Random bistochastic matrices

Valerio Cappellini[1], Hans-Jürgen Sommers[2], Wojciech Bruzda[3] and Karol Życzkowski[3,4]

[1] *"Mark Kac" Complex Systems Research Centre, Uniwersytet Jagielloński, ul. Reymonta 4, 30-059 Kraków, Poland*
[2] *Fachbereich Physik, Universität Duisburg-Essen, Campus Duisburg, 47048 Duisburg, Germany*
[3] *Instytut Fizyki im. Smoluchowskiego, Uniwersytet Jagielloński, ul. Reymonta 4, 30-059 Kraków, Poland*
[4] *Centrum Fizyki Teoretycznej, Polska Akademia Nauk, Al. Lotników 32/44, 02-668 Warszawa, Poland*

Abstract

Ensembles of random stochastic and bistochastic matrices are investigated. While all columns of a random stochastic matrix can be chosen independently, the rows and columns of a bistochastic matrix have to be correlated. We evaluate the probability measure induced into the Birkhoff polytope of bistochastic matrices by applying the Sinkhorn algorithm to a given ensemble of random stochastic matrices. For matrices of order $N = 2$ we derive explicit formulae for the probability distributions induced by random stochastic matrices with columns distributed according to the Dirichlet distribution. For arbitrary N we construct an initial ensemble of stochastic matrices which allows one to generate random bistochastic matrices according to a distribution locally flat at the center of the Birkhoff polytope. The value of the probability density at this point enables us to obtain an estimation of the volume of the Birkhoff polytope, consistent with recent asymptotic results.

e-mails: valerio@ictp.it   h.j.sommers@uni-due.de   w.bruzda@uj.edu.pl   karol@cft.edu.pl

1   Introduction

A *stochastic* matrix M is defined as a square matrix of size N, consisting of non–negative elements, such that the sum in each column is equal to unity. Such matrices provide an important tool often applied in various fields of theoretical physics, since they represent Markov chains. In other words, any stochastic matrix maps the set of probability vectors into itself. Weak positivity of each element of M guarantees that the image vector $p' = Mp$ does not contain any negative components, while the probability is preserved due to the normalization of each column of M.

A stochastic matrix B is called *bistochastic* (or *doubly stochastic*) if additionally each of its rows sums up to unity, so that the map preserves identity and for this reason it is given the name *unital*. Bistochastic matrices are used in the theory of *majorization* [1–3] and emerge in several physical problems [4]. For instance they may represent a transfer process at an oriented graph consisting of N nodes.

The set $\mathcal{B}_N$ of bistochastic matrices of size N can be viewed as a convex polyhedron in $\mathbb{R}^{(N-1)^2}$. Due to the Birkhoff theorem, any bistochastic matrix can be represented as a convex combination of permutation matrices. This $(N-1)^2$ dimensional set is often called *Birkhoff polytope*. Its volume with respect to the Euclidean measure is known [5–7] for $2 \leqslant N \leqslant 10$.

To generate a random stochastic matrix one may take an arbitrary square matrix with non-negative elements and renormalize each of its columns. Alternatively, one may generate independently each column according to a given probability distribution defined on the probability simplex. A standard choice is the Dirichlet distribution (14), which depends on the real parameter $s > 0$ and interpolates between the uniform measure obtained for $s = 1$ and the statistical measure for $s = 1/2$ — see e.g. [8].

Random bistochastic matrices are more difficult to generate, since the constraints imposed for the sums in each column and each row imply inevitable correlations between elements of the entire matrix. In order to obtain a bistochastic matrix one needs to normalize all its rows and columns, and this cannot be performed independently. However, since the both sets of stochastic and unital matrices are convex, iterating such a procedure, converges [9] and yields a bistochastic matrix. Note that initializing the scheme of alternating projections with different ensembles of initial conditions leads to various probability measures on the set.

The aim of this work is to analyze probability measures inside the Birkhoff polytope. In particular we discuss methods of generating random bistochastic matrices according to the uniform (flat) measure in this set. Note that the brute force method of generating random points distributed uniformly inside the unit cube of dimension $(N-1)^2$ and checking if the bistochasticity conditions are satisfied, is not effective even for N of order of 10, since the volume of the Birkhoff polytope $\mathcal{B}_N$ decreases fast with the matrix size.

The paper is organized as follows. In Section 2 we present after Sinkhorn [10] two equivalent algorithms producing a bistochastic matrix out of any square matrix of non-negative elements. An implicit formula (13) expressing the probability distribution in the set of bistochastic matrices for arbitrary N is derived in Sec. 3.1, while exact formulas for the case $N = 2$ are presented in Section 3.2. Furthermore, we obtain its power series expansion around the center $B_N^\star$ of the Birkhoff polytope and for each N we single out a particular initial distribution in the set of stochastic matrices, such that the output distribution is flat (at least locally) in the vicinity of $B_N^\star$. Finally, in section 5 we compute the value of the probability density at this very point and obtain an estimation of the volume of the set of bistochastic matrices, consistent with recent results of Canfield and McKay [12]. In Appendix A we demonstrate equivalence of two algorithms used to generate random bistochastic matrices. The key expression of this paper (36) characterising the probability distribution for random bistochastic matrices in vicinity of the center of the Birkhoff polytope is derived in Appendix B, while the third order expansion is worked out in Appendix C.

2   How to generate a bistochastic matrix?

2.1   Algorithm useful for numerical computation

In 1964 Sinkhorn [10] introduced the following iterative algorithm leading to a bistochastic matrix, based on alternating normalization of rows and columns of a given square matrix with non-negative entries:

Algorithm 1 (rows/columns normalization)

1)   take an input $N \times N$ stochastic matrix M such that each row contains at least one positive element,

2)   normalize each row-vector of M by dividing it by the sum of its elements,

3)   normalize each column-vector as in the previous point 2),

4)   stop if the matrix M is bistochastic up to certain accuracy in some norm $\| \cdot \|$, otherwise go to point 2) .

The above algorithm is symbolically visualized in Fig. 1. For an initial point M one may take an arbitrary matrix with non-negative entries. To fix the scale we may assume that the sum of all entries is equal to N, so M belongs to interior of the $(N^2 - 1)$ dimensional simplex Δ_{N^2-1}. The transformation R of normalization of the rows of M produces a unital matrix, for which the sum of all (non-negative) entries in each row is equal to unity. Subsequent normalization of the columns of $R(M)$ maps this matrix into the set of stochastic matrices. This step can be rewritten as $C = TRT$, where T denotes the transposition of the matrix. Hence the entire map reads $\Pi := CR = (T \circ R)^2$. For instance if $N = 2$ in the limit we aim to get a bistochastic matrix

$$\lim_{n\to\infty} \Pi^n(M) =: M_\infty = \begin{pmatrix} d & 1-d \\ 1-d & d \end{pmatrix} \quad , \qquad \text{for some } d \in [0,1]. \tag{1}$$
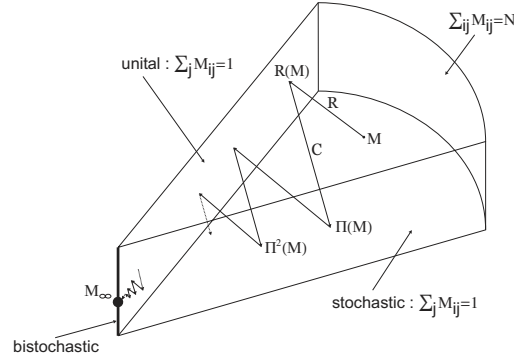

Figure 1: Sketch of the iteration procedure: a matrix M consisting of non-negative entries is sent by the transformation R (normalization of rows) into the set of unital matrices, and then by the transformation C (normalization of columns) into the set of stochastic matrices. Iterating the map $\Pi = (T \circ R)^2$ one arrives at a bistochastic matrix M_∞.

Since both these sets are convex, our procedure can be considered as a particular example of a general construction called 'projections on convex sets'. Due to convexity of these sets the procedure of alternating projections converges to a point belonging to the intersection of both sets [9]. An analogous method was recently used by Audenaert and Scheel to generate quantum bistochastic maps [13].

2.2 Algorithm suitable for analytical calculation

To perform analytical calculations of probability distribution inside the Birkhoff polytope we are going to use yet another algorithm to generate bistochastic matrix, the idea of which is due to Djoković [14]. Already in his earlier paper [10] Sinkhorn demonstrated that for a given positive matrix M there exists exactly one doubly stochastic matrix B such that $B = D^L M D^R$. In order to extend such important result from posistive matrices to non-negative ones, one has to introduce the hypotesis of *fully indecomposability* [11,14]. For the sake of clarity and reading, we prefer to mention here that the set of non-fully indecomposable (stochastic) matrices constitute a zero measure set within the set of all stochastic matrices, instead of going through the details of Sinkhorn's proof. This means that the converge of our algorithms we will assume to hold true from now onwards, has to be intended *almost everywhere* in the compact set of stochastic matrices, with respect to the usual Lebesgue measure.

Here D^L and D^R denote diagonal matrices with positive entries determined uniquely up to a scalar factor.

To set the notation, we will denote with $\mathbb{R}^+$ the positive semi-axis $(0, \infty)$ whereas, the symbol $\mathbb{R}_+$ will be used for $\mathbb{R}^+ \cup \{0\} = [0, \infty)$. Let us now consider the positive cone $\mathbb{R}_+^N$ and the set of endomorphisms over it, $End\left[\mathbb{R}_+^N\right]$, representable

by means of $N \times N$ matrices M consisting of non negative elements $m_{ij} \geqslant 0$. For any given two vectors L and R in $\mathbb{R}_+^N$, one can consider a map $\Gamma_{L,R} \in End\left[End\left[\mathbb{R}_+^N\right]\right]$, given by

$$End\left[\mathbb{R}_+^N\right] \ni M \longmapsto M' = \Gamma_{L,R}(M) \in End\left[\mathbb{R}_+^N\right] \tag{2a}$$

$$\mathbb{R}_+ \ni m_{ij} \longmapsto m'_{ij} = \Gamma_{L,R}\left(m_{ij}\right) := L_i\, m_{ij}\, R_j \in \mathbb{R}_+\,. \tag{2b}$$

Defining the positive diagonal matrices $D_{ij}^L := L_i\, \delta_{ij}$, and $D_{ij}^R := R_i\, \delta_{ij}$ respectively, one can observe that $\Gamma_{L,R}(M) = D^L M\, D^R$. Our purpose is to design an algorithm that takes a generic $M \in End\left[\mathbb{R}_+^N\right]$ as an input and produces as an output an appropriate pair of vectors $L, R \in \mathbb{R}_+^N$ such that $\Gamma_{L,R}(M) =: B$ is bistochastic.

The *stochasticity* condition implies

$$\sum_i B_{ij} = 1 = \sum_i L_i\, m_{ij}\, R_j \quad \Longrightarrow \quad R_j > 0 \quad \text{and} \quad \frac{1}{R_j} = \sum_k L_k\, m_{kj}\,. \tag{3a}$$

Analogously, *unitality* implies

$$\sum_j B_{ij} = 1 = \sum_j L_i\, m_{ij}\, R_j \quad \Longrightarrow \quad L_i > 0 \quad \text{and} \quad \frac{1}{L_i} = \sum_j m_{ij}\, R_j\,, \tag{3b}$$

so that $L, R \in (\mathbb{R}^+)^N \subset \mathbb{R}_+^N$. Both equations (3) can be merged together into a single equation for L,

$$\frac{1}{L_i} = \sum_j m_{ij}\, \frac{1}{\sum_k L_k\, m_{kj}} \tag{4}$$

which can be interpreted as a kind of *equation of the motion* for L, as it corresponds to a stationary solution of the *action–like* functional

$$\Phi[L] = -\sum_i \ln(L_i) + \sum_j \ln\left(\sum_k L_k\, m_{kj}\right)\,. \tag{5}$$

Equations (4–5) imply that if L is a solution, then for any $\lambda \in \mathbb{R}$ the rescaled vector λL is as well a solution of (5). Thus we may fix $L_N = 1$ and try to solve (4) for $L_1, L_2, \ldots, L_{N-1}$. Differentiating eq. (5) we get

$$\frac{\partial \Phi}{\partial L_i} = -\frac{1}{L_i}\left[1 - \sum_j S_{ij}\right] \quad, \quad \text{where} \quad S_{ij} := L_i\, m_{ij}\, \frac{1}{\sum_k L_k\, m_{kj}} \tag{6}$$

is a stochastic matrix. Since $L_i \neq 0$, unitality of S is attained once we impose stationarity to (6). Hence the stationary L implies that S becomes bistochastic. Equation (5) displays convexity of Φ for very small L_i ($i = 1, 2, \ldots, N-1$, $L_N = 1$). The function Φ is convex at the stationary point and starts to become concave for large L_i. Thus there is a unique minimum of the function Φ which can be reached by the following iteration procedure:

$$L_i^{(n)} = \frac{1}{\displaystyle\sum_j m_{ij}\, \frac{1}{\displaystyle\sum_k L_k^{(n-1)}\, m_{kj}}}\,, \tag{7}$$

where we fix L_N and iterate the remaining components $L_1, L_2, \ldots, L_{N-1}$ only. We start with setting $L_k^{(1)} = 1$, $\forall k$ which leads to

Algorithm 2 (convergent sequences of $\mathbb{R}^N$ vectors)

1) take an input $N \times N$ stochastic matrix $M = \{m_{ij}\}_{ij}$ and define the vector $L^{(0)} = (1, 1, \ldots, 1)^\mathsf{T} \in \mathbb{R}^N$,

2) run equation (7) yielding the vector $L^{(n)}$ out of $L^{(n-1)}$,

3) stop if the matrix $S^{(n)} := L_i^{(n)} m_{ij} \dfrac{1}{\sum_k L_k^{(n)} m_{kj}}$ is bistochastic up to a certain accuracy in some norm $\| \cdot \|$,

otherwise go to point 2).

The Algorithm (1) is expected to converge faster than the Algorithm (2), so it can be recommended for numerical implementation. On the other hand Algorithm (2) is useful to evaluate analytically the probability measure induced into the Birkhoff polytope by a given choice of the input ensemble, and it is used for this purpose in further sections. The equivalence of these two algorithms is shown in Appendix A.

3   Probability measures in the Birkhoff polytope

Assume that the algorithm is initiated with a random matrix M drawn according to a given distribution $W[\{m_{ij}\}]$ of matrices of non negative elements $m_{ij} \geqslant 0$. We want to know the distribution of the resulting bistochastic matrices B_{ij} obtained as output of the Algorithm (2). To this end, using eq. (4) and imposing stationarity condition (6), we write the distribution for B by integrating over delta functions

$$P[\{B_{ij}\}] = \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N dL_r \right) \int_0^\infty \cdots \int_0^\infty \left(\prod_{p,q=1}^N dm_{pq} \; W[\{m_{pq}\}] \right) \times \prod_{i,j=1}^N \delta \left(B_{ij} - L_i \, m_{ij} \, \frac{1}{\sum_k L_k \, m_{kj}} \right) \times$$

$$\times \, \delta \, (L_N - 1) \prod_{u=1}^{N-1} \delta \left(-\frac{1}{L_u} + \sum_t m_{ut} \, \frac{1}{\sum_v L_v \, m_{vt}} \right) \times J \{L_1, L_2, \ldots, L_{N-1}\} \,, \quad (8)$$

where the Jacobian factor reads

$$J \{L_1, L_2, \ldots, L_{N-1}\} := \det \left[\frac{\partial^2 \Phi}{\partial L_i \, \partial L_\ell} \right]_{i,\ell=1}^{N-1} = \left(\prod_{i=1}^{N-1} \frac{1}{L_i^2} \right) \times \det \left[\mathbb{1} - BB^\mathsf{T} \right]_{N-1} . \quad (9)$$

Here and in the following $\left[\mathbb{1} - BB^\mathsf{T} \right]_{N-1}$ will indicate the $(N-1) \times (N-1)$ block matrix $\left[\delta_{i\ell} - \sum_{j=1}^N B_{ij} \, B_{\ell j} \right]_{i,\ell=1}^{N-1}$, that is positive defined, and the symbol $P[\{A_{ij}\}]$ will denote the probability density P of matrices $A = \{A_{ij}\}$. This notation will also be used for matrices whose elements are functions of elements of another matrix, namely $P[\{f(A_{ij})\}]$. Plugging eq. (9) into (8) and introducing again the delta functions for variables R_j of (3a) we obtain

$$P[\{B_{ij}\}] = \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N dL_r \right) \int_0^\infty \cdots \int_0^\infty \left(\prod_{s=1}^N dR_s \right) \delta \, (L_N - 1) \int_0^\infty \cdots \int_0^\infty \left(\prod_{p,q=1}^N dm_{pq} \; W[\{m_{pq}\}] \right) \times \prod_{i,j=1}^N \delta \left(B_{ij} - L_i \, m_{ij} \, R_j \right) \times$$

$$\times \prod_{u=1}^{N-1} \delta \left(-\frac{1}{L_u} + \sum_t m_{ut} \, R_t \right) \times \prod_{w=1}^N \delta \left(R_w - \frac{1}{\sum_h L_h \, m_{hw}} \right) \times \prod_{z=1}^{N-1} \frac{1}{L_z^2} \times \det \left[\mathbb{1} - BB^\mathsf{T} \right]_{N-1} . \quad (10)$$

Using the property of the Dirac delta function and making use of the Heaviside step function θ, we perform integration over the variables dm_{pq}. Introducing new variables $\alpha_i := 1/L_i$ and $\beta_i := 1/R_i$, so that $dL_i \, dR_j \mapsto L_i^2 \, R_j^2 \, d\alpha_i \, d\beta_j$, we get

$$P[\{B_{ij}\}] = \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N d\alpha_r \, \alpha_r^{N-1} \right) \int_0^\infty \cdots \int_0^\infty \left(\prod_{s=1}^N d\beta_s \, \beta_s^{N-1} \right) \prod_{p,q=1}^N W\left[\left\{ \alpha_p B_{pq} \beta_q \right\} \right] \delta(\alpha_N - 1) \times$$

$$\times \det\left[\mathbb{1} - BB^\mathsf{T} \right]_{N-1} \times \prod_{u=1}^{N-1} \delta\left(1 - \sum_t B_{ut} \right) \times \prod_{w=1}^N \delta\left(1 - \sum_h B_{hw} \right) \times \prod_{a,c=1}^N \theta(B_{ac}) . \tag{11}$$

The last three factors show that B_{ij} is bistochastic. The factor $\det\left[\mathbb{1} - BB^\mathsf{T} \right]_{N-1}$ indicates that the expression is meaningful only in the case for which the leading eigenvalue 1 of BB^T is non-degenerate.

If the matrix m_{ij} is already stochastic,

$$W[\{m_{pq}\}] = V[\{m_{pq}\}] \times \prod_{w=1}^N \delta\left(1 - \sum_h m_{hw} \right) \times \prod_{a,c=1}^N \theta(m_{ac}) , \tag{12}$$

then the integration over β_j can be performed and we arrive at the final expression for the probability distribution inside the Birkhoff polytope which depends on the initial measure V in the set of stochastic matrices;

$$P[\{B_{ij}\}] = \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N d\alpha_r \, \alpha_r^{N-1} \right) \prod_{t=1}^N \frac{1}{\left(\sum_s \alpha_s B_{st} \right)^N} \prod_{p,q=1}^N V\left[\left\{ \alpha_p B_{pq} \frac{1}{\sum_r \alpha_r B_{rq}} \right\} \right] \delta(\alpha_N - 1) \times$$

$$\times \det\left[\mathbb{1} - BB^\mathsf{T} \right]_{N-1} \times \prod_{u=1}^{N-1} \delta\left(1 - \sum_t B_{ut} \right) \times \prod_{w=1}^N \delta\left(1 - \sum_h B_{hw} \right) \times \prod_{a,c=1}^N \theta(B_{ac}) . \tag{13}$$

The above implicit formula, valid for any matrix size N and an arbitrary initial distribution V, constitutes one of the key results of this paper. It will be now used to yield explicit expressions for the probability distribution inside the set of bistochastic matrices for various particular cases of the problem.

3.1   Measure induced by Dirichlet distribution

Let us now assume that the initial stochastic matrices are formed of N independent columns each distributed according to the Dirichlet distribution [8, 15, 17],

$$D_s(\lambda_1, \ldots, \lambda_{N-1}) = \alpha_s \, \lambda_1^{s-1} \ldots \lambda_{N-1}^{s-1} (1 - \lambda_1 - \cdots - \lambda_{N-1})^{s-1} , \tag{14}$$

where $s > 0$ is a free parameter and the normalization constant reads $\alpha_s = \Gamma[2s] / \Gamma[s]^2$.

Algorithm 3 (Random points in the simplex according to the Dirichlet distribution)

Following [18] we are going to sketch here a useful algorithm for generating random points in a simplex Δ_{N-1} according to the distribution (14).

1)   generate an N–dimensional vector X, whose elements are independent random numbers x_i from the gamma distribution $f(x_i; s, 1)$ of *shape* s and *rate* 1, so that each of them is drawn according to the probability density $x_i^{s-1} e^{-x_i} / \Gamma(s)$;

2)   normalize the vector X by dividing it by its ℓ_1 norm, $X \longmapsto Y := X/\|X\|_1$, so that the entries will become $x_i \longmapsto y_i := x_i / \sum_{k=1}^N x_k$.

A simplified version, suited for (semi)integer s is described in the appendix of [17]. In particular, to get the uniform distribution in the simplex ($s = 1$), it is sufficient to generate N independent complex Gaussian variables (with mean zero and variance equal to unity) and set the probability vector by

$$y_i = |z_i|^2 / \sum_{i=1}^N |z_i|^2. \tag{15}$$

Hence the initial stochastic matrix M is characterized by the vector consisting of N Dirichlet parameters $\boldsymbol{s} = \{s_1, \ldots, s_N\}$, which determine the distribution of each column.

The probability density can be written as

$$V_s[\{m_{ij}\}] := \prod_j D_{s_j}\left(m_{1j}, m_{2j}, \ldots, m_{N-1\,j}\right) = \mathcal{N} \prod_{ij} \left(m_{ij}\right)^{s_j-1}, \tag{16}$$

where the normalisation factor reads

$$\mathcal{N} = \prod_{j=1}^N \frac{\Gamma\left(Ns_j\right)}{\Gamma\left(s_j\right)^N}. \tag{17}$$

Thus one can obtain the probability distribution of the product

$$V_s[\{\alpha_p B_{pq} \beta_q\}] = \mathcal{N} \prod_{pq} \left(\alpha_p B_{pq} \beta_q\right)^{s_q-1} = \mathcal{N} \prod_{pq} B_{pq}{}^{s_q-1} \times \prod_x \alpha_x{}^{\sum_y s_y - N} \times \prod_z \beta_z{}^{N(s_z-1)} \tag{18}$$

and making use of eq. (13) one eventually arrives at a compact expression for the probability distribution in the set of bistochastic matrices

$$P_s[\{B_{ij}\}] = \mathcal{N} \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N d\alpha_r \; \alpha_r{}^{\sum_y s_y - 1}\right) \prod_{t=1}^N \frac{1}{\left(\sum_j \alpha_j B_{jt}\right)^{Ns_t}} \; \delta\left(\alpha_N - 1\right) \times \prod_{p,q=1}^N B_{pq}{}^{s_q-1} \times$$

$$\times \times \prod_{u=1}^{N-1} \delta\left(1 - \sum_t B_{ut}\right) \times \prod_{w=1}^N \delta\left(1 - \sum_h B_{hw}\right) \times \prod_{a,c=1}^N \theta\left(B_{ac}\right). \tag{19}$$

Although the results were obtained under the assumption that the initially random stochastic matrices are characterized by the Dirichlet distributions (16,17), one may also derive analogous results for other initial distributions. As interesting examples, one can consider the one–parameter family $V_{\boldsymbol{s},\lambda}[\{m_{ij}\}]$, in which each j–column of M is drawn according to a different gamma distribution $f\left(m_{ij}; s_j, \lambda\right)$ of *shape* s_j and *rate* λ, that is

$$V_{\boldsymbol{s},\lambda}[\{m_{ij}\}] = \prod_{j=1}^N \frac{\lambda^{Ns_j}}{\Gamma\left(s_j\right)^N} \prod_{ij} e^{-\lambda m_{ij}} \left(m_{ij}\right)^{s_j-1} \tag{20}$$

or, allowing the exponents s to vary through the whole matrix, we can start with

$$V_{\{s_{ij}\},\lambda}[\{m_{ij}\}] = \prod_{ij} \left[e^{-\lambda m_{ij}} \left(m_{ij}\right)^{s_{ij}-1}\right] \frac{\lambda^{s_{ij}}}{\Gamma\left(s_{ij}\right)}. \tag{21}$$

and recover (19), independently on the rate λ labeling the input.

3.2   Probability measures for $N = 2$

In the simplest case, for $N = 2$ and $B_{ij} = \begin{pmatrix} d & 1-d \\ 1-d & d \end{pmatrix}$, formula (19) describes the probability measure $P_{s_1,s_2}(d)$ induced into the set of bistochastic matrices by the ensemble of stochastic matrices with two independent columns distributed according to the Dirichlet measure with parameters s_1 and s_2; after integration on α_2, renaming α_1 into α, and expressing $\det\left[\mathbb{1} - BB^{\mathsf{T}}\right]_{N-1} = 2\,d\,(1-d)$, we arrive at

$$P_{s_1,s_2}(d) = \mathcal{N}\Big|_{N=2} \int_0^\infty d\alpha\, \alpha^{s_1+s_2-1} \left[\frac{1}{\alpha d + 1 - d}\right]^{2s_1} \left[\frac{1}{\alpha(1-d)+d}\right]^{2s_2} 2\,[d\,(1-d)]^{s_1+s_2-1}\,\theta(d)\,\theta(1-d)\ . \quad (22)$$

This expression can be explicitly evaluated for exemplary pairs of the Dirichlet parameters s_1 and s_2,

$$\mathrm{Pr}_{1,1}(r) = \frac{\left(1-4r^2\right)\left[\left(1+4r^2\right)\ln\left(\frac{1+2r}{1-2r}\right) - 4r\right]}{16r^3}\ , \quad (23)$$

$$\mathrm{Pr}_{3/2,3/2}(r) = \frac{\left(1-4r^2\right)^2\left[\left(3+8r^2+48r^4\right)\ln\left(\frac{1+2r}{1-2r}\right) - 12r - 48r^3\right]}{16\pi^2 r^5}\ , \quad (24)$$

$$\mathrm{Pr}_{1/2,1/2}(r) = \frac{2\ln\left(\frac{1+2r}{1-2r}\right)}{\pi^2 r}\ , \quad (25)$$

$$\mathrm{Pr}_{1/2,1}(r) = \mathrm{Pr}_{1,1/2}(r) = 1\ , \quad (26)$$

where $r = d - \frac{1}{2}$. These distributions are plotted in Fig. 2 and compared with the numerical results.

There is another important distribution that we would like to consider. We started our analysis by considering a stochastic matrix as an input state of the renormalization algorithm. However, as an initial point one can also take a generic matrix K whose four entries $\{k_{11},\ k_{12},\ k_{21},\ k_{22}\}$ are just uniformly distributed on some interval. After the first application of the half–step map $T \circ R$, (see Fig. 1) as

$$K = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \xrightarrow{\ T \circ R\ } \begin{pmatrix} \dfrac{k_{11}}{k_{11}+k_{12}} & \dfrac{k_{21}}{k_{21}+k_{22}} \\[2mm] \dfrac{k_{12}}{k_{11}+k_{12}} & \dfrac{k_{22}}{k_{21}+k_{22}} \end{pmatrix} = \begin{pmatrix} a & 1-b \\ 1-a & b \end{pmatrix}\ , \quad (27)$$

matrix K becomes stochastic, so that this problem can be reduced to the framework developed so far.

The joint probability distribution of N independent random numbers y_i', drawn according to the uniform distribution in one interval of $\mathbb{R}^+$, and then rescaled as

$$y_i' \to y_i = \frac{y_i'}{\sum_{i=1}^N y_i'}\ , \quad (28)$$

reads $P(y_1....y_N) = \delta(1 - \sum_i y_i)/\{N\,[\max(y_i)]\}^N$ [17]. In the simplest case, $N = 2$, it gives $\widetilde{p}(y) = 1/2y^2$ for $y \in (1/2,1]$, (where $y := y_1 = 1 - y_2$) and symmetrically for $y \in [0, 1/2]$. Using this and assuming independence between the entries of the matrix K, the distribution for the variable a and b of (27) reads

$$\widetilde{P}(a,b) := \widetilde{p}(a) \times \widetilde{p}(b) = \left(\frac{1}{2\,\max\{a,1-a\}\,\max\{b,1-b\}}\right)^2\ . \quad (29)$$

Plugging the last expression into the r.h.s. of (22) we obtain (see Fig. 2(e))

$$\widetilde{\mathrm{Pr}}(d) = \ = \frac{2\,(1-2\,|r|)\left[1 + 2\ln\left(\frac{1+2|r|}{1-2|r|}\right)\right]}{(1+2\,|r|)^3}\ , \quad (30)$$
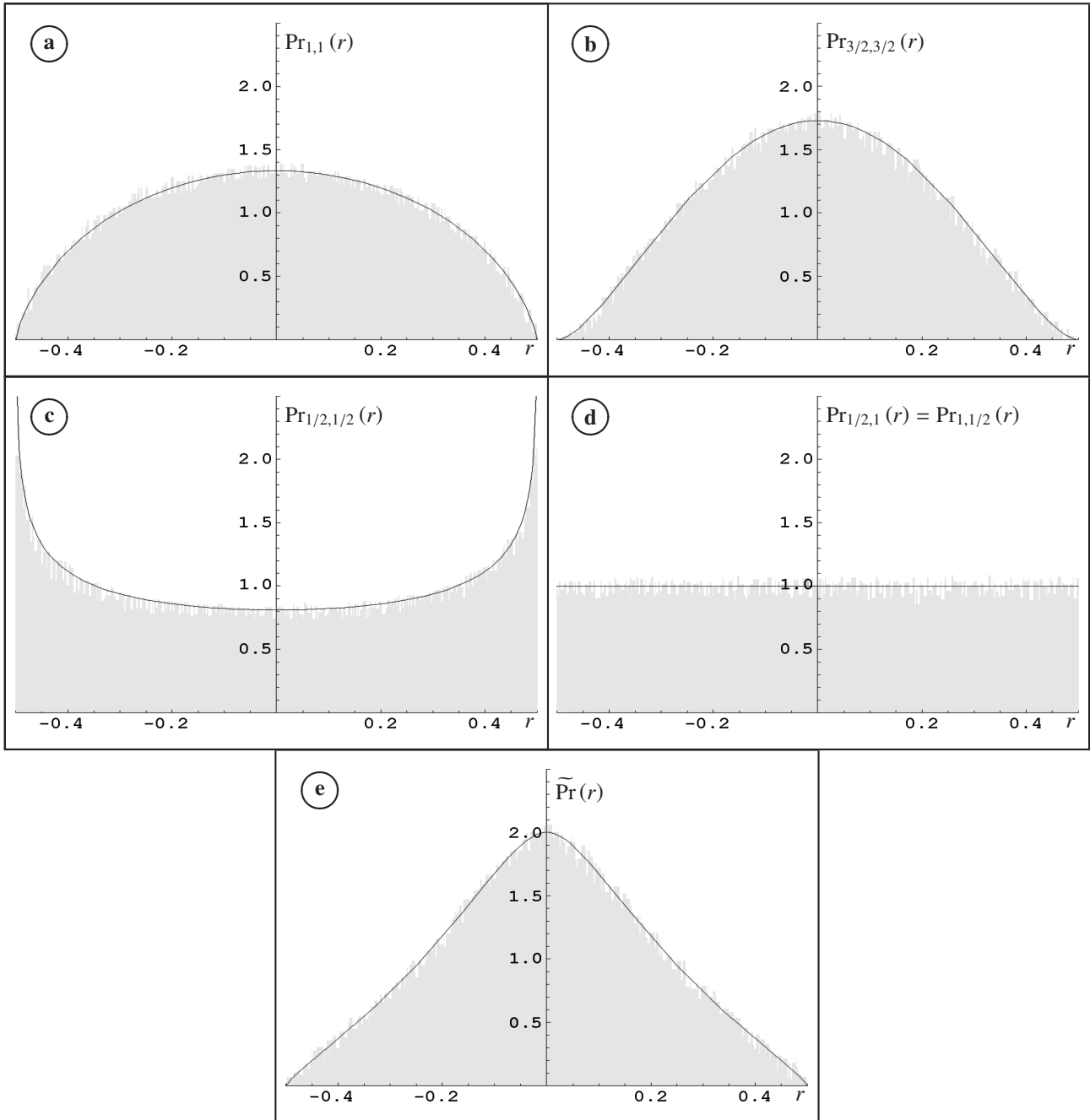
where again $r = d - 1/2$.

Figure 2: Probability distribution $\Pr(r)$ in the set of $N = 2$ bistochastic matrices for various initial measures. Histograms obtained numerically for a sample of 10^6 initial matrices by applying Algorithm (1) are compared with analytical probability distributions (solid lines); **(a)** semicircle–like (23) for $\Pr_{1,1}$; **(b)** Gaussian–like (24) for $\Pr_{3/2,3/2}$; **(c)** convex distribution (25) for $\Pr_{1/2,1/2}$; **(d)** flat distribution (26) for $\Pr_{1/2,1}$ and **(e)** distribution (30).

3.3   Symmetries and relations with the unistochastic matrices for $N = 2$

Consider the map $T \circ R$ defined in (27) acting on an initially stochastic matrix $\begin{pmatrix} a & 1-b \\ 1-a & b \end{pmatrix}$. The symmetry of this map with respect to diagonal lines $a = b$ and $a = -b$ implies that:

- the limit distribution $\mathrm{Pr}_{s_a, s_b}(r)$ is an even function of r,

- $\mathrm{Pr}_{s_a, s_b} = \mathrm{Pr}_{s_b, s_a}$, for any s_a and s_b. The final accumulation point $d \in [0, 1]$ can be achieved from the point (a, b) as well as from (b, a).

In particular the second point implies that **if** $\mathrm{Pr}_{s_a, s_b}(r)$ is the output probability density when the (a, b)–distribution is given by $P_{s_a, s_b}(a, b)$ and **if** $s_a \neq s_b$ **then** for any given $\lambda \in [0, 1]$ the distribution $\lambda\, P_{s_a, s_b}(a, b) + (1 - \lambda)\, P_{s_b, s_a}(a, b)$ will give the same output. Using this we can restore the symmetry between (a, b) simply by picking $\lambda = 1/2$.

$$P^{\mathrm{sym}}_{[1/2, 1]}(a, b) := \frac{1}{2} P_{1/2, 1}(a, b) + \frac{1}{2} P_{1, 1/2}(a, b) = \frac{1}{2\pi \sqrt{a(1-a)}} + \frac{1}{2\pi \sqrt{b(1-b)}} \tag{31}$$

is a symmetric distribution for a and b which produce, at a long run, the uniform distribution $P(d) = 1$. Note that the above formula is not of a product form, so the distribution in both columns are correlated. In fact such a probability distribution can be interpreted as a classical analogue of the quantum entangled state [8, 16].

Random pairs (a, b) distributed according to distribution (31) can be generated by means of the following algorithm,

1)   generate the number a according to $D_{1/2}(a)$ and b according to $D_1(b)$

2)   flip a coin: on tails do nothing, on heads exchange a with b

For $N = 2$ there exists an equivalence between the set of bistochastic and unistochastic matrices [20]. The latter set is defined as the set of 2×2 matrices whose entries are squared moduli of entries of unitary matrices. The Haar measure on $U(2)$ induces a natural, uniform measure in the set of unistochastic matrices: if U is random then $P(|U_{11}|^2) = P(|U_{22}|^2) = 1$ on $[0, 1]$. Hence initiating Algorithm (1) with stochastic matrices distributed according to eq. (31) we produce the same measure in the set of bistochastic matrices as it is induced by the Haar measure on $U(2)$ by the transformation $B_{ij} = |U_{ij}|^2$.

4   In search for the uniform distribution for an arbitrary N

For an arbitrary N we shall compute the probability density at the center $B_N^\star$ of the Birkhoff polytope,

$$B_N^\star = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \cdots \cdots \cdots \cdots \cdots \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}. \tag{32}$$

Let us begin our analysis by expanding $P_s[\{B_{ij}\}]$ around the center $B_N^\star$ (32) of the Birkhoff polytope. We start from (19) with $\mathcal{N}$ given by equation (17), so that

$$P_s[\{B_{ij}\}] = \widetilde{P}_s[\{B_{ij}\}] \times \prod_{u=1}^{N-1} \delta\left(1 - \sum_t B_{ut}\right) \times \prod_{w=1}^{N} \delta\left(1 - \sum_h B_{hw}\right) \times \prod_{a,c=1}^{N} \theta(B_{ac}) \tag{33}$$

and

$$\widetilde{P}_s[\{B_{ij}\}] = \mathcal{N} \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N d\alpha_r \, \alpha_r^{\sum_y s_y - 1} \right) \prod_{w=1}^N \frac{1}{\left(\sum_s \alpha_s B_{st}\right)^{N s_w}} \, \delta\left(\alpha_N - 1\right) \times$$

$$\times \prod_{p,q=1}^N B_{pq}^{\,s_q-1} \times \det\left[\mathbb{1} - BB^\mathsf{T}\right]_{N-1} , \tag{34}$$

on the manifold $\sum_t B_{ut} = \sum_h B_{hw} = 1$, $B_{ac} \geqslant 0$.

4.1   Expansion of probability distribution around the center of the polytope

Expanding $\widetilde{P}_s[\{B_{ij}\}]$ in power of δB_{ij} with

$$B_{ij} = \frac{1}{N} + \delta B_{ij} , \quad \sum_i \delta B_{ij} = \sum_j \delta B_{ij} = 0 . \tag{35}$$

we obtain, as shown in Appendix B, the following result

$$\widetilde{P}_s[\{B_{ij}\}] = P_N^\star \left\{ 1 + \left(\frac{N^2}{2} - 1 \right) \sum_{pq} \left(\delta B_{pq} \right)^2 - \frac{\sigma N^3}{2\left(\sigma N + 1\right)} \sum_{pq} s_q \left(\delta B_{pq} \right)^2 + \frac{N^3}{2\left(\sigma N + 1\right)} \sum_p \left(\sum_q s_q \delta B_{pq} \right)^2 + O\left((\delta B)^3\right) \right\} , \tag{36}$$

where $\sigma = \sum_{j=1}^N s_j$ denotes the sum of the Dirichlet parameters for each column and the factor

$$P_N^\star := P_s[\{ B_{ij} = 1/N , \; \forall ij \}] = N^{N^2-1} \frac{\Gamma\left(\sum_m s_m\right)^N}{\Gamma\left(N \sum_m s_m\right)} \prod_{n=1}^N \frac{\Gamma\left(N s_n\right)}{\Gamma\left(s_n\right)^N} , \tag{37}$$

is equal to the value of the probability distribution at the center of the polytope $\mathcal{B}_N$, which corresponds to $\delta B = 0$.

Assume now that there exists a set of Dirichlet exponents s_i, such that $\widetilde{P}_s[\{B_{pq}\}]$ is constant on the required manifold (35). Then the quadratic form in δB_{pq} must be identically zero. For $N = 2$ this yields only one equation for two exponents, $2s_1 + 2s_2 + 1 = 8s_1 s_2$, which can e.g. be fulfilled by $s_1 = 1/2$ and $s_2 = 1$ (compare with Section 3.3).

For $N \geqslant 3$, however, this gives more independent equations, in general $(N-1)^2$, namely the number of independent variables parameterizing the Birkhoff polytope. Being N the number of exponents to be determined, if a solution exists, then it is unique. Actually the solution exists, and corresponds to take all s_i equal to each other: let's call s this collective exponent. Within this constraint, the last term in (36) drops out, because of equation (35), and the entire quadratic form can be zero, provided that we choose

$$\left(\frac{N^2}{2} - 1 \right) = \frac{\sigma N^3}{2\left(\sigma N + 1\right)} s^\star . \tag{38}$$

Now, setting $\sigma = N s^\star$, we arrive at $N^4 s^{\star 2} = \left(N^2 s^\star + 1\right)\left(N^2 - 1\right)$, whose unique positive solution is

$$s^\star = \frac{1}{2N^2} \left(N^2 - 2 + \sqrt{N^4 - 4}\right) = 1 - \frac{1}{N^2} - \frac{1}{N^4} + O\left(\frac{1}{N^8}\right) . \tag{39}$$

The distribution generated by the choice $s = s^\star$ will be flat at the center of the polytope but it needs not to be globally uniform.

It is not possible to find an initial Dirichlet distribution which gives the output distribution uniform in the vicinity of the center of the Birkhoff polytope up to the third order — see Appendix C.

4.2   Numerical results for $N = 3$

Properties of the measures induced in the space of bistochastic matrices by applying the iterative Algorithm (1) were analyzed for $N = 3$. As a starting point we took a random stochastic matrix M generated according to the Dirichlet distribution (14) with the same parameter for all three columns, $s_1 = s_2 = s_3 = s$. The resulting bistochastic matrix, $B = \lim_{n\to\infty} \Pi^n(M)$, can be parameterized by

$$B = \left[\begin{array}{cc|c} B_{11} & B_{12} & * \\ B_{21} & B_{22} & * \\ \hline * & * & * \end{array}\right],$$

where the $*$-marked entries depend on the entries B_{jk}, with $j,k \in \{1,2\}$. A sample of initial points consisted of 10^8 stochastic matrices generated according to the Dirichlet distribution with the optimal value $s^\star = \frac{1}{18}(7 + \sqrt{77})$ which follows from eq. (39)). It produces an ensemble covering the entire $4D$ Birkhoff polytope formed by the convex hull of the six different permutation matrices of order three.
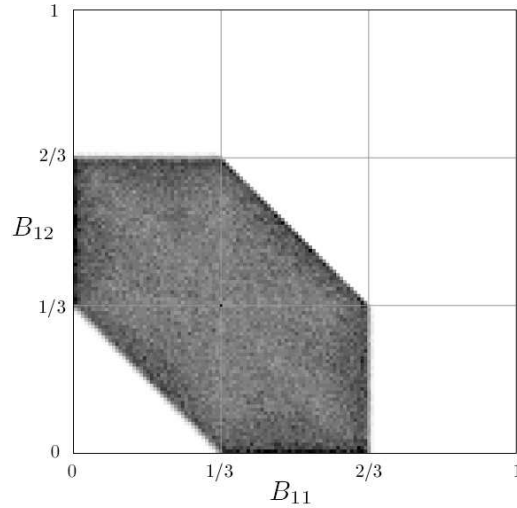

Figure 3: Probability density at a subset of the Birkhoff polytope for $N = 3$, the "fat" hexagon characterized by $\left[B_{11}, B_{12}, \frac{1}{3} \pm 0.01, \frac{1}{3} \pm 0.01\right]$, for initially stochastic matrices generated with the Dirichlet parameter $s^\star$ given by eq. (39).

To visualize numerical results we selected the cases for which $B_{21} = B_{22} = 1/3 \pm 0.01$. Such a two dimensional cross-section of the Birkhoff polytope has a shape of a hexagon at the plane (B_{11}, B_{12}), centered at the center of the body, $B_3^\star = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}; \ldots, \frac{1}{3}\right]$. Figure 3 shows the probability distribution along this section, obtained from these 4×10^6 realizations of the algorithm which produce bistochastic matrices inside a layer of width 0.02 along the section.

As expected for the critical value $s^\star$ of the Dirichlet parameter, the resulting distribution is flat in the vicinity of the center of the polytope. However, this distribution is not globally uniform and shows a slight enhancement of the probability (darker color) along the boundary of the polytope.

This feature is further visible in Fig 4, which shows a comparison of the results obtained for two different initial measures on a one–dimensional cross section of Fig. 3. Although the measure obtained for the critical parameter $s^\star$

is indeed uniform in the vicinity of the center, namely around $B_{11} = 1/3$, the measure induced by random stochastic matrices with the flat measure, $s = 1$, displays similar properties. Since for larger matrix size N the value of the optimal parameter $s^\star$ tends to unity as $1 - 1/N^2$, it seems reasonable to generate random bistochastic matrices of a larger size initiating the iterative Algorithm (1) with random stochastic matrices distributed according to the uniform measure, (i.e. each column is generated independently according to the Dirichlet distribution with $s = 1$).
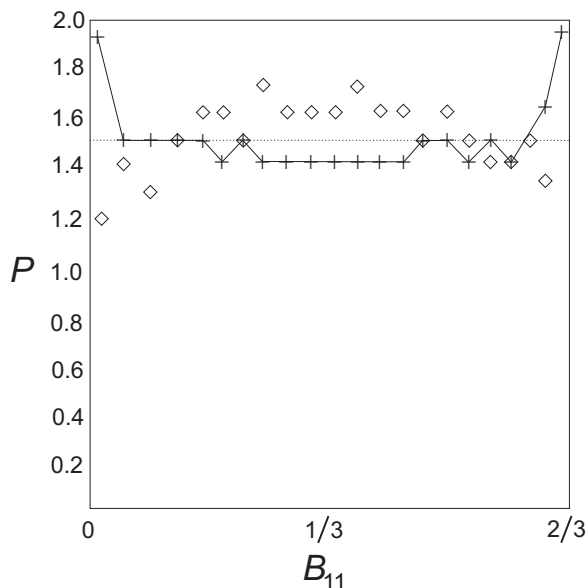

Figure 4: Probability density along the line $B_{12} = \frac{1}{3}$ of Fig. 3 obtained from 5×10^3 events for two initial measures: **(a)** the critical parameter $s = s^\star$ (marked by $+$ and decorated by a solid line to guide the eye) and **(b)** the flat measure $s = 1$ (marked by $\diamond$).

5 Estimation of the volume of the Birkhoff polytope

The set $\mathcal{B}_N$ of bistochastic matrices of size N forms a convex polytope in $\mathbb{R}^{(N-1)^2}$. Its volume with respect to the Euclidean measure is known for $N = 2, ..., 10$ [5, 6]. The concrete numbers depend on the normalization chosen. For instance, in the simplest case the set $\mathcal{B}_2$ forms an interval $d \in [0, 1]$, any point of which corresponds to the bistochastic matrix, $B(d) = \left(\begin{smallmatrix} d & 1-d \\ 1-d & d \end{smallmatrix} \right)$. If the range of the single, independent element is concerned, the *relative volume* of the polytope reads $v(\mathcal{B}_2) = 1$. On the other hand, if we regard this set as an interval in $\mathbb{R}^4$, its length is equal to the *volume* of the Birkhoff polytope, $\mathrm{Vol}(\mathcal{B}_2) = \sqrt{4} = 2$. In general, both definitions of the volumes are related by [12]

$$\mathrm{Vol}(\mathcal{B}_N) = N^{N-1} v(\mathcal{B}_N) . \tag{40}$$

In Section 3 we derived formula (37), giving the probability distribution $P_N^\star$ at the center $B_N^\star$ of the Birkhoff polytope induced by the Dirichlet measure on the space of input stochastic matrices. If all Dirichlet parameters are equal to $s_i = s$ for $i = 1, \ldots N$ then formula (37) simplifies to

$$P_N^\star(s) = \frac{\Gamma(Ns)^{2N}}{\Gamma(N^2 s) \, \Gamma(s)^{N^2}} \, N^{N^2-1} , \tag{41}$$

Making use of the Stirling expansion

$$\Gamma(x) \approx \sqrt{2\pi}\, x^{x-1/2}\, e^{-x}\, \left[1 + \frac{1}{12x} + O\!\left(\frac{1}{x^2}\right)\right],$$
(42)

and plugging it into eq. (41) we obtain an approximation valid for a large matrix size N,

$$P_N^\star \approx N^{N^2-N}\,(2\pi)^{N-1/2}\, s^{sN^2-N+1/2}\,[\Gamma(s)]^{-N^2}\,\exp\left\{-sN^2 + \frac{1}{6s} + O\!\left(\frac{1}{N}\right)\right\}.$$
(43)

For $s = s^\star = 1 - 1/N^2 + O(1/N^4)$ this distribution is flat in the vicinity of the center $B_N^\star$ – compare eq. (39). Assuming it is close to uniform in the entire Birkhoff polytope, we obtain an approximation of its relative volume, $v(\mathcal{B}_N) \approx 1/P_N^\star$. Substituting $s^\star$ into (43) we arrive at

$$v(\mathcal{B}_N) \approx N^{N-N^2}\,(2\pi)^{1/2-N}\,\exp\left\{N^2 + C + O\!\left(\frac{1}{N}\right)\right\}.$$
(44)

Making use of the expansion $\Gamma(1 + x) = 1 - \gamma x + O(x^2)$ we can express the value of C by the Euler gamma constant $\gamma \approx 0.577\,215\,665\ldots$ The result is $C = \gamma - 1/6 \approx 0.410\,548\,998\ldots$.

Interestingly, the above approximation is identical, up to a value of this constant, with the recent result of Canfield and Mackay [12]. Making use of the relation (40) we see that their asymptotic formula for the volume vol($\mathcal{B}_N$) of the Birkhoff polytope is consistent with eq. (44) for $C = 1/3$. This fact provides a strong argument that the distribution generated by the Dirichlet measure with $s = s^\star$, is close (but not equal) to the uniform distribution inside the Birkhoff polytope. Furthermore, the initially flat distribution of the stochastic matrices, obtained for $s = 1$, leads to yet another reasonable approximation for the relative volume of $\mathcal{B}_N$, equivalent to (44) with $C = -1/6$.

6   Concluding Remarks

In this paper we introduced several ensembles of random stochastic matrices. Each of them can be considered as an ensemble of initial points used as input data for the Sinkhorn Algorithm, which generates bistochastic matrices. Thus any probability measure $W[M]$ in the set of stochastic matrices induces a certain probability measure $P[B]$ in the set of bistochastic matrices.

Let us emphasize that the iterative procedure of Sinkhorn [10] applied in this work, covers the entire set of bistochastic matrices. This is not the case for the ensemble of *unistochastic matrices*, which are obtained from a unitary matrix by squaring moduli of its elements. Due to unitarity of U the matrix $B_{ij} = \left|U_{ij}\right|^2$ is bistochastic, and the Haar measure on $U(N)$ induces a certain measure inside the Birkhoff polytope [20]. However, for $N \geqslant 3$, this measure does not cover the entire Birkhoff polytope since in this case there exist bistochastic matrices which are not unistochastic [1, 20].

In the general case of arbitrary N we derive an integral expression representing the probability distribution inside the $(N-1)^2$–dimensional Birkhoff polytope $\mathcal{B}_N$ of bistochastic matrices. In the simplest case of $N = 2$ it is straightforward to obtain explicit formulae for the probability distribution in the set of bistochastic matrices induced by the ensemble of stochastic matrices, in which both columns are independent. Furthermore, we find that to generate the uniform (flat) measure, $P[B] = \text{const}$, one needs to start with random stochastic matrices of size 2 distributed according to eq. (31), for which both columns are correlated.

For an arbitrary N the integral form for the probability distribution can be explicitly worked out for a particular point — the flat, van der Waerden matrix (32) located at the center of the Birkhoff polytope. In this case we obtain an explicit formula for the probability distribution at this point as a function of the parameters $\{s_i \mid 1 \leqslant i \leqslant N\}$ defining the Dirichlet distribution for each column of the initially random stochastic matrix. Expanding the probability density in the vicinity of

$B_N^\star$ we find the condition for the optimal parameters $s_i = s^\star$, for which the density $P[B]$ is flat in this region. Discrepancy of the measure constructed in this way from the uniform distribution is numerically analyzed in the case $N = 3$.

This measure is symmetric with respect to permutations of rows and columns of the matrix and for large N it tends to the uniform measure in the set of bistochastic matrices. For large N the optimal Dirichlet parameter $s^\star$ tends to unity as $1 - 1/N^2$. Thus we may suggest a simplified procedure of taking the initial stochastic matrices according to the flat measure, ($s = 1$). Each column of such a random stochastic matrix is drawn independently and it consists of N numbers distributed uniformly in the simplex Δ_{N-1}. With an initial matrix constructed in this way we are going to run Algorithm (1). Such a procedure is shown to work fine already for $N = 3$. We tend to believe that this scheme of generating random bistochastic matrices could be useful for several applications in mathematics, statistics and physics.

Assuming that a given probability measure in a compact set is flat, the value of the probability density P at an arbitrary point x gives us an information about the Euclidean volume of this set, $V = 1/P(x)$. We were pleased to find that the optimal algorithm for generating random bistochastic matrices is characterized by an inverse probability $1/P_s[B_N^\star]$ at the center $B_N^\star$ of the polytope which displays the same dependence on the dimension N as the volume of the Birkhoff polytope, $\text{Vol}(\mathcal{B}_N)$, derived in [12].

Although in this paper we analyzed dynamics in the classical probability simplex, the main idea of the algorithm may be generalized for the quantum dynamics. In such a case a stochastic matrix corresponds to a stochastic map (so called quantum operation), which sends the set of quantum states (Hermitean, positive matrices of trace one) into itself [8]. A quantum stochastic map is called *bistochastic*, if it preserves the maximally mixed state, $\mathbb{1}/N$. To generate random bistochastic maps one can use an analogous technique of alternating projection onto the subspaces in which a given map or its dual is stochastic. Such an algorithm suitable for the quantum problem, was proposed independently by Audenaert and Scheel [13]. First results concerning various measures induced into the set of quantum stochastic maps are presented in [25].

Acknowledgements

Appendix A

In this appendix we demonstrate that the Algorithm (2) suitable for analytical calculations is equivalent with the Sinkhorn Algorithm (1).

To apply the former Algorithm (2) one takes some initial matrix $M \in End\left[\mathbb{R}_+^N\right]$ and makes it bistochastic by means of left– and right–multiplication by two matrices D^L, and D^R. The latter are limits of convergent sequences of diagonal matrices $D^L = \lim_n D_n^L$ and $D^R = \lim_n D_n^R$ and the finally $B = D^L M D^R$.

In a similar way, Algorithm (1) performs the same task of transforming the initially stochastic matrix $M \in End\left[\mathbb{R}_+^N\right]$ into a bistochastic matrix B by alternating rows– and columns–normalization (R and C, for short), which in turn is the same of left–, respectively right–multiplication by diagonal matrices. Once a matrix $M = \{m_{pq} \geqslant 0\}$ is given to renormalize the pth row means to divide each of its elements by the factor $\sum_q m_{pq}$,

$$m_{pq} \longmapsto m'_{pq} = \hat{L}_p\, m_{pq} \quad , \quad \text{with} \qquad \frac{1}{\hat{L}_p} = \sum_q m_{pq} . \tag{45a}$$

Analogously, to renormalize the qth column means to divide each of its elements by the factor $\sum_p m_{pq}$,

$$m_{pq} \longmapsto m'_{pq} = m_{pq}\,\hat{R}_q \quad , \quad \text{with} \qquad \frac{1}{\hat{R}_q} = \sum_p m_{pq}\,. \tag{45b}$$

Let us now run the Algorithm 1, taking as an input a generic $M^{(0)} = \{m_{pq}^{(0)} \geqslant 0\}$, and set $\mathsf{1CRCRCR\ldots}$ to be the row–column renormalization sequence, where the first symbol $\mathsf{1}$ denotes the dummy operation

$$\frac{1}{\hat{L}_p^{(0)}} = 1 \qquad \blacklozenge \qquad m_{pq}^{(0)} \longmapsto \overline{m}_{pq}^{(0)} = \hat{L}_p^{(0)}\, m_{pq}^{(0)}\,. \tag{46a}$$

Now we start with equations (45b)

$$\frac{1}{\hat{R}_q^{(0)}} = \sum_p \overline{m}_{pq}^{(0)} = \sum_p \hat{L}_p^{(0)}\, m_{pq}^{(0)} \qquad \blacklozenge \qquad \overline{m}_{pq}^{(0)} \longmapsto m_{pq}^{(1)} = \overline{m}_{pq}^{(0)}\,\hat{R}_q^{(0)} = \hat{L}_p^{(0)}\, m_{pq}^{(0)}\,\hat{R}_q^{(0)}\,, \tag{46b}$$

followed by (45a)

$$\frac{1}{\hat{L}_p^{(1)}} = \sum_q m_{pq}^{(1)} = \sum_q \hat{L}_p^{(0)}\, m_{pq}^{(0)}\,\hat{R}_q^{(0)} \qquad \blacklozenge \qquad m_{pq}^{(1)} \longmapsto \overline{m}_{pq}^{(1)} = \hat{L}_p^{(1)}\, m_{pq}^{(1)}\,. \tag{46c}$$

The next two steps are

$$\frac{1}{\hat{R}_q^{(1)}} = \sum_p \overline{m}_{pq}^{(1)} = \sum_p \hat{L}_p^{(1)}\, m_{pq}^{(1)} \qquad \blacklozenge \qquad \overline{m}_{pq}^{(1)} \longmapsto m_{pq}^{(2)} = \overline{m}_{pq}^{(1)}\,\hat{R}_q^{(1)} = \hat{L}_p^{(1)}\, m_{pq}^{(1)}\,\hat{R}_q^{(1)} = \hat{L}_p^{(1)}\,\hat{L}_p^{(0)}\, m_{pq}^{(0)}\,\hat{R}_q^{(0)}\,\hat{R}_q^{(1)} \tag{46d}$$

and

$$\frac{1}{\hat{L}_p^{(2)}} = \sum_q m_{pq}^{(2)} = \sum_q \quad \hat{L}_p^{(1)}\,\hat{L}_p^{(0)}\, m_{pq}^{(0)}\,\hat{R}_q^{(0)}\,\hat{R}_q^{(1)} \qquad \blacklozenge \qquad \cdots \tag{46e}$$

so that the iteration procedure can be written as

$$\begin{cases} \dfrac{1}{\hat{L}_p^{(n)}\hat{L}_p^{(n-1)}\cdots\hat{L}_p^{(1)}\hat{L}_p^{(0)}} = \sum_q m_{pq}^{(0)}\,\,\hat{R}_p^{(0)}\hat{R}_q^{(1)}\cdots\hat{R}_q^{(n-1)} \\[3mm] \dfrac{1}{\hat{R}_p^{(n)}\hat{R}_q^{(n-1)}\cdots\hat{R}_q^{(1)}\hat{R}_q^{(0)}} = \sum_p \hat{L}_p^{(n)}\hat{L}_p^{(n-1)}\cdots\hat{L}_p^{(1)}\hat{L}_p^{(0)}\,\,m_{pq}^{(0)}\,. \end{cases} \tag{47}$$

The latter form can be rewritten more compactly,

$$\frac{1}{\check{L}_p^{(n)}} = \sum_q \,\,m_{pq}^{(0)}\,\,\frac{1}{\displaystyle\sum_s \,\check{L}_s^{(n-1)}\,\,m_{sq}^{(0)}}\,, \tag{48}$$

where we introduced new variables

$$\check{L}_s^{(n)} := \prod_{\ell=1}^n \hat{L}_s^{(\ell)} \qquad \text{and} \qquad \check{R}_s^{(n)} := \prod_{\ell=1}^n \hat{R}_s^{(\ell)}\,. \tag{49}$$

Equation (47) is formally equivalent to (7), the only difference being in the number of component of L vectors, respectively $\check{L}$, that are processed: in Algorithm (1) one iterates all $\check{L}_s^{(n)}$, whereas in Algorithm (2) the element $L_N^{(n)}$ is fixed to unity in each step. We know that the solution of the limit equation for $L^{(n)}$ is not unique. But the only non-uniqueness is due to multiplication by a fixed factor $\eta > 0$.

Appendix B

In this appendix we present the basic steps allowing one to derive the central result of this work - the second order expansion (36) around the center of the Birkhoff polytope of the probability distribution generated by Dirichlet random stochastic matrices.

Since $\widetilde{P}_s[\{B_{ij}\}] \geqslant 0$, it is convenient to expand $\ln \widetilde{P}_s[\{B_{ij}\}]$. We denote the sum of the Dirichlet parameters for each column by $\sigma = \sum_{j=1}^{N} s_j$ and start with the following integral

$$
Q_s[\{B_{ij}\}] := \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N d\alpha_r \, \alpha_r^{\sigma-1} \right) \prod_{w=1}^N \frac{1}{\left[\sum_h \alpha_h \frac{1}{N} + \sum_h \alpha_h \delta B_{hw} \right]^{Ns_w}} \, \delta(\alpha_N - 1) =
$$

$$
= \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N d\alpha_r \, \alpha_r^{\sigma-1} \right) \frac{1}{\left(\sum_h \alpha_h \frac{1}{N} \right)^{N\sigma}} \exp\left[-\sum_{w=1}^N Ns_w \, \ln\left(1 + \frac{\sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h \frac{1}{N}} \right) \right] \delta(\alpha_N - 1) . \quad (50)
$$

Expanding the function $\ln(1 + x) = x - \frac{x^2}{2} + O(x^3)$,

$$
\exp\left[-\sum_{w=1}^N Ns_w \, \ln\left(1 + \frac{\sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h \frac{1}{N}} \right) \right] = \exp\left\{ -\sum_{w=1}^N Ns_w \left[\frac{\sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h \frac{1}{N}} - \frac{1}{2} \left(\frac{\sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h \frac{1}{N}} \right)^2 + O\left((\delta B)^3 \right) \right] \right\}, \quad (51)
$$

and then $e^{-x} \approx 1 - x + x^2/2$ we get

$$
Q_s[\{B_{ij}\}] = N^{N\sigma} \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N d\alpha_r \, \alpha_r^{\sigma-1} \right) \frac{\delta(\alpha_N - 1)}{\left(\sum_h \alpha_h \right)^{N\sigma}} \left\{ 1 - N^2 \frac{\sum_w s_w \sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h} + \right.
$$

$$
\left. + \frac{N^3}{2} \sum_w s_w \left(\frac{\sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h} \right)^2 + \frac{N^4}{2} \left(\frac{\sum_w s_w \sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h} \right)^2 + O\left((\delta B)^3 \right) \right\} . \quad (52)
$$

Thus we have to integrate the following expression for an arbitrary vector of parameters $\vartheta_w \geqslant 0$

$$
I := \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N d\alpha_r \, \alpha_r^{\sigma-1} \right) \frac{\delta(\alpha_N - 1)}{\left(\sum_h \alpha_h \right)^{N\sigma}} \prod_{w=1}^N \left(\frac{\alpha_w}{\sum_h \alpha_h} \right)^{\vartheta_w} =
$$

$$
= \frac{1}{\Gamma(\sigma N + m)} \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^N d\alpha_r \, \alpha_r^{\sigma-1} \right) \alpha_1^{\vartheta_1} \alpha_2^{\vartheta_2} \cdots \alpha_{N-1}^{\vartheta_{N-1}} \, \delta(\alpha_N - 1) \int_0^\infty dt \, e^{-t \sum_h \alpha_h} \, t^{N\sigma + m - 1} , \quad (53)
$$

Here $m := \vartheta_1 + \vartheta_2 + \cdots + \vartheta_{N-1} + \vartheta_N$, so the integral reads

$$
I = \frac{\Gamma(\sigma + \vartheta_1)\Gamma(\sigma + \vartheta_2)\ldots\Gamma(\sigma + \vartheta_{N-1})\Gamma(\sigma + \vartheta_N)}{\Gamma(\sigma N + m)} = \frac{\Gamma(\sigma)^N}{\Gamma(\sigma N)} \left\langle \prod_{w=1}^N \left(\frac{\alpha_w}{\sum_s \alpha_s} \right)^{\vartheta_w} \right\rangle, \quad (54)
$$

with

$$
\left\langle \prod_{w=1}^N \left(\frac{\alpha_w}{\sum_h \alpha_h} \right)^{\vartheta_w} \right\rangle := \frac{\Gamma(\sigma N)}{\Gamma(\sigma N + \sum_i \vartheta_i)} \prod_{j=1}^N \frac{\Gamma(\sigma + \vartheta_j)}{\Gamma(\sigma)} . \quad (55)
$$

This expression, completely symmetric in all variables α_1, $\alpha_2 \ldots \alpha_{N-1}$, α_N allows us to calculate the expansion of the integral (52):

$$Q_s[\{B_{ij}\}] = N^{N\sigma} \frac{\Gamma(\sigma)^N}{\Gamma(\sigma N)} \left\{ 1 - N^2 \left\langle \frac{\sum_w s_w \sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h} \right\rangle + \right.$$
$$\left. + \frac{N^3}{2} \sum_w s_w \left\langle \left(\frac{\sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h} \right)^2 \right\rangle + \frac{N^4}{2} \left\langle \left(\frac{\sum_w s_w \sum_h \alpha_h \delta B_{hw}}{\sum_h \alpha_h} \right)^2 \right\rangle + O\left((\delta B)^3 \right) \right\}. \quad (56)$$

Therefore we need

$$\begin{cases} \left\langle \dfrac{\alpha_u}{\sum_h \alpha_h} \right\rangle &= \dfrac{\Gamma(\sigma+1)}{\Gamma(\sigma)} \cdot \dfrac{\Gamma(\sigma N)}{\Gamma(\sigma N+1)} = \dfrac{\sigma}{\sigma N} = \dfrac{1}{N} \\[2ex] \left\langle \dfrac{\alpha_u^2}{\left(\sum_h \alpha_h\right)^2} \right\rangle &= \dfrac{\Gamma(\sigma+2)}{\Gamma(\sigma)} \cdot \dfrac{\Gamma(\sigma N)}{\Gamma(\sigma N+2)} = \dfrac{\sigma(\sigma+1)}{\sigma N(\sigma N+1)} = \dfrac{\sigma+1}{N(\sigma N+1)} \\[2ex] \left\langle \dfrac{\alpha_u \alpha_v}{\left(\sum_h \alpha_h\right)^2} \right\rangle &= \dfrac{\Gamma(\sigma+1)\,\Gamma(\sigma+1)}{\Gamma(\sigma)\;\Gamma(\sigma)} \cdot \dfrac{\Gamma(\sigma N)}{\Gamma(\sigma N+2)} = \dfrac{\sigma \cdot \sigma}{\sigma N(\sigma N+1)} = \dfrac{\sigma}{N(\sigma N+1)} \end{cases} \quad (57)$$

Thus the second term in (56) is

$$\left\langle \frac{\sum_w s_w \sum_h \alpha_h \delta B_{hw}}{\sum_i \alpha_i} \right\rangle = \frac{1}{N} \sum_{hw} \delta B_{hw} s_w = 0 \quad , \quad \text{because of (35)} \cdot \quad (58)$$

For the third term we have

$$J := \sum_w s_w \left\langle \left(\frac{\sum_h \alpha_h \delta B_{hw}}{\sum_i \alpha_i} \right)^2 \right\rangle = \sum_w s_w \sum_h \sum_{h' \neq h} \left\langle \frac{\alpha_h \alpha_{h'}}{\left(\sum_h \alpha_h\right)^2} \right\rangle \delta B_{hw} \delta B_{h'w} + \sum_w s_w \sum_h \left\langle \frac{\alpha_h^2}{\left(\sum_h \alpha_h\right)^2} \right\rangle (\delta B_{hw})^2$$
$$= \sum_w \frac{\sigma s_w}{N(\sigma N+1)} \sum_h \delta B_{hw} \sum_{h' \neq h} \delta B_{h'w} + \sum_w \frac{(\sigma+1) s_w}{N(\sigma N+1)} \sum_h (\delta B_{hw})^2$$

and using from (35) the relation $\sum_{h' \neq h} \delta B_{h'w} = -\delta B_{hw}$ we get

$$J = \sum_w s_w \sum_h (\delta B_{hw})^2 \left(\frac{(\sigma+1)}{N(\sigma N+1)} - \frac{\sigma}{N(\sigma N+1)} \right) = \frac{1}{N(\sigma N+1)} \sum_w s_w \sum_h (\delta B_{hw})^2 \cdot \quad (59)$$

For the fourth term we obtain

$$\left\langle \left(\frac{\sum_h \alpha_h \sum_w s_w \delta B_{hw}}{\sum_h \alpha_h} \right)^2 \right\rangle = \frac{1}{N(\sigma N+1)} \sum_h \left(\sum_w s_w \delta B_{hw} \right)^2 \cdot \quad (60)$$

Finally, expression (56) yields:

$$Q_s[\{B_{ij}\}] = N^{N\sigma} \frac{\Gamma(\sigma)^N}{\Gamma(\sigma N)} \left\{ 1 + \frac{N^2}{2(\sigma N+1)} \sum_{hw} s_w (\delta B_{hw})^2 + \frac{N^3}{2(\sigma N+1)} \sum_h \left(\sum_w s_w \delta B_{hw} \right)^2 + O\left((\delta B)^3 \right) \right\} \cdot \quad (61)$$

In principle we are able to calculate all higher terms. There are two other terms to be expanded: $\prod_{p,q=1}^{N} B_{pq}{}^{s_q-1}$ and $\det\left[\mathbb{1} - BB^{\mathsf{T}}\right]_{N-1}$. For the latter we have

$$\det\left[\mathbb{1} - BB^{\mathsf{T}}\right]_{N-1} = \det\left[\delta_{ik} - \sum_{j=1}^{N}\left(\frac{1}{N} + \delta B_{ij}\right)\left(\frac{1}{N} + \delta B_{kj}\right)\right]_{N-1} = \exp\left\{\ln\det\left[D_{ik} - \sum_{j=1}^{N}\delta B_{ij}\,\delta B_{kj}\right]_{N-1}\right\}. \quad (62)$$

In the last line we made use of equation (35) and we introduced the $(N-1)\times(N-1)$ *circulant* [24] matrix $D_{ik} := \delta_{ik} - 1/N$. As it can be verified by direct matrix multiplication, the inverse of D reads $D_{ik}^{-1} = \delta_{ik}+1$. Hence, factorizing the determinant of the product in the product of determinants, it follows from (62)

$$\det\left[\mathbb{1} - BB^{\mathsf{T}}\right]_{N-1} = \det\left[D\right]_{N-1} \times \det\left[\delta_{i\ell} - \sum_{j=1}^{N-1}\sum_{k=1}^{N}\left(\delta_{ij} + 1\right)\delta B_{jk}\,\delta B_{\ell k}\right]_{N-1}. \quad (63)$$

Observe that the index j labels the $(N-1)$ columns of the matrix $\left[D^{-1}\right]_{N-1}$, whereas k runs from 1 to N, since we are going to consider $\left[(\delta B)\,(\delta B)^{\mathsf{T}}\right]_{N-1}$ and not $\left[(\delta B)\right]_{N-1}\left[(\delta B)^{\mathsf{T}}\right]_{N-1}$. Using the property of circulant matrices [24], we can determine the spectrum of D, consisting of a simple eigenvalue $1/N$ and another one equal to 1, of multiplicity $(N-2)$. Thus $\det\left[D\right]_{N-1} = 1/N$ and eq. (35) and (63) yield

$$\det\left[\mathbb{1} - BB^{\mathsf{T}}\right]_{N-1} = \frac{1}{N} \times \det\left[\delta_{i\ell} - \sum_{k=1}^{N}\delta B_{ik}\,\delta B_{\ell k} + \sum_{k=1}^{N}\delta B_{Nk}\,\delta B_{\ell k}\right]_{N-1}. \quad (64)$$

From the identity $\det\left[\exp\left(A\right)\right] = \exp\left[\mathrm{Tr}\left(A\right)\right]$, with the substitution $A \leftarrow \log\left(\mathbb{1} + X\right)$ we get

$$\det\left(\mathbb{1} - X\right) = \exp\left\{\mathrm{Tr}\left[\log\left(\mathbb{1} - X\right)\right]\right\} = \exp\left\{\mathrm{Tr}\left[-X + O\left(X^2\right)\right]\right\}$$

$$= \exp\left\{-\mathrm{Tr}\left(X\right) + O\left[\mathrm{Tr}\left(X^2\right)\right]\right\} = 1 - \mathrm{Tr}\left(X\right) + \frac{[\mathrm{Tr}\left(X\right)]^2}{2} + O\left[\mathrm{Tr}\left(X^2\right)\right]$$

so that, choosing for X the (δB)'s contributions in equation (64), we get $\mathrm{Tr}\left(X\right) = \sum_{\ell,k=1}^{N}\delta B_{\ell k}\,\delta B_{\ell k}$ and therefore

$$\det\left[\mathbb{1} - BB^{\mathsf{T}}\right]_{N-1} = \frac{1}{N}\left\{1 - \sum_{\ell k}(\delta B_{\ell k})^2 + O\left((\delta B)^4\right)\right\}. \quad (65)$$

Finally we use the expansion

$$\prod_{p,q=1}^{N} B_{pq}{}^{s_q-1} = \exp\left\{\sum_{p,q=1}^{N}\left(s_q - 1\right)\ln\left(\frac{1}{N} + \delta B_{pq}\right)\right\} = \frac{1}{N^{N(\sigma-N)}} \times \left\{1 + \frac{N^2}{2}\sum_{p,q=1}^{N}\left(\delta B_{pq}\right)^2 - \frac{N^2}{2}\sum_{p,q=1}^{N}s_q\left(\delta B_{pq}\right)^2 + O\left((N\,\delta B)^3\right)\right\}. \quad (66)$$

Now, substituting (61) and (65–66) into (34), we obtain the final formula for the resulting probability distribution around the center of the Birkhoff polytope $\mathcal{B}_N$ given by ((36)).

Appendix C

In this appendix we provide the third order expansion of the probability distribution $P_s(B)$ at $B = B_\star$. The result obtained implies that it is not possible to find an ensemble of stochastic matrices characterised by the Dirichlet distribution,

which induces a distribution flat up to the third order at the center of the Birkhoff polytope. Furthermore, we provide an estimation, that is how the asymmetry of the optimal distribution around $B_\star$ changes with N.

For general s the output distribution behaves like $\widetilde{P}_s[\{B_{ij}\}] \propto \exp\left(\lambda \sum_{pq} \left(\delta B_{pq}\right)^2\right)$ at the center, with

$$\lambda = \frac{N^2}{2} - 1 - \frac{N^4 s^2}{2\left(N^2 s + 1\right)}. \tag{67}$$

From now on, symbols like $\widetilde{P}_s$, P_s, V_s, W_s denote the probability densities obtained from the input described by the string
$\boldsymbol{s} = \{s_1 = s, s_2 = s, \ldots, s_N = s\}$ consisting of N Dirichlet exponents equal. Since $\frac{d\lambda}{ds} < 0$, the distribution is Gaussian for $s > s^\star$.

In order to study the deviations from the Gaussian distribution, we now study the third order contribution to $\widetilde{P}_s[\{B_{ij}\}]$ of eq. (34), in the case $s_i = s$. Under the latter hypothesis, many terms of the kind $\sum_q s_q \delta B_{pq}$ vanish for (35) so such terms will be omitted.

The distribution (34) can be factorized into a product of three factors:

- $\prod_{p,q=1}^{N} B_{pq}{}^{s-1}$ gives a contribution

$$\frac{1}{N^{N(\sigma-N)}} \times \left[\frac{N^3}{3}(s-1)\sum_{p,q=1}^{N}\left(\delta B_{pq}\right)^3 + O\left((N\,\delta B)^4\right)\right]; \tag{68}$$

- $\det\left[\mathbb{1} - BB^{\mathsf{T}}\right]_{N-1}$ gives no 3$^{\text{rd}}$ order contribution (just the overall factor $1/N$ already present in (65));

- the integral $Q_s[\{B_{ij}\}]$ of (50) gives

$$\Delta_3 Q_s[\{B_{ij}\}] = N^{N\sigma} \int_0^\infty \cdots \int_0^\infty \left(\prod_{r=1}^{N} d\alpha_r\, \alpha_r^{\sigma-1}\right) \frac{\delta\left(\alpha_N - 1\right)}{\left(\sum_s \alpha_s\right)^{N\sigma}} \left\{-\frac{1}{3}Ns\sum_j\left(\frac{N\sum_i \alpha_i\,\delta B_{ij}}{\sum_k \alpha_k}\right)^3 + O\left((\delta B)^4\right)\right\}$$

$$= N^{N\sigma} \frac{\Gamma(\sigma)^N}{\Gamma(\sigma N)} \left\{-\frac{N^4 s}{3}\sum_j\left\langle\left(\frac{\sum_i \alpha_i\,\delta B_{ij}}{\sum_k \alpha_k}\right)^3\right\rangle + O\left((\delta B)^4\right)\right\}, \tag{69}$$

where we made use of the symbol $\langle\cdot\rangle$ introduced through eqs. (53–55).

Using the same reasoning as in Appendix B, including now the new contributions (68–69), we arrive at the 3$^{\text{rd}}$ order contribution for $\widetilde{P}_s[\{B_{ij}\}]$,

$$\Delta_3 \widetilde{P}_s[\{B_{ij}\}] = P_N^\star \left\{\frac{N^3}{3}(s-1)\sum_{p,q=1}^{N}\left(\delta B_{pq}\right)^3 - \frac{N^4 s}{3}\sum_j\left\langle\left(\frac{\sum_i \alpha_i\,\delta B_{ij}}{\sum_k \alpha_k}\right)^3\right\rangle\right\}. \tag{70}$$

The last term reads

$$\left\langle\left(\frac{\sum_i \alpha_i\,\delta B_{ij}}{\sum_k \alpha_k}\right)^3\right\rangle = \left\langle\frac{\alpha_1\,\alpha_2\,\alpha_3}{\left(\sum_k \alpha_k\right)^3}\right\rangle \sum_{\substack{\mu\neq\nu,\,\nu\neq\tau \\ \tau\neq\mu}} \delta B_{\mu j}\,\delta B_{\nu j}\,\delta B_{\tau j} + 3\left\langle\frac{\alpha_1^2\,\alpha_2}{\left(\sum_k \alpha_k\right)^3}\right\rangle \sum_{\mu\neq\tau}\left(\delta B_{\mu j}\right)^2\,\delta B_{\tau j} + \left\langle\frac{\alpha_1^3}{\left(\sum_k \alpha_k\right)^3}\right\rangle \sum_\mu\left(\delta B_{\mu j}\right)^3.$$

$$\tag{71}$$

It follows from (35), that $\sum_\tau \delta B_{\tau j} = 0 = \sum_{\tau \neq \mu} \delta B_{\tau j} + \delta B_{\mu j}$, so $\sum_{\tau \neq \mu} \delta B_{\tau j} = -\delta B_{\mu j}$ Multiplying this equality by $\left(\delta B_{\mu j}\right)^2$ and summing over μ one gets

$$\sum_{\mu \neq \tau} \left(\delta B_{\mu j}\right)^2 \delta B_{\tau j} = -\sum_\mu \left(\delta B_{\mu j}\right)^3 . \tag{72a}$$

Similarly

$$\left(\sum_\mu \delta B_{\mu j}\right)^3 = 0 = \sum_{\mu, \nu, \tau} \delta B_{\mu j} \delta B_{\nu j} \delta B_{\tau j} = \sum_{\substack{\mu \neq \nu, \nu \neq \tau \\ \tau \neq \mu}} \delta B_{\mu j} \delta B_{\nu j} \delta B_{\tau j} + 3 \sum_{\mu \neq \tau} \left(\delta B_{\mu j}\right)^2 \delta B_{\tau j} + \sum_\mu \left(\delta B_{\mu j}\right)^3$$

and using (72a) we arrive at

$$\sum_{\substack{\mu \neq \nu, \nu \neq \tau \\ \tau \neq \mu}} \delta B_{\mu j} \delta B_{\nu j} \delta B_{\tau j} = 2 \sum_\mu \left(\delta B_{\mu j}\right)^3 . \tag{72b}$$

Substituting eqs. (72) into (71) one obtains

$$\left\langle \left(\frac{\sum_i \alpha_i \delta B_{ij}}{\sum_k \alpha_k}\right)^3 \right\rangle = \left[\left\langle \frac{\alpha_1^3}{\left(\sum_k \alpha_k\right)^3} \right\rangle - 3 \left\langle \frac{\alpha_1^2 \alpha_2}{\left(\sum_k \alpha_k\right)^3} \right\rangle + 2 \left\langle \frac{\alpha_1 \alpha_2 \alpha_3}{\left(\sum_k \alpha_k\right)^3} \right\rangle \right] \sum_\mu \left(\delta B_{\mu j}\right)^3 .$$

Now we use (55)

$$\left\langle \left(\frac{\sum_i \alpha_i \delta B_{ij}}{\sum_k \alpha_k}\right)^3 \right\rangle = \left\{ \frac{\Gamma(\sigma + 3)}{\Gamma(\sigma)} - 3 \frac{\Gamma(\sigma + 2)\Gamma(\sigma + 1)}{[\Gamma(\sigma)]^2} + 2 \left[\frac{\Gamma(\sigma + 1)}{\Gamma(\sigma)}\right]^3 \right\} \frac{\Gamma(N\sigma)}{\Gamma(N\sigma + 3)} \sum_\mu \left(\delta B_{\mu j}\right)^3$$

$$= \frac{2}{N(N\sigma + 1)(N\sigma + 2)} \sum_\mu \left(\delta B_{\mu j}\right)^3 . \tag{73}$$

Thus, from (70), the third order contribution to $\widetilde{P}_s[\{B_{ij}\}]$ is

$$\Delta_3 \widetilde{P}_s[\{B_{ij}\}] = P_N^\star \left\{ \frac{(s-1)N^3}{3} - \frac{2N^3 s}{3(N\sigma + 1)(N\sigma + 2)} \right\} \sum_{pq} \left(\delta B_{pq}\right)^3 \tag{74}$$

and, near the center $B_N^\star$, $\widetilde{P}_s[\{B_{ij}\}]$ has the following structure:

$$\widetilde{P}_s[\{B_{ij}\}] = P_N^\star \exp \left\{ -c_2 \sum_{pq} \left(\delta B_{pq}\right)^2 - c_3 \sum_{pq} \left(\delta B_{pq}\right)^3 + O\left[(\delta B)^4\right] \right\} . \tag{75}$$

Assuming that $s_i = s$ (so $\sigma = Ns$) we may then find from (36) the value of the constant c_2,

$$c_2 = 1 - \frac{N^2}{2} + \frac{\sigma^2 N^2}{2(\sigma N + 1)} . \tag{76a}$$

Similarly eq. (74) implies that the third constant reads,

$$c_3 = \frac{N^2}{3}\left(N - \sigma + \frac{2\,\sigma}{(N\sigma + 1)(N\sigma + 2)}\right). \tag{76b}$$

Adjusting $s = \sigma/N$ appropriately to the size N of the matrix one may find such a value of the Dirichlet parameter s that c_2 or c_3 are equal to zero. However, if we set c_2 to zero, the parameter c_3 is non zero, so the third order terms remain in eq. (75). Thus we have shown that it is not possible to find an initial Dirichlet distribution which gives the output distribution uniform in the vicinity of the center of the Birkhoff polytope up to the third order. A power expansion of c_3 gives

$$c_3 = \frac{N}{3} + \frac{1}{N} - \frac{4}{3}\left(\frac{1}{N}\right)^3 + O\left[\left(\frac{1}{N}\right)^5\right], \tag{77}$$

Thus the scale of the asymmetry is $\delta B \propto N^{-1/3}$ so it cannot be seen for $|\delta B| \lesssim N^{-1/3}$ that means if $\left|\delta B/B_N^\star\right| \lesssim N^{2/3}$.

References

[1] A. W. Marshall and I. Olkin, *Inequalities: theory of majorization and its applications*, Academic Press, New York, NY, 1979.

[2] R. Bhatia, *Matrix analysis*, Springer–Verlag, New York, NY, 1997.

[3] J. E. Cohen, Kemperman J. H. B. and G. Zbaganu, *Comparisons of stochastic matrices, with applications in information theory, statistics, economics, and population sciences*, Birkhäuser, Boston, MA, 1998.

[4] R. A. Brualdi *Combinatorial matrix classes*, Encyclopedia of mathematics and its applications, Cambridge University Press, 2006.

[5] Chan C.S. and Robbins D.P., On the volume of the polytope of doubly stochastic matrices, *Exp. Math.* **8(3)**, 291–300 (1999).

[6] M. Beck and D. Pixton, The Ehrhart Polynomial of the Birkhoff Polytope, *Discrete & Computational Geometry* **30(4)**, 623–637 (2003).

[7] J. A. De Loera, F. Liu and R. Yoshida, Formulas for the volumes of the polytope of doubly–stochastic matrices and its faces, `arXiv:math/0701866v1` [math.CO], 2007.

[8] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*, Cambridge University Press, Cambridge, 2006.

[9] H. H. Bauschke and J.M. Borwein, *SIAM Rev.* **38**, 367-426 (1996).

[10] R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices, Ann. Math. Stat. **35**, 876-879 (1964).

[11] R. Sinkhorn and P. Knopp, Concerning non-negative matrices and doubly stochastic matrices, Pacific J. Math. **21(2)**, 343-348 (1967).

[12] E. R. Canfield and B. D. McKay, The asymptotic volume of the Birkhoff polytope, `arXiv:0705.2422v1` [math.CO], 2007.

[13] K. M. R. Audenaert and S. Scheel, On random unitary channels, *N. J. Phys.* **10**, 023011 (2008).

[14] D. Djoković, Note on nonnegative matrices, Proc. Amer. Math. Soc. **25**, 82-90 (1970).

[15] P. B. Slater, A priori probabilities of separable quantum states, *J. Phys. A: Math. Gen.* **32(28)**, 5261–5275 (1999).

[16] R. R. Tucci, Entanglement of Formation and Conditional Information Transmission, preprint quant-ph/0010041 (2000).

[17] K. Życzkowski and H.-J. Sommers, Induced measures in the space of mixed quantum states, *J. Phys. A: Math. Gen.* **34(35)**, 7111–7125 (2001).

[18] L. Devroye, *Non–Uniform Random Variate Generation*, Springer–Verlag, New York, Berlin, Heidelberg, Tokyo, 1986.

[19] Mathematica 5.2, Wolfram Research, Inc., 2005.

[20] W. Słomczyński, K. Życzkowski, M. Kuś, and H.-J. Sommers, Random unistochastic matrices, *J. Phys. A: Math. Gen.* **36(12)**, 3425–3450 (2003).

[21] S. Barnett, *Matrices: methods and applications*, Oxford applied mathematics and computing science series, Clarendon Press, Oxford, 1990.

[22] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, Classics in applied mathematics. v.9., Ann. of Phys., Philadelphia, 1979.

[23] D. S. Bernstein, *Matrix mathematics: theory, facts, and formulas with application to linear systems theory*, Princeton University Press, Princeton, 2005.

[24] J. Hofbauer and K. Sigmund, *Evolutionary games and population dynamics*, Cambridge University Press, Cambridge, 1998.

[25] W. Bruzda, V. Cappellini, H.-J. Sommers and K. Życzkowski, Random quantum operations, *Phys. Lett.* **A 373**, 320-324 (2009).