Bayes estimate of the transition matrix is easy to compute.

1.3. Elements of Random Matrices. The present paper has many points of contact with the ongoing study of the behavior of entries of a uniformly chosen random matrix in one of the classical compact groups O_n or U_n . These problems we originally studied to understand the ‘equivalence of ensembles’ in statistical mechanics. Indeed, the first row of a random matrix in O_n is uniformly distributed on the n -sphere—the micro-canonical ensemble. The entries multiplied by \sqrt{n} are approximately independent standard normal—the canonical ensemble. This is an early theorem of Borel; see [22] for a historical review, sharp statements and pointers to the work of Lévy and others. Later these theorems were extended and used to prove sharp finite forms of de Finetti’s theorems and many extensions [24].

For M chosen uniformly on U_n , the entries multiplied by \sqrt{n} are approximately independent standard complex normal. This has been proved in various sense. For example [31] shows that an $m \times m$ block is close to normal in total variation if $m = o(\sqrt{n})$. For other topologies [29] shows independent normal behaviour persists for $m = o(n/\log n)$. Other global features, such as the maximum entry [30], traces of powers of M [21, 17] and arbitrary linear combinations of the entries [15] behave like normals as well. Of course there are differences. The eigenvalues of a random element of U_n lie on the unit circle while the eigenvalues of independent normals fill out the disk uniformly. For refinements, see [36, 38].

Yuval Peres suggested that these results may have a close connection to the Birkhoff polytope. Let M be uniform in U_n and set $N_{ij} = |M_{ij}|^2$. Then N is doubly stochastic with entries approximately independent and exactly exponentially distributed. While we show in Section 3.1 that these distributions are not the same it seems likely that they share many properties.

Classical results for equivalence of ensembles show equivalence of micro-canonical and canonical ensembles which result from fixing low dimensional sufficient statistics. The results above, and in the present paper, show that equivalences of various sorts persist after conditioning on high dimensional statistics: If $\{E_{ij}\}$ is a matrix of independent exponentials, the conditional distribution given that all the row and column sums are equal to one is uniform on \mathcal{M}_n . More background on equivalence of ensembles can be found in [42] and [32].

1.4. Magic squares and contingency tables. There is a close connection between the Birkhoff polytope \mathcal{M}_n and $MS(n, c)$ the set of $n \times n$ matrices with non-negative integer entries and all row and column sums equal to c . Elements of $MS(n, c)$ are called magic squares in the enumerative literature. It is known that $|MS(n, c)|$ is a polynomial in c of degree $(n-1)^2$. The leading coefficient of this polynomial is a simple multiple of the volume of \mathcal{M}_n [41]. See also [18].

Generalizing, the set of $m \times n$ matrices with non-negative entries and fixed row and column sums is intensively studied both in combinatorics and statistics where they are called contingency tables. It is known that exact enumerations of the size of this set is $\#P$ -complete even when $n = 2$. A host of techniques for approximate counting and random generation have been developed as well as a remarkable collection of asymptotic formulae. See [23] and [6] for surveys.

Questions of the properties of random contingency tables or randomly chosen points in polytopes are closely connected to the problem of estimating the volume of the polytopes. Important recent work by Barvinok and Hartigan has given asymptotic formulas for the number of contingency tables and the volumes of polytopes of such matrices [8, 9, 7] as well as the closely related problem of the

number of graphs with a given degree sequence [10]. A central idea in their analysis is the maximum entropy distribution which for the Birkhoff polytopes corresponds to independent exponentials for the vertices of the matrix. This maximum entropy distribution provides a good approximation to the distribution yielding (after much work) an asymptotic calculation of the volume.

Beyond asymptotic volume calculations Barvinok [6] also asked the question of “what does a random contingency table look like”? In [7] a precise sense was given to the statement that “in many respects a random matrix behaves as a matrix X of independent geometric random variables”, a direction pursued independently in this paper. One result of this equivalence given in [6] is that the sum of large subsets of the entries of such contingency tables are concentrated around their expectation given under the maximum entropy distribution. Barvinok [7] posed the natural question of determining the marginals of the entries of such random matrices. In the case of doubly stochastic matrices we answer this question determining that they are asymptotically independent exponentials.

2. MARGINALS OF UNIFORM DOUBLY STOCHASTIC MATRICES

Let $X = (X_{ij})_{i,j=1,\dots,n}$ be a uniform doubly stochastic matrix, that is chosen uniformly from the Birkhoff polytope. Since the sum of the rows and columns add to 1, it satisfies $2n - 1$ linear constraints and the matrix is determined by the $(n - 1)^2$ entries $(X_{ij})_{i,j=1,\dots,n-1}$. Let $\Gamma : \mathbb{R}^{(n-1)^2} \rightarrow \mathbb{R}^{n^2}$ denote the function

$$\Gamma(X) = \Gamma(X)_{ij} = \begin{cases} X_{ij} & 1 \leq i, j \leq n - 1, \\ 1 - \sum_{k=1}^{n-1} X_{ik} & 1 \leq i \leq n - 1, j = n \\ 1 - \sum_{k=1}^{n-1} X_{kj} & 1 \leq j \leq n - 1, i = n \\ 1 - \sum_{l=1}^{n-1} (1 - \sum_{k=1}^n X_{kl}) & i = j = n. \end{cases}$$

Let $\Phi : \mathbb{R}^{(n-1)^2} \rightarrow \mathbb{R}^{n^2}$ be the projection $X \mapsto (X_{ij})_{1 \leq i,j \leq n-1}$. By an abuse of notation we will also use Γ as a function from \mathbb{R}^{n^2} to itself by $\Gamma(\Phi(X))$. Then the doubly stochastic matrices correspond to the $(n - 1) \times (n - 1)$ -matrices in the set

$$S_n = \left\{ (x_{ij})_{i,j=1,\dots,n-1} \in [0, 1]^{(n-1)^2} : \min_{1 \leq i,j \leq n} x_{ij} - \Gamma(x)_{ij} \geq 0 \right\}.$$

The distribution of $(X_{ij})_{i,j=1,\dots,n-1}$ is given by the uniform distribution on S_n . Let Z_n denote the volume of S_n , that is

$$Z_n = \int_{[0,1]^{(n-1)^2}} I(x \in S_n) dx$$

where I denotes the indicator function. Canfield and McKay [13] showed that asymptotically the volume of the Birkhoff polytope (in units of basic cells of the lattice which is equivalent to our usage) is

$$Z_n = \frac{1}{n^{n-1}} \cdot \frac{1}{(2\pi)^{n-1/2} n^{(n-1)^2}} \exp\left(\frac{1}{3} + n^2 + o(1)\right). \quad (2.1)$$

Also define

$$\mathcal{D}_n = \left\{ (y_{ij})_{i,j=1,\dots,n} \in \mathbb{R}^{n^2} : \Phi\left(\frac{1}{n}y\right) \in S_n, \min_{i,j} (y - \Gamma\left(\frac{1}{n}y\right))_{ij} \geq 0 \right\}.$$

As we observed in the introduction, the uniformly distributed stochastic matrix shares many properties with matrices of independent exponentials so let us define $(Y_{ij})_{1 \leq i,j \leq n}$ as a matrix of iid exponential mean 1 random variables.

Lemma 2.1. *Conditional on $Y \in \mathcal{D}_n$ we have that $\frac{1}{n}(Y_{ij})_{1 \leq i, j \leq n-1}$ is uniform on S_n . Further, for large n we have that,*

$$P(Y \in \mathcal{D}_n) \geq n^{-4n}. \quad (2.2)$$

Proof. Let \mathcal{W} be the product of the intervals $\mathcal{W} = \prod_{1 \leq i, j \leq n} I_{ij}$ where

$$I_{ij} = \begin{cases} [0, \infty) & \text{if } \max\{i, j\} = n \\ \{0\} & \text{o.w.} \end{cases}$$

Then for each fixed $\bar{Y} \in S_n$ the set $\{Y \in \mathcal{D}_n : (\frac{1}{n}Y)_{i,j=1,\dots,n-1} = \bar{Y}\}$ is $n\Gamma(\bar{Y}) + \mathcal{W}$. Since the density of Y depends only on $\sum_{ij} Y_{ij}$ and since $\sum_{ij} n\Gamma(\bar{Y})_{ij} \equiv n^2$ it follows that $\Gamma(\frac{1}{n}Y)$ is uniform on S_n . Now

$$\begin{aligned} P(Y \in \mathcal{D}_n) &= \int_{\mathbb{R}^{n^2}} \exp\left(-\sum_{i=1}^n \sum_{j=1}^n y_{ij}\right) I(Y \in \mathcal{D}_n) dy_{11} \dots dy_{nn} \\ &= \int_{nS_n} \int_{\mathbb{R}^{2n-1}} \exp\left(-\sum_{i=1}^n \sum_{j=1}^n n(\Gamma(\frac{1}{n}Y)_{ij} - [y_{ij} - n(\Gamma(\frac{1}{n}Y)_{ij})])\right) \end{aligned} \quad (2.3)$$

$$\begin{aligned} &\cdot I\left(\min_{i,j} y_{ij} - \Gamma(Y)_{ij} \geq 0\right) dy_{11} \dots dy_{nn} \\ &= \text{Vol}_{n^2}(nS_n) \exp(-n^2) \end{aligned} \quad (2.4)$$

$$\cdot \int_{[0, \infty)^{2n-1}} \int_{\mathbb{R}^{2n-1}} \exp\left(-\sum_{i=1}^n y_{in} - \sum_{j=1}^{n-1} y_{in}\right) dy_{1n} \dots dy_{nn} dy_{n1} \dots dy_{n,n-1} \quad (2.5)$$

$$= \text{Vol}_{n^2}(nS_n) \exp(-n^2). \quad (2.6)$$

Combining equations (2.1), (2.7), (4.6) we have that

$$P(Y \in \mathcal{D}_n) = \frac{1}{n^{n-1}} \cdot \frac{1}{(2\pi)^{n-1/2} n^{(n-1)^2}} \exp\left(\frac{1}{3} + n^2 + o(1)\right) n^{n^2} \exp(-n^2) \geq n^{-4n}, \quad (2.7)$$

for large n . ■

In particular this means for X uniform on \mathcal{M}_n , for any measurable set $\mathcal{B} \subset \mathbb{R}^{(n-1)^2}$, by equation (2.7) we have that

$$P(X \in \mathcal{B}) \leq n^{4n} P(\Phi(Y) \in \mathcal{B}). \quad (2.8)$$

This equation is only meaningful when $P(\Phi(Y) \in \mathcal{B}) \leq n^{-4n}$. However, for a number of important large deviation events we can effectively translate results about Y to results about X . In particular using the exchangeability of the entries of X we can establish the asymptotic marginal distribution of the entries of the X given in Theorem 1.

Proof of Theorem 1. Let \mathcal{A} be a measurable subset of $[0, \infty)$. By the Azuma–Hoeffding inequality

$$P\left(\left|\frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} I(nY_{ij} \in \mathcal{A}) - P(Y_{11} \in \mathcal{A})\right| > \frac{1}{2} n^{-1/2+\epsilon}\right) \leq \exp(-cn^{1+2\epsilon}).$$

Then by equation (2.8) we have that,

$$P\left(\left|\frac{1}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}I(nX_{ij}\in\mathcal{A})-P(Y_{11}\in\mathcal{A})\right|>\frac{1}{2}n^{-1/2+\epsilon}\right)\leq n^{4n}\exp(-cn^2)\leq\exp(-c'n^2)$$

and so since the entries of X are exchangeable,

$$|P(nX_{11}\in\mathcal{A})-P(Y_{11}\in\mathcal{A})|<n^{-1/2+\epsilon}+\exp(-c'n^2).$$

As this holds uniformly over all \mathcal{A} it follows that $d_{\text{tv}}(X_{11}, Y_{11}) < n^{-1/2+\epsilon}$ for large n which establishes the result. \blacksquare

2.1. Marginal distributions of submatrices. In this subsection we go beyond marginal distributions and investigate the asymptotic distribution of sub-arrays of the matrix, in particular showing that for boxes of sidelength almost \sqrt{n} the entries are close to iid exponentials after rescaling.

Fix some $k = k(n) = O(\frac{n^{1/2}}{\log n})$. Define $W^{\ell_1\ell_2} \in \mathbb{R}^{k^2}$ as the $k \times k$ -submatrix of entries of the matrix Y_{ij} for $i \in \{(\ell_1 - 1)k + 1, \dots, \ell_1 k\}$ and $j \in \{(\ell_2 - 1)k + 1, \dots, \ell_2 k\}$, i.e.,

$$W^{\ell_1\ell_2} = \begin{pmatrix} Y_{(\ell_1-1)k+1,(\ell_2-1)k+1} & \cdots & Y_{(\ell_1-1)k+1,\ell_2k} \\ \vdots & \ddots & \vdots \\ Y_{\ell_1k,(\ell_2-1)k+1} & \cdots & Y_{\ell_1k,\ell_2k} \end{pmatrix}.$$

Let $\epsilon > 0$ and let A be a measurable subset of \mathbb{R}^{k^2} . By the Azuma–Hoeffding inequality we have the following large deviations bound.

$$P\left(\left|\left[\frac{n-1}{k}\right]^{-2}\sum_{\ell_1=1}^{\lfloor n-1/k \rfloor}\sum_{\ell_2=1}^{\lfloor n-1/k \rfloor}I(W^{\ell_1\ell_2}\in A)-P(W^{11}\in A)\right|>\frac{1}{2}\epsilon\right)\leq\exp\left(-\frac{\epsilon^2}{8}\left[\frac{n-1}{k}\right]^2\right). \quad (2.9)$$

Now define $V^{\ell_1\ell_2} \in \mathbb{R}^{k^2}$ as the $k \times k$ -submatrix of X_{ij} with $i \in \{(\ell_1 - 1)k + 1, \dots, \ell_1 k\}$ and $j \in \{(\ell_2 - 1)k + 1, \dots, \ell_2 k\}$, i.e.,

$$V^{\ell_1\ell_2} = \begin{pmatrix} X_{(\ell_1-1)k+1,(\ell_2-1)k+1} & \cdots & X_{(\ell_1-1)k+1,\ell_2k} \\ \vdots & \ddots & \vdots \\ X_{\ell_1k,(\ell_2-1)k+1} & \cdots & X_{\ell_1k,\ell_2k} \end{pmatrix}.$$

We now prove Theorem 4 showing that $d_{\text{tv}}(nV^{11}, W^{11})$ converges to 0.

Proof of Theorem 4. By equation (2.8) and (2.9) we have that,

$$P\left(\left|\frac{1}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}I(nV^{ij}\in\mathcal{A})-P(W^{11}\in\mathcal{A})\right|>\frac{1}{2}\epsilon\right)\leq n^{4n}\exp\left(-\frac{\epsilon^2}{8}\left[\frac{n-1}{k}\right]^2\right)=o(1).$$

Since the entries of X are exchangeable this implies that,

$$|P(nV^{11}\in\mathcal{A})-P(W^{11}\in\mathcal{A})|<\frac{1}{2}\epsilon+o(1).$$

As this holds uniformly over all \mathcal{A} it follows that $d_{\text{tv}}(nV^{11}, W^{11}) < \epsilon$ for large n which establishes the result. \blacksquare

3. FURTHER PROPERTIES OF UNIFORM DOUBLY STOCHASTIC MATRICES

In this section we establish further properties of the matrices including convergence of moments and the mixing time of such matrices.

3.1. Non-constant row sums. It will be important to consider the generalized case of $m \times n$ -matrices with fixed but non-constant row and column sums. For a sequence of positive row sums $\{a_i\}_{i=1}^m$ and columns sums $\{b_i\}_{i=1}^n$ where $\sum_{i=1}^m a_i = \sum_{i=1}^n b_i = t$ we define the transportation polytope $\mathbf{p} = \mathbf{p}((a_i), (b_i))$ to be the polytope of $m \times n$ -matrices with nonnegative entries, row sums a_i and column sums b_i . Let $\mathcal{P}_{m,n,t}$ denote the set of all such polytopes and let $\mathbf{p}^* = \mathbf{p}_{m,n,t}^*$ denote the special case of polytopes with constant row sums t/m and column sums t/n . We will let $\text{Vol}_{(m-1)(n-1)}(\mathbf{p})$ denote the volume of the image of the set \mathbf{p} under the map

$$(X_{ij})_{i=1,\dots,m,j=1,\dots,n} \mapsto (X_{ij})_{i=1,\dots,m-1,j=1,\dots,n-1}$$

in $\mathbb{R}^{(m-1)(n-1)}$. The following lemma shows that amongst all $m \times n$ -matrices \mathbf{p}^* has the largest volume.

Lemma 3.1. *We have that*

$$\text{Vol}_{(m-1)(n-1)}(\mathbf{p}_{m,n,t}^*) = \max_{\mathbf{p} \in \mathcal{P}_{m,n,t}} \text{Vol}_{(m-1)(n-1)}(\mathbf{p})$$

Proof. We begin by proving the following simpler claim.

Claim 3.2. *Let $\{a_i\}_{i=1}^m$ be a collection of row sums with $\sum_{i=1}^m a_i = t$ and let $\mathbf{p}(r)$ denote the polytope of $m \times 2$ -matrices with row sums (a_i) and column sums $r, t-r$ for $0 \leq r \leq t$. Then*

$$\text{Vol}_{(m-1)}\mathbf{p}(t/2) = \max_{0 \leq r \leq t} \text{Vol}_{(m-1)}\mathbf{p}(r).$$

Let $X = (X_{ij})_{i=1,\dots,m,j=1,2}$ be chosen uniformly according to $\mathbf{p}(r)$. Let $(Y_i)_{i=1,\dots,m}$ be independent random variables with the uniform distribution $[0, a_i]$. It is easy to verify that $(X_{i1})_{i=1,\dots,m}$ is equal in distribution to $(Y_i)_{i=1,\dots,m}$ conditional on $\sum_{i=1}^m Y_i = r$ and moreover that the volume $\text{Vol}_{(m-1)}\mathbf{p}(r)$ is proportional to the density of $\sum_{i=1}^m Y_i$ at r .

It remains to show that this density is maximized at $t = r/2$. We say a distribution is log-concave if the logarithm of its density is concave. This clearly includes the uniform distribution on an interval. Moreover, the sum of independent random variables with log-concave distributions itself has a log-concave distribution [12]. Since the density of $\sum_{i=1}^m Y_i$ is symmetric about $t/2$ it follows that it is maximized at $t/2$ which completes the claim.

We now complete the proof of Lemma 3.1. Let $\mathbf{p} = \mathbf{p}((a_i), (b_i))$ and $\mathbf{p}' = \mathbf{p}((a_i), (b'_i))$ where $b'_1 = b'_2 = \frac{b_1+b_2}{2}$ and $b'_i = b_i$ for $i \geq 3$. Further define the set

$$\hat{A} = \{(\hat{a}_i)_{i=3}^m : 0 \leq \hat{a}_i \leq a_i, \sum_{i=1}^m \hat{a}_i = t - b_1 - b_2\}$$

which represent possible values for the sum of the entries of the rows of a matrix in \mathbf{p} excluding the first two columns. Then by first conditioning on these sums we have the following integral for the volumes

$$\begin{aligned} & \text{Vol}_{(m-1)(n-1)}\mathbf{p}((a_i), (b_i)) \\ &= \int_{\hat{A}} \text{Vol}_{(m-1)}\mathbf{p}((a_i - \hat{a}_i), (b_i)_{i=1,2}) \text{Vol}_{(m-1)(n-3)}\mathbf{p}((\hat{a}_i), (b_i)_{i=3,\dots,n}) \mu(d(\hat{a}_i)) \end{aligned}$$

where μ is the uniform distribution over \mathcal{A}^* . Similarly

$$\begin{aligned} & \text{Vol}_{(m-1)(n-1)} \mathbf{p}((a_i), (b'_i)) \\ &= \int_{\hat{\mathcal{A}}} \text{Vol}_{(m-1)} \mathbf{p}((a_i - \hat{a}_i), (b'_i)_{i=1,2}) \text{Vol}_{(m-1)(n-3)} \mathbf{p}((\hat{a}_i), (b'_i)_{i=3,\dots,n}) \mu(d(\hat{a}_i)) \end{aligned}$$

Applying Claim 3.2 we, therefore, have that

$$\text{Vol}_{(m-1)(n-1)} \mathbf{p}((a_i), (b_i)) \leq \text{Vol}_{(m-1)(n-1)} \mathbf{p}((a_i), (b'_i))$$

which says that replacing the first two column sums by their average can only increase the volume of the polytope. This is true of course for any pair of columns and similarly for any pair of rows. It is easy to show that the volume of polytopes in $\mathcal{P}_{m,n,t}$ are symmetric and continuous in the row and column sums $(a_i), (b_i)$ and hence it follows that \mathbf{p}^* must be a maxima of the volume. \blacksquare

Canfield and McKay [13] give an asymptotic formula for the volume of matrices with constant row and column sums as

$$\begin{aligned} & \text{Vol}_{(m-1)(n-1)}(\mathbf{p}_{m,n,m}^*) \\ &= \frac{1}{m^{(n-1)/2} n^{(m-1)/2}} \cdot \frac{1}{(2\pi)^{(m+n-1)/2} n^{(m-1)(n-1)}} \exp\left(\frac{1}{3} + mn - \frac{(m-n)^2}{12mn} + o(1)\right). \end{aligned} \quad (3.1)$$

Note that our definition of volume corresponds to their notion of volume in units of basic cells of the lattice induced by \mathbb{Z}^{mn} .

Let $\mathcal{R} = \mathcal{R}_{r,n}$ denote the $r(n-1)$ -dimensional polytope of nonnegative matrices whose rows sum to 1. Let ν_r denote the measure on \mathcal{R} induced by the first r rows of a uniform doubly stochastic $n \times n$ -matrix (X_{ij}) and let μ_r denote uniform probability measure on \mathcal{R} . Equivalently μ_r is the measure induced by the first r rows of a uniform stochastic matrix (one where the rows are independent and conditioned to sum to 1).

Lemma 3.3. *For a fixed integer $r \geq 1$ and $n > r$ the Radon-Nikodym derivative of the measures μ_r and ν_r satisfies*

$$\frac{d\nu_r}{d\mu_r} \leq (1 + o(1))e^{r/2}.$$

as $n \rightarrow \infty$.

Proof. Conditioned on the first r rows of a uniform doubly stochastic $n \times n$ -matrix (X_{ij}) the remainder of the matrix is a uniformly chosen matrix from the polytope of $(n-r) \times n$ -matrices

$$\mathbf{p}\left(1_{n-r}, \left(1 - \sum_{i=1}^r X_{ij}\right)_{j=1,\dots,n}\right)$$

where 1_m represents the vectors of 1's of length m . Since μ_r is the uniform distribution over $\mathcal{R} = \mathcal{R}_{r,n}$ it follows that

$$\frac{d\nu_r}{d\mu_r}(X_{ij}) \propto \text{Vol}_{(n-r-1)(n-1)} \mathbf{p}\left(1_{n-r}, \left(1 - \sum_{i=1}^r X_{ij}\right)_{j=1,\dots,n}\right)$$

where \propto denote proportionality. To determine the constant of proportionality note that

$$Z_n = \text{Vol}_{r(n-1)}(\mathcal{R}) \int_{\mathcal{R}} \text{Vol}_{(n-r-1)(n-1)} \mathbf{p}\left(1_{n-r}, \left(1 - \sum_{i=1}^r X_{ij}\right)_{j=1,\dots,n}\right) \mu_r(d(X_{ij}))$$

recalling that Z_n is the volume of the Birkhoff polytope. It follows that

$$\begin{aligned} \frac{d\nu_r}{d\mu_r}(X_{ij}) &= \frac{1}{Z_n} \text{Vol}_{r(n-1)}(\mathcal{R}) \text{Vol}_{(n-r-1)(n-1)} \mathfrak{p}\left(1_{n-r}, \left(1 - \sum_{i=1}^r X_{ij}\right)_{j=1, \dots, n}\right) \\ &\leq \frac{1}{Z_n} \text{Vol}_{r(n-1)}(\mathcal{R}) \text{Vol}_{(n-r-1)(n-1)} \mathfrak{p}_{m,n,m}^* \end{aligned}$$

by Lemma 3.1. Hence substituting the formulas for the volumes of the polytopes and applying Stirling's formula we have that

$$\begin{aligned} \frac{d\nu_r}{d\mu_r}(X_{ij}) &\leq (1 + o(1)) \frac{n^{n-1}}{(n-r)^{(n-1)/2} n^{(n-r-1)/2}} \cdot \frac{1}{((n-1)!)^r} \cdot \frac{(2\pi)^{n-1/2} n^{(n-1)^2} e^{-rn}}{(2\pi)^{(2n-r-1)/2} n^{(n-r-1)(n-1)}} \\ &= (1 + o(1)) n^{r/2} e^{r/2} \cdot \frac{n^r}{(\sqrt{2\pi n} n^n e^{-n})^r} \cdot (2\pi)^{r/2} n^{r(n-1)} e^{-rn} \\ &= (1 + o(1)) e^{r/2} \end{aligned}$$

which completes the proof. \blacksquare

This proof also shows that the uniformly distributed stochastic matrix is not given exactly by the square of the absolute value of a random unitary matrix. In such a random matrix the rows are distribution according to μ_1 while we have that

$$\frac{d\nu_1}{d\mu_1}(\frac{1}{n} \mathbf{1}_n) = \frac{1}{Z_n} \text{Vol}_{r(n-1)}(\mathcal{R}) \text{Vol}_{(n-r-1)(n-1)} \mathfrak{p}_{m,n,m}^* = (1 + o(1)) e^{r/2}.$$

Hence at least for large n the models are not the same (in the trivial case of $n = 2$ they are equal).

3.2. Convergence of Moments. Using Lemma 3.3 we may now establish convergence of the moments of the entries of a doubly stochastic matrix to those of independent exponentials. We will let (V_k) be a sequence of iid exponentially distributed mean 1 random variables.

Lemma 3.4. *Let $(i_1, j_1), \dots, (i_L, j_L)$ be a fixed sequence of pairs of positive integers and $\alpha_1, \dots, \alpha_L$ be fixed a sequence of positive integers. Then if $(X_{ij})_{i,j=1, \dots, n}$ are distributed as a uniform doubly stochastic matrix then*

$$E \prod_{k=1}^L (nX_{i_k, j_k})^{\alpha_k} \rightarrow E \prod_{k=1}^L V_k^{\alpha_k}.$$

Proof. By Theorem 4 the joint distribution of the $(nX_{i_k, j_k})_{k=1, \dots, L}$ converges to iid exponential random variables. It follows that

$$E \left[\prod_{k=1}^L (nX_{i_k, j_k})^{\alpha_k} I(\max_{1 \leq k \leq L} nX_{i_k, j_k} < M) \right] \rightarrow E \left[\prod_{k=1}^L V_k^{\alpha_k} I(\max_{1 \leq k \leq L} V_k < M) \right],$$

and hence we can complete the proof by showing that

$$\lim_{M \rightarrow \infty} \limsup_n E \left[\prod_{k=1}^L (nX_{i_k, j_k})^{\alpha_k} I(\max_{1 \leq k \leq L} nX_{i_k, j_k} \geq M) \right] \rightarrow 0. \quad (3.2)$$

By the exchangeability of X we may assume without loss of generality that $\max_{1 \leq k \leq L} i_k \leq L$ and that $\max_{1 \leq k \leq L} j_k \leq L$. In particular this assumption implies that each of the entries X_{i_k, j_k} appear in the first L rows of the matrix. Let \tilde{Y}_{ij} denote a uniform stochastic matrix, that is one whose rows are independent and chosen according to μ_1 .

Now by Lemma 3.3 it follows that

$$\begin{aligned} & E \left[\prod_{k=1}^L (nX_{i_k, j_k})^{\alpha_k} I(\max_{1 \leq k \leq L} nX_{i_k, j_k} \geq M) \right] \\ & \leq (e^{L/2} + o(1)) E \left[\prod_{k=1}^L (n\tilde{Y}_{i_k, j_k})^{\alpha_k} I(\max_{1 \leq k \leq L} n\tilde{Y}_{i_k, j_k} \geq M) \right] \end{aligned}$$

and hence it is sufficient to establish equation (3.2) replacing the X_{i_k, j_k} with \tilde{Y}_{i_k, j_k} . Now the Y_{i_k, j_k} are given by Beta distributions $B(1, n-1)$. It follows that

$$E\tilde{Y}_{i_k, j_k}^{\alpha_k} = \prod_{\ell=1}^{\alpha_k} \frac{\ell}{n-1+\ell} = (1+o(1))\alpha_k!n^{-\alpha_k} \quad (3.3)$$

By the power mean inequality and the fact that $E|\tilde{Y}|^\alpha I(Y > M) \leq M^{-1}E|\tilde{Y}|^{\alpha+1}$

$$\begin{aligned} & E \prod_{k=1}^L (n\tilde{Y}_{i_k, j_k})^{\alpha_k} I(\max_{1 \leq k \leq L} n\tilde{Y}_{i_k, j_k} \geq M) \\ & \leq E \frac{1}{\sum_{k=1}^L \alpha_k} \sum_{k=1}^L \alpha_k (n\tilde{Y}_{i_k, j_k})^{\sum_{k=1}^L \alpha_k} I(\max_{1 \leq k \leq L} n\tilde{Y}_{i_k, j_k} \geq M) \\ & \leq M^{-1} E \frac{1}{\sum_{k=1}^L \alpha_k} \sum_{k=1}^L \alpha_k (n\tilde{Y}_{i_k, j_k})^{1+\sum_{k=1}^L \alpha_k} \end{aligned}$$

and hence by equation (3.3),

$$\lim_{M \rightarrow \infty} \sup_n E \prod_{k=1}^L (nY_{i_k, j_k})^{\alpha_k} I(\max_{1 \leq k \leq L} nY_{i_k, j_k} \geq M) = 0$$

which completes the proof. \blacksquare

We may also examine the maximal element of the matrix. For an $n \times n$ -matrix of iid exponential random variables with mean 1 the maximum entry is at most $(2+o(1))\log n$ with high probability and we show that this is also the case for the renormalized uniform doubly stochastic matrix.

Proof of Theorem 2. By Lemma 3.3 we have that

$$P(nX_{11} > (2+\epsilon)\log n) \leq (e^{1/2} + o(1))P(nY_{11} > (2+\epsilon)\log n)$$

Now since Y_{11} has $B(1, n-1)$ distribution

$$\begin{aligned} P(nY_{11} > (2+\epsilon)\log n) &= (n-1) \int_{\frac{(2+\epsilon)\log n}{n}}^1 (1-y)^{n-2} \\ &= \left(1 - \frac{(2+\epsilon)\log n}{n}\right)^{n-1} \\ &= (1+o(1))n^{-2-\epsilon}. \end{aligned} \quad (3.4)$$

The exchangeability of the entries and a union bound completes the proof. \blacksquare

3.3. Mixing Time. As uniformly distributed stochastic matrices correspond to the transition matrices of Markov chains one can ask about the mixing time of such matrices.

Proof of Theorem 5. By Lemma 3.3 the mixing time cannot be 1 since it implies that the rows of the matrix are not close to being constant. We show at time 2, however, they are almost constant. Let $X_{ij}^{(2)}$ denote the ij -th entry of the matrix X^2 . The total variation distance from stationarity of the Markov chain at time 2 is given by

$$\max_i \frac{1}{2} \sum_{j=1}^n \left| \frac{1}{n} - X_{ij}^{(2)} \right|$$

which is equal to

$$\max_i \sum_{j=1}^n \max\left\{ \frac{1}{n} - X_{ij}^{(2)}, 0 \right\}.$$

Since the rows are exchangeable, by taking a union bound it is sufficient to show that for each $\epsilon > 0$,

$$P \left(\sum_{j=1}^n \max\left\{ \frac{1}{n} - X_{1j}^{(2)}, 0 \right\} > \epsilon \right) = o(1/n).$$

We will again work first in the independent entries model (Y_{ij}) . Let \mathcal{F} denote the σ -algebra generated by $(Y_{1j})_{j=1, \dots, n-1}$ and let \mathcal{H} denote the event

$$\left\{ \max_{1 \leq j \leq n-1} Y_{1j} \leq 3 \log n \right\} \cap \left\{ \sum_{j=1}^{n-1} Y_{1j} \leq n - 3 \log n \right\}.$$

The sums $\sum_{k=2}^{n-1} Y_{1k} Y_{kj}$ are conditionally independent given \mathcal{F} . Further for $\delta, \lambda > 0$,

$$\begin{aligned} P \left(\frac{1}{n} \sum_{k=2}^{n-1} Y_{1k} Y_{kj} < 1 - \delta \text{ and } \mathcal{H} \mid \mathcal{F} \right) &= P \left(\frac{1}{n} \sum_{k=2}^{n-1} Y_{1k} (1 - Y_{kj}) > \delta - 6 \log n \text{ and } \mathcal{H} \mid \mathcal{F} \right) \\ &= P \left(\exp \left(\frac{\lambda}{n} \sum_{k=2}^{n-1} Y_{1k} (1 - Y_{kj}) \right) > \exp \left(\frac{\lambda}{n} (\delta - 6 \log n) \right) \text{ and } \mathcal{H} \mid \mathcal{F} \right) \end{aligned}$$

Now if $Y_{1k} \leq \frac{3}{n} \log n$ and $\lambda = \frac{n}{\log^3 n}$ then by Taylor series for large n and $1 \leq j \leq n-1$,

$$E \left[\exp \left(\frac{\lambda}{n} Y_{1k} (1 - Y_{kj}) \right) \mid Y_{1k} \right] = \frac{\exp \left(\frac{\lambda}{n} Y_{1k} \right)}{1 + \frac{\lambda}{n} Y_{1k}} \leq \frac{1 + \frac{\lambda}{n} Y_{1k} + \left(\frac{\lambda}{n} Y_{1k} \right)^2}{1 + \frac{\lambda}{n} Y_{1k}} \leq \exp \left(\left(\frac{\lambda}{n} Y_{1k} \right)^2 \right).$$

Hence by Markov's inequality for large n ,

$$P \left(\frac{1}{n} \sum_{k=2}^{n-1} Y_{1k} Y_{kj} < 1 - \delta \text{ and } \mathcal{H} \mid \mathcal{F} \right) \leq \frac{\exp \left(\frac{9n}{\log^4 n} \right)}{\exp \left(\frac{n(\delta - \frac{6}{n} \log n)}{\log^3 n} \right)} \leq \exp \left(-\frac{n\delta}{2 \log^3 n} \right)$$

with room to spare. By the conditional independence of the sums we have that

$$P \left(\# \left\{ 1 \leq j \leq n-1 : \frac{1}{n} \sum_{k=2}^{n-1} Y_{1k} Y_{kj} < 1 - \delta \right\} > \delta n \text{ and } \mathcal{H} \mid \mathcal{F} \right) \leq \binom{n}{n\delta} \exp \left(-\frac{n^2 \delta^2}{2 \log^3 n} \right). \quad (3.5)$$

This implies that

$$P\left(\sum_{j=1}^{n-1} \max\left\{\frac{1}{n} - \sum_{k=2}^{n-1} \frac{Y_{1k} Y_{kj}}{n}, 0\right\} > 2\delta \text{ and } \mathcal{H}\right) \leq \binom{n}{n\delta} \exp\left(-\frac{n^2\delta^2}{2\log^3 n}\right).$$

We can now return to the doubly stochastic matrix setting. By equation (2.8) we have that

$$P\left(\sum_{j=1}^{n-1} \max\left\{\frac{1}{n} - \sum_{k=2}^{n-1} X_{1k} X_{kj}, 0\right\} > 2\delta, \max_{1 \leq j \leq n} X_{1n} \geq \frac{3 \log n}{n}\right) \leq n^{4n} \binom{n}{n\delta} \exp\left(-\frac{n^2\delta^2}{2\log^3 n}\right).$$

and hence since $X_{1j}^{(2)} = \sum_{k=1}^n X_{1k} X_{kj} \geq \sum_{k=2}^{n-1} X_{1k} X_{kj}$ and so

$$P\left(\sum_{j=1}^n \max\left\{\frac{1}{n} - X_{1j}^{(2)}, 0\right\} > 2\delta + \frac{1}{n}, \max_{1 \leq j \leq n} X_{1n} \geq \frac{3 \log n}{n}\right) = o(1/n).$$

By equation (3.4) we have that

$$P\left(\max_{1 \leq j \leq n} X_{1n} \geq \frac{3 \log n}{n}\right) = O(n^{-2})$$

so it follows that

$$P\left(\sum_{j=1}^n \max\left\{\frac{1}{n} - X_{1j}^{(2)}, 0\right\} > 2\delta + \frac{1}{n}\right) = o(1/n)$$

for any $\delta > 0$. Letting δ go to 0 completes the proof. \blacksquare

4. SINGULAR VALUES

In this section we give the proof of Theorem 3. Let $0 \leq \sigma_1^n \leq \dots \leq \sigma_n^n$ denote the singular values of $n^{1/2}(X - EX)$. These correspond to the square roots of the eigenvalues of the matrix $n(X - EX)(X - EX)^*$ which is a Hermitian matrix. For a Hermitian matrix A let $\lambda_1(A) \leq \dots \leq \lambda_n(A)$ denote its eigenvalues and let $\hat{\mu}(A) = \sum_{i=1}^n \delta_{\lambda_i(A)}$ denote the empirical spectrum of A .

Let $(\tilde{Y}_{ij})_{i,j=1,\dots,n}$ denote the $n \times n$ -matrix with i.i.d. entries supported in $[0, K]$ and consider the Wishart Matrix $\Xi_n = n^{-1}(\tilde{Y} - E\tilde{Y})(\tilde{Y} - E\tilde{Y})^*$ which is Hermitian and hence has real eigenvalues. Marčenko and Pastur [34] showed that $\hat{\mu}(\Xi_n) \rightarrow \mu'$ weakly in probability as $n \rightarrow \infty$ where μ' is the distribution on $[0, 2]$ with density $\frac{\sqrt{x(4-x)}}{2\pi x}$.

As with our previous results we use large deviation results on random matrices to transfer results to uniform doubly stochastic matrices. In this case we use results of Guionnet and Zeitouni [26] who establish concentration of measure results for the spectrum of large Wishart matrices. In Corollary 1.8 and the remarks that follow they show that for any $\epsilon > 0$ there exists $c(\epsilon) > 0$ such that for large n and $K > 1$,

$$P(d_W(\hat{\mu}(\Xi_n), E\hat{\mu}(\Xi_n)) > \epsilon) \leq \exp(-cK^{-2}n^2). \quad (4.1)$$

where d_W denotes the Wasserstein distance. We will take the entries of \tilde{Y} to have density given by

$$\rho_n(x) = \begin{cases} \frac{1}{1-n^{10}} e^{-x} & x \in [0, 10 \log n], \\ 0 & \text{o.w.} \end{cases} \quad (4.2)$$

That is the entries are mean 1 exponentials conditioned to be less than $10 \log n$ and so it follows that

$$P(d_W(\hat{\mu}(\Xi_n), E\hat{\mu}(\Xi_n)) > \epsilon) \leq \exp(-c'n^2 \log^{-2}). \quad (4.3)$$

Now let

$$\tilde{S}_n = \left\{ (x_{ij})_{i,j=1,\dots,n-1} \in S_n : \max_{1 \leq i,j \leq n} \Gamma(x)_{ij} \leq \frac{6}{n} \log n \right\}$$

which corresponds to the doubly stochastic matrices whose maximum entry is at most $\frac{6}{n} \log n$. Also define

$$\tilde{\mathcal{D}}_n = \left\{ (x_{ij})_{i,j=1,\dots,n} \in [0, 8 \log n]^{n^2} : \frac{1}{n}(x_{ij})_{i,j=1,\dots,n-1} \in \tilde{S}_n, \forall 1 \leq i, j \leq n, 0 \leq (x - \Gamma(x))_{ij} \leq n^{-4} \right\}.$$

The following lemma is the analogue of Lemma 2.1 for \tilde{Y} .

Lemma 4.1. *With \tilde{Y} as above with marginals given by (4.2), conditional on $\tilde{Y} \in \tilde{\mathcal{D}}_n$ we have that $\Gamma(\frac{1}{n}Y)$ is uniform on \tilde{S}_n . Further, for large n we have that,*

$$P(\tilde{Y} \in \tilde{\mathcal{D}}_n) \geq n^{-8n}. \quad (4.4)$$

Proof. Let \mathcal{W} be the product of the intervals $\mathcal{W} = \prod_{1 \leq i,j \leq n} I_{ij}$ where

$$I_{ij} = \begin{cases} [0, n^{-4}] & \text{if } \max\{i, j\} = n \\ \{0\} & \text{o.w.} \end{cases}$$

Then for each fixed $\bar{Y} \in \tilde{S}_n$ the set $\{\tilde{Y} \in \tilde{\mathcal{D}}_n : \Phi(\frac{1}{n}\tilde{Y}) = \bar{Y}\}$ is $n\Gamma(\bar{Y}) + \mathcal{W}$. Since the density of \tilde{Y} depends only on $\sum_{ij} \tilde{Y}_{ij}$ and since $\sum_{ij} n\Gamma(\bar{Y})_{ij} \equiv n^2$ it follows that $\Gamma(\frac{1}{n}\tilde{Y})$ is uniform on \tilde{S}_n .

Now

$$\begin{aligned} P(\tilde{Y} \in \tilde{\mathcal{D}}_n) &= (1 - n^{-10})^{-n^2} \int_{\tilde{\mathcal{D}}_n} \exp\left(-\sum_{i=1}^n \sum_{j=1}^n y_{ij}\right) dy_{11} \dots dy_{nn} \\ &= (1 + o(1)) \exp(-n^2) n^{n^2} \text{Vol}_{n^2}(\mathcal{D}_n) \end{aligned} \quad (4.5)$$

as for all $\tilde{Y} \in \tilde{\mathcal{D}}_n$ we have that

$$n^2 \leq \sum_{i=1}^n \sum_{j=1}^n \tilde{Y}_{ij} \leq n^2 + (2n+1)n^{-4}.$$

The volume of \mathcal{W} is clearly $n^{-4(2n-1)}$ so we have that

$$\text{Vol}_{n^2}(\tilde{\mathcal{D}}_n) = \text{Vol}_{(n-1)^2}(\tilde{S}_n) n^{-4(2n-1)}.$$

Now interpreting \tilde{S}_n as a subset of S_n it corresponds to the set of doubly stochastic matrices whose maximum entry is at most $6 \log n$. Hence by Theorem 2 we have that

$$\frac{\text{Vol}_{(n-1)^2}(\tilde{S}_n)}{\text{Vol}_{(n-1)^2}(S_n)} = P(\max_{ij} X_{ij} \leq 6 \log n) = 1 - o(1). \quad (4.6)$$

Combining equations (2.1), (2.7), (4.6) we have that

$$\begin{aligned} P(\tilde{Y} \in \tilde{\mathcal{D}}_n) &= (1 + o(1)) \frac{\exp(-n^2) n^{n^2} n^{-4(2n-1)}}{n^{n-1} (2\pi)^{n-1/2} n^{(n-1)^2}} \exp\left(\frac{1}{3} + n^2\right) \\ &\geq n^{-8n} \end{aligned} \quad (4.7)$$

for large n . ■

Now the Courant-Fischer Minimax Theorem says that for an $n \times n$ Hermitian matrix X the k -th eigenvalue of X is given by

$$\lambda_k(X) = \min_{U: \dim(U)=k} \max_{x \in U} \frac{x^* X x}{x^* x}$$

where the minimum is over all k -dimensional subspaces of \mathbb{R}^n . It follows that for Hermitian matrices X, Y that

$$|\lambda_k(X) - \lambda_k(Y)| \leq \|X - Y\|_{\text{op}} \leq n \max_{i,j} |X_{ij} - Y_{ij}|$$

where $\|\cdot\|_{\text{op}}$ is the operator norm (see e.g. [28]). For $\bar{Y} \in \tilde{\mathcal{S}}_n$ and $\tilde{Y} \in \tilde{\mathcal{D}}_n$ such that $\Gamma(\frac{1}{n}\tilde{Y}) = \Gamma(\bar{Y})$ we compare the eigenvalues of the matrices

$$\begin{aligned} A &= n(\Gamma(\bar{Y}) - n^{-1}\mathbf{1})(\Gamma(\bar{Y}) - \frac{1}{n}\mathbf{1})^* \\ B &= n^{-1}(\tilde{Y} - y\mathbf{1})(\tilde{Y} - y\mathbf{1})^* \end{aligned}$$

where $y = \frac{1-(1+10\log n)n^{-10}}{1-n^{-10}} = E\tilde{Y}_{11}$ and $\mathbf{1}$ is the $n \times n$ -matrix of all 1's. By the above bound we have that for $1 \leq k \leq n$,

$$|\lambda_k(A) - \lambda_k(B)| \leq n \max_{i,j} |A_{ij} - B_{ij}| \quad (4.8)$$

Breaking $A - B$ into parts we first have that

$$\begin{aligned} \sup_{i,j} \left| \left(n\Gamma(\bar{Y})^2 - n^{-1}\tilde{Y}^2 \right)_{ij} \right| &= \sup_{i,j} n^{-1} \left| \left(2(n\Gamma(\bar{Y}))(\tilde{Y} - n\Gamma(\bar{Y})) + (\tilde{Y} - n\Gamma(\bar{Y}))^2 \right)_{ij} \right| \\ &= O(n^{-3}) \end{aligned} \quad (4.9)$$

since $\max_{i,j} (n\Gamma(\bar{Y}))_{ij} \leq 6 \log n$ and $\max_{i,j} |(\tilde{Y} - n\Gamma(\bar{Y}))_{ij}| \leq n^{-4}$. Also

$$\begin{aligned} \sup_{i,j} \left| \left(n\Gamma(\bar{Y}) \cdot n^{-1}\mathbf{1} - n^{-1}\tilde{Y} \cdot y\mathbf{1} \right)_{ij} \right| &= \sup_{i,j} \left| \left((\Gamma(\bar{Y}) - n^{-1}y\tilde{Y}) \mathbf{1} \right)_{ij} \right| \\ &= \sup_{i,j} \left| \left((\Gamma(\bar{Y}) - n^{-1}\tilde{Y}) \mathbf{1} \right)_{ij} \right| + O(n^{-10}) \\ &= O(n^{-4}) \end{aligned} \quad (4.10)$$

since $1 - y = O(n^{-10})$. Finally we have that

$$\sup_{i,j} \left| \left(n^{-1}\mathbf{1} - n^{-1}y^2\mathbf{1} \right)_{ij} \right| = O(n^{-10}) \quad (4.11)$$

since $1 - y^2 = O(n^{-10})$. Combining (4.12), (4.9), (4.10) and (4.11) it follows that

$$|\lambda_k(A) - \lambda_k(B)| \leq O(n^{-2}). \quad (4.12)$$

In particular we have that for large n if $d_W(\hat{\mu}(A), \hat{\mu}(B)) = o(1)$ uniformly in \bar{Y} and \tilde{Y} . With Ξ_n defined above and X a uniform doubly stochastic matrix by Lemma 4.1 we have that for any $\epsilon > 0$ and large enough n that

$$\begin{aligned} &P \left(d_W(\hat{\mu}(n(X - EX)(X - EX)^*), E\hat{\mu}(\Xi_n)) > 2\epsilon \mid \Phi(X) \in \tilde{\mathcal{S}}_n \right) \\ &\leq P \left(d_W(\hat{\mu}(\Xi_n), E\hat{\mu}(\Xi_n)) > \epsilon \mid \tilde{Y} \in \tilde{\mathcal{D}}_n \right) \\ &\leq P(d_W(\hat{\mu}(\Xi_n), E\hat{\mu}(\Xi_n)) > \epsilon) P(\tilde{Y} \in \tilde{\mathcal{D}}_n)^{-1} \\ &\leq n^{8n} \exp(-c'n^2 \log^{-2}) = o(1) \end{aligned} \quad (4.13)$$

where the final inequality follows from Lemma 4.1 and equation (4.3). Now by Theorem 2,

$$P\left(\Phi(X) \in \tilde{S}_n\right) \rightarrow 1$$

so

$$P(d_W(\hat{\mu}(n(X - EX)(X - EX)^*), E\hat{\mu}(\Xi_n)) > 2\epsilon) \rightarrow 0$$

as $n \rightarrow \infty$. As $E\hat{\mu}(\Xi_n) \rightarrow \mu'$ (see e.g. [34, 3]) it follows that

$$\hat{\mu}(n(X - EX)(X - EX)^*) \rightarrow \mu'$$

weakly in probability as $n \rightarrow \infty$. Since the singular values of $n^{1/2}(X - EX)$ are the positive square roots of the eigenvalues of $n^{1/2}(X - EX)(X - EX)^*$ and the map $x \mapsto x^2$ maps μ to μ' this completes the proof of Theorem 3.

REFERENCES

- [1] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly*, 93:333–348, 1986.
- [2] Hans C. Andersen and Persi Diaconis. Hit and run as a unifying device. *J. Soc. Fr. Stat. & Rev. Stat. Appl.*, 148:5–28, 2007.
- [3] G. Anderson, A. Guionnet, and O. Zeitouni. An introduction to random matrices. *Cambridge University Press*, 2009.
- [4] S. Bacallado, J.D. Chodera, and V. Pande. Bayesian comparison of Markov models of molecular dynamics with detailed balance constraint. *The Journal of chemical physics*, 131:045106, 2009.
- [5] S. Bacallado and V. Pande. Bayesian analysis of higher order reversible Markov chains. *preprint*, 2010.
- [6] A. Barvinok. What does a random contingency table look like? Available at <http://arxiv.org/abs/0806.3910>, 2009.
- [7] A. Barvinok. Matrices with prescribed row and column sums. Available at <http://arxiv.org/abs/1010.5706>, 2010.
- [8] A. Barvinok and JA Hartigan. An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums. Available at <http://arxiv.org/abs/0910.2477>, 2009.
- [9] A. Barvinok and JA Hartigan. Maximum entropy Gaussian approximation for the number of integer points and volumes of polytopes. Available at <http://arxiv.org/abs/0903.5223>, 2009.
- [10] A. Barvinok and JA Hartigan. The number of graphs and a random graph with a given degree sequence. Available at <http://arxiv.org/abs/1003.0356>, 2010.
- [11] L.J. Billera and A. Sarangarajan. All 0–1 polytopes are traveling salesman polytopes. *Combinatorica*, 16:175–188, 1996.
- [12] S.P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge Univ Pr, 2004.
- [13] E.R. Canfield and B.D. McKay. The asymptotic volume of the Birkhoff polytope. *preprint arXiv*, 2007.
- [14] A.B. Cruse. A note on symmetric doubly-stochastic matrices. *Discrete Mathematics*, 13:109–119, 1975.
- [15] A. D’Aristotile, P. Diaconis, and C.M. Newman. Brownian motion and the classical groups. *Probability, Statistics and their applications: Papers in Honor of Rabii Bhattacharaya. Lecture Notes-Monograph Series*, 41:97–116, 2003.
- [16] J.A. De Loera, F. Liu, and R. Yoshida. A generating function for all semi-magic squares and the volume of the Birkhoff polytope. *Journal of Algebraic Combinatorics*, 30:113–139, 2009.
- [17] P. Diaconis and S.N. Evans. Linear functionals of eigenvalues of random matrices. *Transactions of the American Mathematical Society*, pages 2615–2633, 2001.
- [18] P. Diaconis and A. Gamburd. Random matrices, magic squares and matching polynomials. *Electronic Journal of Combinatorics*, 11:R2, 2004.
- [19] P. Diaconis and P. Matchett-Wood. Random doubly stochastic tridiagonal matrices. *preprint*, 2010.
- [20] P. Diaconis and S.W.W. Rolles. Bayesian analysis for reversible Markov chains. *The Annals of Statistics*, 34:1270–1292, 2006.
- [21] P. Diaconis and M. Shahshahani. On the eigenvalues of random matrices. *Journal of Applied Probability*, 31:49–62, 1994.

- [22] Persi Diaconis and David Freedman. A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.*, 23:397–423, 1987.
- [23] Persi Diaconis and Anil Gangolli. Rectangular arrays with fixed margins. In *Discrete probability and algorithms (Minneapolis, MN, 1993)*, volume 72 of *IMA Vol. Math. Appl.*, pages 15–41. Springer, New York, 1995.
- [24] Persi W. Diaconis, Morris L. Eaton, and Steffen L. Lauritzen. Finite de Finetti theorems in linear models and multivariate analysis. *Scand. J. Statist.*, 19:289–315, 1992.
- [25] VA Emelichev, MM Kovalev, and MK Kravtsov. Polytopes, graphs and optimization. *Cambridge University Press, New York*, 1984.
- [26] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136, 2000.
- [27] Martin Hildebrand. A survey of results on random random walks on finite groups. *Probab. Surv.*, 2:33–63, 2005.
- [28] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge Univ. Pr., 1990.
- [29] T. Jiang. The Entries of Haar-invariant Matrices from the Classical Compact Groups. *Journal of Theoretical Probability*, pages 1–17.
- [30] T. Jiang. Maxima of entries of Haar distributed matrices. *Probability Theory and Related Fields*, 131:121–144, 2005.
- [31] T. Jiang. How many entries of a typical orthogonal matrix can be approximated by independent normals? *The Annals of Probability*, 34:1497–1529, 2006.
- [32] O. Lanford. Entropy and equilibrium states in classical statistical mechanics. *Statistical mechanics and mathematical problems*, pages 1–113, 1973.
- [33] L. Lovász and M.D. Plummer. *Matching theory*. Elsevier Science Ltd, 1986.
- [34] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72:507–536, 1967.
- [35] James John Martin. *Bayesian decision problems and Markov chains*. Robert E. Krieger Publishing Co., Huntington, N.Y., 1975. Reprint of the 1967 edition.
- [36] Elizabeth Meckes. Linear functions on the classical matrix groups. *Trans. Amer. Math. Soc.*, 360:5355–5366, 2008.
- [37] E. Melilli and G. Petris. Bayesian inference for contingency tables with given marginals. *Statistical Methods and Applications*, 4:215–233, 1995.
- [38] F. Mezzadri. Howto Generate Random Matrices from the Classical Compact Groups. *Notices of the AMS*, 54:592–604, 2007.
- [39] I. Pak. Four questions on Birkhoff polytope. *Annals of Combinatorics*, 4:83–90, 2000.
- [40] VN Sačkov. On extremal points of the space of symmetric stochastic matrices. *Sbornik: Mathematics*, 25:419–428, 1975.
- [41] Richard P. Stanley. *Enumerative combinatorics. Vol. 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.
- [42] S. L. Zabell. Characterizing Markov exchangeable sequences. *J. Theoret. Probab.*, 8:175–178, 1995.

COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012.

E-mail address: `sourav@cims.nyu.edu`

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, STANFORD, CA 94305

E-mail address: `diaconis@math.stanford.edu`

THEORY GROUP, MICROSOFT RESEARCH, ONE MICROSOFT WAY, REDMOND, WA 98052

E-mail address: `allansly@microsoft.com`