Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

The ensemble of random Markov matrices

Martin Horvat

Physics Department, Faculty of Mathematics and Physics,
University of Ljubljana, Slovenia
and
CAMTP—Centre for Applied Mathematics and Theoretical Physics,
University of Maribor, Krekova 2, SI-2000 Maribor, Slovenia
E-mail: martin.horvat@fmf.uni-lj.si

Received 10 February 2009
Accepted 4 May 2009
Published 1 July 2009

Online at stacks.iop.org/JSTAT/2009/P07005
doi:10.1088/1742-5468/2009/07/P07005

Abstract. The ensemble of random Markov matrices is introduced as a set of Markov or stochastic matrices with the maximal Shannon entropy. The statistical properties of the stationary distribution π, the average entropy growth rate h and the second-largest eigenvalue ν across the ensemble are studied. It is shown and heuristically proven that the entropy growth rate and second-largest eigenvalue of Markov matrices scale on average with the dimension of the matrices d as $h \sim \log(\mathrm{O}(d))$ and $|\nu| \sim d^{-1/2}$, respectively, yielding the asymptotic relation $h\tau_\mathrm{c} \sim 1/2$ between the entropy h and the correlation decay time $\tau = -1/\log|\nu|$. Additionally, the correlation between h and τ_c is analysed; it decreases with increasing dimension d.

Keywords: dynamical processes (theory), dynamical processes (experiment), stochastic processes (theory), stochastic processes (experiment)

 1742-5468/09/P07005+14$30.00

Contents

J. Stat. Mech. (2009) P07005

1. Introduction

In information theory and mathematical modelling of physical processes we often stumble upon Markov chains [1] and Markov or stochastic matrices [2], which determine the evolution of the former. Let us assume that a Markov chain is based on the set of states $\mathcal{S} = \{s_i\}_{i=1}^d$ and $P(s_j|s_i)$ is the conditional probability for a transition from state s_i to s_j; then the corresponding Markov matrix $M \in \mathbb{R}_+^{d\times d}$ is a collection of conditional probabilities

$$M_{i,j} = P(s_j|s_i) \geq 0, \qquad \sum_{j=1}^{d} M_{i,j} = 1,$$

where d is the dimension of the Markov matrix. Notice that the sum of elements in each row is normalized to 1. The applications of Markov matrices and their construction are very diverse. Particularly interesting is their use in dynamical systems for giving a probabilistic description of the dynamics. For a general introduction in this direction see e.g. [3]. For example, let us consider a discrete dynamical system $\phi^t\colon X \to X$, where $t \in \mathbb{N}$, with the phase space X and the invariant measure μ. By choosing disjoint subsets of phase space $\{X_i \subset X\colon X_i \cap X_j = 0 \text{ for } i \neq j\}$, which satisfy $\bigcup_i X_i = X$, the Markov matrix $M = [M_{i,j}]_{i,j=1}^d$ corresponding to the dynamical system can be defined as

$$M_{i,j} = \frac{\mu(\phi(X_i) \cap X_j)}{\mu(X_i)},$$

and describes a single time step of the dynamical system. In this way a paradigmatic example of a dynamical system with an algebraic decay of correlation—the triangle map [4]—was examined in [5]. Besides the method presented for constructing a Markov matrix and other methods producing matrices with specific properties, there is often a need for a way to construct a Markov matrix ad hoc, i.e. without incorporating any information about the system except the number of states d. If the construction procedure is a stochastic process, then the resulting matrix is called *the random Markov matrix* and the set of such matrices form *the ensemble of random Markov matrices*. These matrices are usually used, without much theoretical background, for testing purposes. For example testing of algorithms or certain statistical hypotheses, where applications in the field of dynamical systems, connected to ergodicity and mixing properties, are the most interesting to us. In information theory, random Markov matrices are used to test the algorithms for recognition or attribution processes, compression algorithms etc.

The present work is strongly related to the work of [6] discussing the general properties of the Markov ensemble and to [7] and [8], where a closer look at the second-largest (sub-dominant) eigenvalue of random Markov matrices was taken. In contrast to the past works, ours tries to emphasize the physical application of results, in particular in the field of dynamical systems. Interestingly, recently an ensemble of random Markov matrices was applied in the spectral analysis of the random quantum super-operators [9].

2. Preliminaries

The set of all possible Markov matrices M of dimension d is defined as

$$\mathcal{M}(d) = \{M \in \mathbb{R}_+^{d\times d}\colon M\underline{1} = \underline{1}\}, \qquad \underline{1} = (1,\ldots,1) \in \mathbb{R}^d,$$

and it is isomorphic to the direct product of d convex sets

$$\bigotimes^{d} \{x \in \mathbb{R}_+^d\colon x^{\mathrm{T}}\underline{1} = 1\}.$$

The set $\mathcal{M}(d)$ forms together with the matrix product a semi-group, whereas the set of non-singular Markov matrices form a group of stochastic matrices. The ensemble of random Markov matrices is defined as a set of Markov matrices $\mathcal{M}(d)$ with the probability measure of matrix elements $M_{i,j} \in \mathbb{R}_+$ reading

$$\mathrm{d}P(M) = [(d-1)!]^d \delta^d(M\underline{1} - \underline{1})\mathrm{d}M, \qquad \mathrm{d}M := \prod_{i,j=1}^{d} \mathrm{d}M_{i,j}, \tag{1}$$

which incorporates minimal information about the set, i.e. only constraints due to the probability conservation in the Markov process. The ensemble of random Markov matrices, denoted by the pair $(\mathrm{d}P(M), \mathcal{M}(d))$, is also referred to as the Dirichlet ensemble [6] and corresponding matrices are called doubly (row and column) stochastic or bi-stochastic Markov matrices [2].

The rows of the Markov matrix from the ensemble $(\mathrm{d}P(M), \mathcal{M}(d))$ are independent random vectors $X = (X_i \geq 0)_{i=1}^d$ with the distribution

$$P_{\mathrm{rows}}(X) = (d-1)!\, \delta(\underline{1}^{\mathrm{T}}X - 1). \tag{2}$$

It can be rather awkward to numerically generate components of vector-rows X directly and so a different approach for doing that is taken. By following the notes on the exponential distribution in [10] (p 76), we find that the vector-rows X of Markov matrices can be expressed using vectors $Y = (Y_i)_{i=1}^d$ of independent variables Y_i with a common exponential distribution in the following way:

$$X = \frac{Y}{\underline{1}^{\mathrm{T}}Y}. \tag{3}$$

In this way we numerically generate the pseudo-random rows in Markov matrices, where each row is generated independently. Consequently, the distribution of the rows can be written as

$$P_{\mathrm{rows}}(X) = \int_{\mathbb{R}_+^d} \mathrm{d}^d Y \delta^d \left(X - \frac{Y}{\underline{1}^{\mathrm{T}} Y} \right) \mathrm{e}^{-\underline{1}^{\mathrm{T}} Y}. \tag{4}$$

This identity of expressions is checked by calculating the moment-generating function of the expression above and expression (2), yielding the same result equal to

$$\int_{\mathbb{R}_+^d} \mathrm{d}^d X \mathrm{e}^{-\lambda^{\mathrm{T}} X} P_{\mathrm{rows}}(X) = (-1)^{d-1}(d-1)! \sum_{i=1}^{d} \frac{\mathrm{e}^{-\lambda_i}}{w'(\lambda_i)},$$

where $\lambda = (\lambda_i)_{i=1}^d$ and $w(x) = \prod_{i=1}^d (x - \lambda_i)$. Let us denote the sum of variables Y_i by $S = \underline{1}^{\mathrm{T}} Y$ and examine its statistics. The ratio between the standard deviation $\sigma_S = \sqrt{\langle (S - \langle S \rangle)^2 \rangle}$ of S and its average value $\langle S \rangle$ is equal to $\sigma_S / \langle S \rangle = d^{-1/2}$ and it is decreasing with increasing dimension d. This means that the renormalization of variables Y_i in the expression (3) has less and less effect on the functional dependence between X and Y as d is increased. We conclude that in the limit for large dimensions $d \gg 1$, variables X_i are approximately independent and exponentially distributed with distribution $P_{\mathrm{m}}(X_i) = d \exp(-d X_i)$. Following this idea we write the asymptotic approximation of the probability measure of Markov ensemble as

$$\mathrm{d}P(M) \sim \mathrm{d}P_{\mathrm{asym}}(M) = d^{d^2} \mathrm{e}^{-d \underline{1}^{\mathrm{T}} M \underline{1}} \mathrm{d}M, \tag{5}$$

which is valid in the limit $d \to \infty$. It can be verified that the averages w.r.t. the distributions $\mathrm{d}P(M)$ and $\mathrm{d}P_{\mathrm{asym}}(M)$ of a well behaved observable on $\mathcal{M}$, which depends only on a finite number of matrix elements, are asymptotically equal.

The probability measure of the Markov ensemble $\mathrm{d}P(M)$ has several unique properties that makes the defined ensemble of random Markov matrices $(\mathrm{d}P(M), \mathcal{M}(d))$ interesting and potentially useful.

For instance the probability measure $\mathrm{d}P(M)$ has a maximal Shannon entropy and in this information sense it is unique. The set of Markov matrices is merely a direct product of planes restricted to $\mathbb{R}_+^d$, and $\mathrm{d}P(M)$ is uniform on them. The Shannon entropy of the measure $\mathrm{d}P(M)$ on the set $\mathcal{M}(d)$ is just the sum of Shannon entropies of the uniform distribution on the planes, which are themselves minimal. Hence the Shannon entropy of $\mathrm{d}P(M)$ is also minimal. Any modification of the measure would necessarily increase the Shannon entropy and therefore it is unique.

It is also interesting to notice that the measure $\mathrm{d}P(M)$ is not invariant w.r.t. matrix multiplication. However, for a given non-singular Markov matrix $A \in \mathcal{M}(d)$, the measure $\mathrm{d}P(M)$ is invariant under matrix multiplication up a constant

$$P(A\mathcal{B}) = |\det(A)|^{-d} P(\mathcal{B}) \qquad \forall \mathcal{B} \subset \mathcal{M}(d).$$

In fact there is no measure of Markov matrices with the matrix elements $M_{i,j}$ approximately independent in the limit $d \gg 1$ which would be invariant w.r.t. matrix multiplication. To show this let us consider two large matrices $A = [A_{i,j}]_{i,j=1}^d$ and $B = [B_{i,j}]_{i,j=1}^d$ with the matrix elements being i.i.d. variables with the distribution $P_{\mathrm{m}}(x)$. Here we denote the ith central moment of some distribution $Q(x)$ for $i > 1$ as $\mu_i(Q) = \int \mathrm{d}x\, (x - \mu_1(Q))^i Q(x)$ and for $i = 1$ by $\mu_1(Q) = \int \mathrm{d}x\, Q(x) x$. We assume that

J. Stat. Mech. (2009) P07005

the first three central moments of $P_{\rm m}(x)$ are finite. The matrix elements of the product $AB = [C_{i,j} = \sum_{k=1}^{d} A_{i,k}B_{k,j}]_{i,j=1}^{d}$ are distributed as

$$P_{\rm C}(x) = \underbrace{P_{\rm AB} * \cdots * P_{\rm AB}}_{d}(x), \qquad P_{\rm AB}(x) = \int_{\mathbb{R}_+} \frac{{\rm d}a}{a} P_{\rm m}(a)\, P_{\rm m}\left(\frac{x}{a}\right) {\rm d}a,$$

where the sign $*$ denotes the convolution and $P_{\rm AB}$ is the distribution of the product $A_{i,j}C_{j,k}$ with the first two central moments reading

$$\mu_1(P_{\rm AB}) = \mu_1(P_{\rm m})^2 \qquad \text{and} \qquad \mu_2(P_{\rm AB}) = 2\left(\mu_1(P_{\rm m})^2 + \mu_2(P_{\rm m})\right)\mu_2(P_{\rm m}).$$

According to the central limit theorem (CLT) [10] the distribution $P_{\rm C}(x)$ converges in the limit $d \to \infty$ to a Gaussian distribution with the first two central moments equal to $\mu_1(P_{\rm C}) = d \cdot \mu_1(P_{\rm AB})$ and $\mu_2(P_{\rm C}) = d \cdot \mu_2(P_{\rm AB})$. For distribution $P_{\rm m}(x)$ to be invariant w.r.t. to the matrix multiplication it has to be asymptotically equal to $P_{\rm C}(x)$, meaning that $P_{\rm m}(x)$ is also a Gaussian distribution for large dimension $d \gg 1$. By comparing the first two central moments of $P_{\rm C}(x)$ and $P_{\rm m}(x)$ we find that the average value of matrix elements of the Markov matrix and their variance are asymptotically equivalent to

$$\mu_1(P_{\rm m}) \sim \frac{1}{d} \qquad \text{and} \qquad \mu_2(P_{\rm m}) \sim \frac{1}{d} - \frac{2}{d^2},$$

respectively. The ratio between the standard deviation and the average scales with dimension as $\sqrt{\mu_2(P_{\rm m})}/\mu_1(P_{\rm m}) = {\rm O}(d^{1/2})$ and diverges in the limit $d \to \infty$. This indicates that a measure of Markov matrices by which the matrix elements are asymptotically independent and distributed via $P_{\rm m}(x)$ does not exist.

In the following we discuss the properties of random Markov matrices from the Markov ensemble $({\rm d}P(M), \mathcal{M})$. We focus on the entropy growth rate and correlation decay in the Markov chains generated by these Markov matrices, and examine their asymptotic behaviour for $d \gg 1$.

3. The entropy and stationary distribution of the random Markov matrices

We consider a Markov chain defined on the set of states $\mathcal{S} = \{s_i\}_{i=1}^{d}$ and with the conditional probabilities $P(s_j|s_i) = M_{i,j}$ given in the Markov matrix $M = [M_{i,j}]_{i,j=1}^{d}$. The initial probability distribution over the states is $(P(s_i))_{i=1}^{d}$. The probability that the Markov chain has evolved up to time t following a specific route $(e_1, \ldots, e_t) \in \mathcal{S}^t$ is given with the product of conditional probabilities reading

$$P(e_1, \ldots, e_t) = P(e_1)P(e_2|e_1)P(e_3|e_2)\cdots P(e_t|e_{t-1}).$$

Then the dynamic entropy S of the Markov chain at time t is given by the sum

$$S(t) = -\sum_{e \in \mathcal{S}^t} P(e) \log P(e),$$

taken over all different routes up to time t. In ergodic Markov chains we expect that the entropy in the limit $t \to \infty$ increases linearly with increasing time t as

$$S(t) \sim ht,$$

where $h \in \mathbb{R}$ denotes the asymptotic entropy growth rate of the Markov chain. The entropy growth rate is given by the formula [11]

$$h = -\sum_{i=1}^{d} \pi_i \sum_{j=1}^{d} M_{i,j} \log M_{i,j}, \tag{6}$$

where we use the stationary distribution $\pi = (\pi_i \geq 0)_{i=1}^d$ defined as the eigenvector of the Markov matrix corresponding to the unit eigenvalue,

$$\pi^{\mathrm{T}} M = \pi^{\mathrm{T}}, \qquad \sum_{i=1}^{d} \pi_i = 1. \tag{7}$$

In the following we discuss the distribution $P_\pi(x)$ of elements π_i corresponding to a stationary distribution π of a random Markov matrix $M = [M_{i,j}]_{i,j=1}^d$. In particular we are interested in the asymptotic limit as $d \to \infty$, where the matrix elements $M_{i,j}$ are approximately independent variables with an exponential distribution $P_{\mathrm{m}}(M_{i,j}) = d\exp(-d\, M_{i,j})$. Further we assume that p_i and $M_{i,j}$ have no correlations. On doing this the eigenvalue equation (7) written in components $\pi_i = \sum_j \pi_j M_{j,i}$ can be translated into an asymptotic self-consistent condition for the distribution $P_\pi(x)$ reading

$$P_\pi(x) \sim \underbrace{P_{\pi\mathrm{M}} * \cdots * P_{\pi\mathrm{M}}}_{d}(x) \qquad \text{as } d \to \infty, \tag{8}$$

with the distribution $P_{\pi\mathrm{M}}(x)$ of the products $p_j M_{j,i}$ depending again on distribution $P_\pi(x)$ and written as

$$P_{\pi\mathrm{M}}(x) = d \int_{\mathrm{R}_+} \frac{\mathrm{d}b}{b} \exp\left(-d\frac{x}{b}\right) P_\pi(b),$$

where $*$ denotes the convolution. Assuming that the first three central moments of the distribution $P_{\pi\mathrm{M}}(x)$ are finite, we can use the CLT and state that the distribution $P_\pi(x)$ converges to a Gaussian distribution as $d \to \infty$. By inserting the ansatz

$$P_\pi(x) = \frac{1}{\sqrt{2\pi\sigma_\pi^2}} \exp\left(-\frac{(x-\overline{\pi})^2}{2\sigma_\pi^2}\right), \qquad \overline{\pi} = \frac{1}{d},$$

into equation (8) and imposing the asymptotic equality (8) we obtain the variance $\sigma_\pi^2 \sim d^{-3}$ of the elements π_i. On appropriately rescaling the coefficients π_i, their cumulative distribution is independent of dimension d in the limit $d \to \infty$ and reads

$$\mathrm{Prob}\left(\frac{d\pi_i - 1}{\sqrt{2}} < \frac{x}{\sqrt{d}}\right) \sim G(x) = \frac{1}{2}\left(\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right) + 1\right). \tag{9}$$

This result is compared in figure 1 with numerically obtained distributions of rescaled coefficients π_i for different large dimensions d and we find a very good agreement. We continue the analysis of the entropy growth rate h (6) of an typical random Markov matrix from the ensemble by decomposing it into an average term h_{ave} and an oscillating term h_{osc} reading

$$h = h_{\mathrm{ave}} + h_{\mathrm{osc}}, \qquad h_{\mathrm{ave}} = \frac{1}{d}\sum_{i=1}^{d} U_i, \qquad h_{\mathrm{osc}} = \sum_{i=1}^{d}\left(\pi_i - \frac{1}{d}\right)(U_i - h_{\mathrm{ave}}),$$

J. Stat. Mech. (2009) P07005

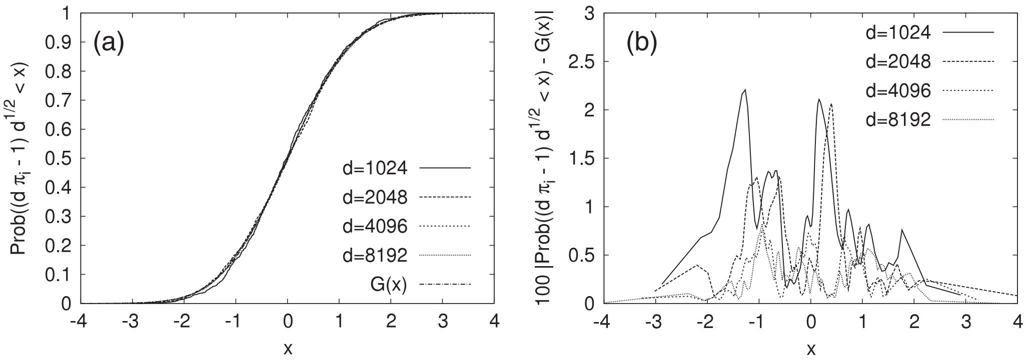

Figure 1. The cumulative distribution of the rescaled coefficients π_i corresponding to a stationary distribution $\pi = (\pi_i)_{i=1}^{d}$ (a) and its deviation from the expected limiting form $G(x)$ (9) (b) calculated for individual random Markov matrices of different dimensions d taken from the ensemble.

where we introduce auxiliary variables

$$U_i = -\sum_{j=1}^{d} M_{i,j} \log M_{i,j}.$$

In the asymptotic limit $d \to \infty$ the variables U_i have, according to the CLT, a Gaussian distribution with the first two central moments reading

$$\langle U_i \rangle \sim \log(\mathrm{e}^{\gamma-1} d),$$
$$\sigma_U^2 \sim \left[1 + (\gamma - 4)\gamma + \pi^2/3 + (2\gamma - 4 + \log d)\log d\right] \frac{1}{d} = \mathrm{O}(d^{-1}),$$

with γ being the Euler constant. The average $\langle\cdot\rangle$ in the expressions above is taken w.r.t. the asymptotic distribution $P_{\mathrm{asym}}(M)$ (5). It is easy to see that the average term converges with increasing d to $\langle U_i \rangle$ as

$$h_{\mathrm{ave}} = \langle U_i \rangle + \mathrm{O}(d^{-1/2}),$$

where the last term on the rhs denotes the statistical deviation. The oscillating term h_{osc} can be treated as a scalar product of vectors $(\pi_i - 1/d)_{i=1}^{d}$ and $(U_i - h_{\mathrm{osc}})_{i=1}^{d}$ and by applying the Schwarz–Cauchy inequality it be bounded from above:

$$h_{\mathrm{osc}}^2 \le \sum_{i=1}^{d} \left(\pi_i - \frac{1}{d}\right)^2 (U_i - h_{\mathrm{ave}})^2 = \mathrm{O}(d^{-2}).$$

The last term on the rhs denotes again the statistical deviation obtained by taking into account the statistical deviations $\langle(\pi_i - 1/d)^2\rangle = \mathrm{O}(d^{-3})$ and $\langle(U_i - h_{\mathrm{ave}})^2\rangle = \mathrm{O}(d^{-1})$. We conclude from this analysis that h behaves over the ensemble, in the leading order in d, approximately as the Gaussian variable h_{ave} or U_i. This is supported by the numerical results in figure 2(a), where we show that the cumulative distribution of $(h - \langle h \rangle)/\sigma_h$ converges to the cumulative normal distribution $G(x)$ (9). By averaging h over the ensemble, the fluctuations from $\langle U_i \rangle$ are eliminated and we obtain the asymptotic equivalence

$$\langle h \rangle \sim \log(\mathrm{e}^{\gamma-1} d) \qquad \text{as } d \to \infty.$$

J. Stat. Mech. (2009) P07005

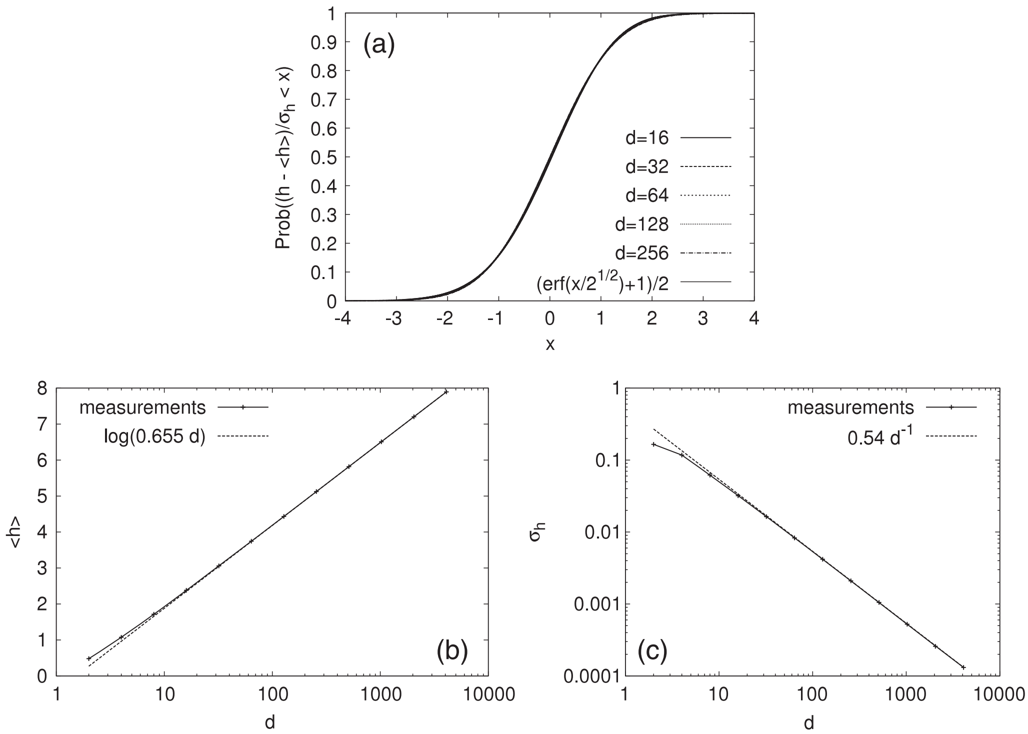

Figure 2. The cumulative distribution of the entropy growth rate h of random Markov matrices for different dimensions d (a), and the average entropy growth rate $\langle h \rangle$ (b) and its standard deviation σ_h (c) as a function of dimension d calculated using $N = 10^6$ random Markov matrices from the ensemble.

The latter agrees very well with the numerically obtained $\langle h \rangle$ as we can see in figure 2(b). This result supports the rule of thumb saying that in order to describe a dynamical system with an entropy rate h accurately via a Markov process we need to have a Markov matrix of the size $d \sim \mathrm{e}^h$. The standard deviation σ_h scales as $\mathrm{O}(d^{-1})$, as seen in figure 2(c), and this agrees with the expected statistical deviation of the form $\sqrt{(\sigma_U^2 + h_{\mathrm{osc}}^2)/d}$ in our asymptotic approximation.

4. The correlation decay induced by random Markov matrices

A state of a Markov chain defined using a Markov matrix $M = [M_{i,j}]_{i,j=1}^d \in \mathbb{R}_+^{d\times d}$ is described by a probability distribution

$$p = (p_i)_{i=1}^d \in \mathbb{R}_+^d, \qquad \sum_{i=1}^d p_i = 1,$$

over a given set of states $\{s_i\}_{i=1}^d$. Some initial probability distribution $p \in \mathbb{R}_+^d$ is evolved in time to $p(t)$ by the Markov matrix in the following way:

$$p(t)^{\mathrm{T}} = p^{\mathrm{T}} M^t,$$

where $t \in \mathbb{N}_0$ denotes the discrete time. We find that a Markov chain generated by a typical random Markov matrix M is mixing and consequently ergodic [1]. We assume

J. Stat. Mech. (2009) P07005

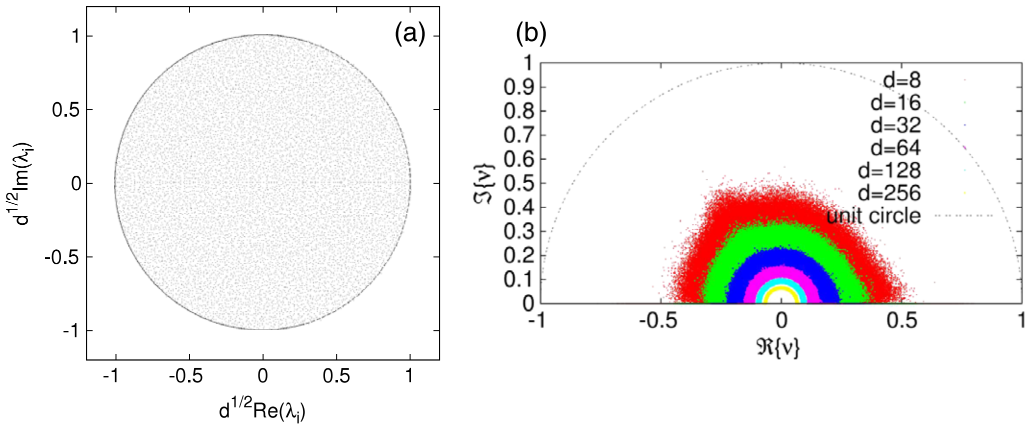

Figure 3. The spectrum of a random Markov matrix of dimension $d = 10^4$ (a) without the eigenvalue 1, which corresponds to the stationary distribution, and the second-largest eigenvalues $\nu \in \mathbb{C}$ (b) of approximately $N = 10^6$ random Markov matrices calculated for different dimensions d.

that the measure of Markov matrices in the ensemble corresponding to a non-mixing Markov chain is zero.

The discrete analogue of the time correlation function $C_{f,g}(t)$ between two real observables $f = (f_i \in \mathbb{R})_{i=1}^d$ and $g = (g_i \in \mathbb{R})_{i=1}^d$ is defined as

$$C_{f,g}(t) = \langle f_i, (g(t))_i \rangle_i - \langle f_i \rangle_i \langle g_i \rangle_i,$$

where we introduce a time propagated observable $g(t)^{\mathrm{T}} = g^{\mathrm{T}} M^t$ and averaging over the stationary distribution $\langle u_i \rangle_i = \sum_i \pi_i u_i$. The second-largest eigenvalue (called also the sub-dominant eigenvalue) of the Markov matrix $\nu \in \mathbb{C}$ determines the decay of correlation between almost all pairs of observables (f, g) following the formula

$$|C_{f,g}(t)| = \mathrm{O}(|\nu|^t) = \mathrm{O}(\mathrm{e}^{-t/\tau_{\mathrm{c}}}) \qquad \text{as } t \to \infty,$$

with $\tau_{\mathrm{c}} = -\log|\nu|$ called the correlation decay time. It is important to notice that the spectrum $\Lambda = \{\lambda\colon\ \det(M - \lambda \mathrm{id.}) = 0\}$ of a Markov matrix M has the symmetry

$$\Lambda^* = \Lambda,$$

where $(\cdot)^*$ represents the complex conjugation. The symmetry can be noticed in figure 3, where we show a spectrum of a typical random Markov matrix in the complex plane. In the limit of large dimensions $d \gg 1$ the eigenvalues are distributed symmetrically around the origin with the constant distribution of the square absolute value of the form

$$\mathrm{Prob}(x \le \|\lambda\|^2 < x + \mathrm{d}x) \approx \mathrm{O}(d^{-1/2})\,\mathrm{d}x.$$

This feature is along the lines of Girko's circular law [12] and its generalizations [13], but this particular case is not yet proved to the best of our knowledge. Here we are mainly interested in the second-largest eigenvalues $\nu \in \mathbb{C}$ of the random Markov matrices. These are depicted for $N = 10^6$ matrices sampled uniformly across the ensemble in figure 3(b) for several different dimensions d. For large d the values of ν are distributed radially symmetrically around the origin with the average radius and dispersion decreasing with increasing d. Further we examine the distribution of the magnitudes of the second-largest

J. Stat. Mech. (2009) P07005

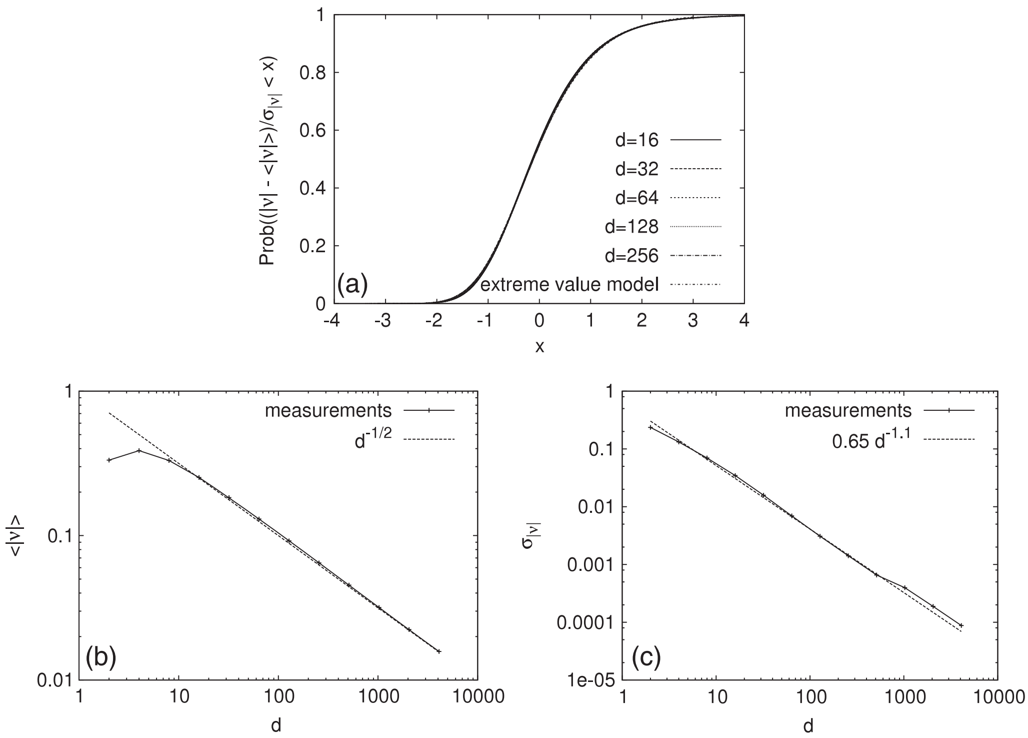

Figure 4. The cumulative distribution of the magnitude of the second-largest eigenvalue $|\nu|$ of random Markov matrices at different dimensions d (a) and the average $\langle|\nu|\rangle$ (b) and the standard deviation of the amplitude $\sigma_{|\nu|}$ (b) as a function of the dimension d calculated from approximately 10^6 and 10^3 random Markov matrices from the ensemble for $d < 512$ and $d \geq 512$, respectively.

eigenvalue $|\nu|$ denoted by $P_{|\nu|}$ and its first two central moments: average magnitude $\langle|\nu|\rangle$ and standard deviation $\sigma_{|\nu|}$. The cumulative distribution of the rescaled magnitude

$$\xi = \frac{|\nu| - \langle\nu\rangle}{\sigma_{|\nu|}}$$

is depicted in figure 4(a). From the figure we conclude that the distribution of ξ is basically independent of dimension d for large d and we find that it agrees well with the extreme value statistics of type 1 (Gumbel) [14]. Let us assume x_i are i.i.d. standard Gaussian variables. Then the maximal value of d variables $y = \max\{x_i\}_{i=1}^d$ is distributed according to the cumulative distribution

$$P_{\max}(y, d) = [G(y)]^d,$$

where we use the cumulative Gaussian distribution $G(y)$ (9). It is known that under simple linear transformation of the variable y, which depends on d, the distribution of transformed y converges in the limit $d \to \infty$ to the Gumbel or double-exponential distribution. To avoid certain problems of slow convergence towards the limiting distribution outlined in [15] we compare in figure 4(a) the numerically obtained distribution directly with $P_{\max}((y - \overline{y})/\sigma_y, d)$ for several large enough d, where $\overline{y}$ and σ_y are the average maximal value and its standard deviation, respectively. We find a very good agreement, suggesting

J. Stat. Mech. (2009) P07005

that the eigenvalues of the Markov matrix behave as i.i.d. random complex variables inside some disc in the complex plane with the radius $\mathrm{O}(d^{-1/2})$. The first two central moments of numerical results as a function of d are shown in figures 4(b) and (c). The average magnitude of the second-largest eigenvalue $\langle|\nu|\rangle$ fits very well to the asymptotic formula found empirically:

$$\langle|\nu|\rangle \sim C_0\, d^{-1/2}, \qquad d \to \infty, \tag{10}$$

where $C_0 \approx 1$. This result supports the conjecture that Girko's circular law is valid for the random Markov matrices stated in [6] and it can be better understood through the asymptotics of the upper bound for $|\nu|$ obtained in the following. By taking into account that all left-hand eigenvectors except the stationary distribution π are perpendicular to the vector $\underline{1}$, we can upper bound the second-largest eigenvalue $|\nu|$ as

$$|\nu|^2 \le \limsup_{x\in\mathcal{S}} \|x^\dagger M\|_2^2, \qquad \mathcal{S} = \{x \in \mathbb{C}^d\colon\, \|x\|_2 = 1 \wedge x \perp \underline{1}\}.$$

Here we write the expression $\|x^\dagger M\|_2^2 = x^\dagger N x$ using the matrix $N = MM^{\mathrm{T}}$. In the asymptotic limit the matrix N takes the form

$$N_{i,j} = \sum_{k=1}^{d} M_{i,k} M_{j,k} \sim \frac{1}{d}\delta_{i,j} + \underline{1}^{\mathrm{T}}\underline{1} + \mathrm{O}(d^{-2}),$$

where the last term denotes the statistical error of the expression. From here we immediately obtain the asymptotic expression for the upper bound:

$$\limsup_{x\in\mathcal{S}} \|Mx\|_2^2 \sim d^{-1/2}.$$

This means that the second-largest eigenvalue in a typical random Markov matrix is bounded from below by $d^{-1/2}$ in the limit $d \to \infty$. This is true also in the averaging over the ensemble yielding the relation $\langle\nu\rangle \le d^{-1/2}$ and setting the value of the constant $C_0 = 1$. The asymptotic behaviour of the standard deviation $\sigma_{|\nu|}$ is not as clear as in the case of the average value $\langle|\nu|\rangle$. The numerical results suggest the power law decay

$$\sigma_{|\nu|} \sim C_1\, d^{-\alpha}, \qquad \alpha \approx 1.1.$$

For the Markov approximations of dynamical systems with the mixing property [16], as described in the introduction, it is interesting to know about the correlation decay time τ_{c} and entropy growth rate h and their dependence on the cardinality of the state space. In the random Markov matrices from the ensemble we find that the average correlation decay time $\overline{t} := -1/\log\langle|\nu|\rangle$ and average entropy growth rate $\langle h\rangle$ obey the asymptotics

$$\langle h\rangle \overline{\tau}_{\mathrm{c}} \sim \tfrac{1}{2} \qquad \text{as } d \to \infty.$$

In dynamical systems there are strong indications that the correlation decay time and the entropy growth rate, given as the sum of positive Lyapunov exponents, are correlated, but this connection is not well understood, yet; see e.g. [17, 18]. We address this question for the random Markov matrices and calculate pairs $(|\nu|, h)$ corresponding to the Markov matrices sampled uniformly over the ensemble. The result is graphically depicted in figure 5, where one can note that in particular at small dimensions d there is clearly some correlation between the amplitude of the second-largest eigenvalue $|\nu|$ and the entropy growth rate h of random Markov matrices.

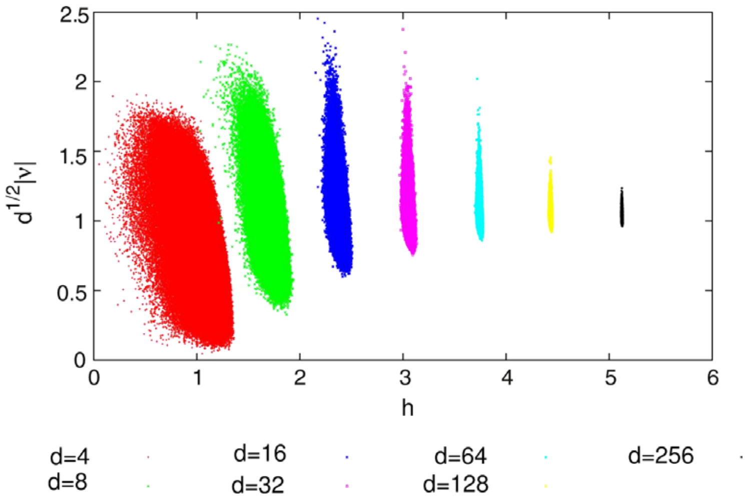

Figure 5. The amplitude of the second-largest eigenvalue $|\nu|$ and corresponding entropy h calculated for some number of random Markov matrices sampled uniformly with respect to the measure $\mathrm{d}P(M)$ (1) at different dimensions d. For details on the statistics see the caption of figure 4.

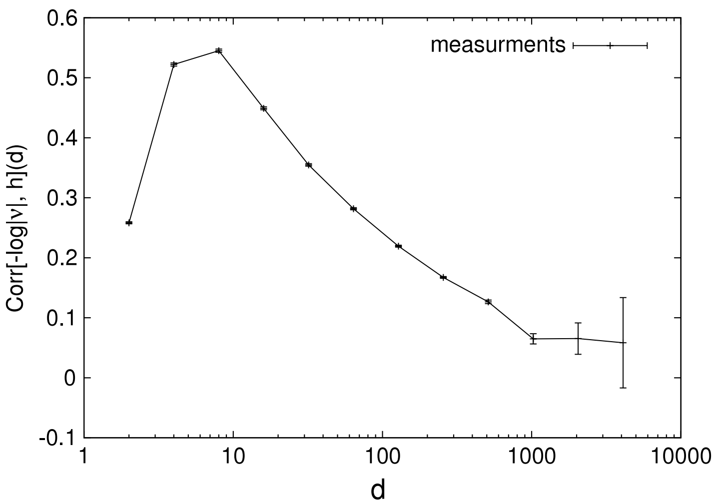

Figure 6. The normalized correlation $\mathrm{Corr}(d)/\mathrm{Corr}(0)$ between the reciprocal correlation time $\tau_\mathrm{c}^{-1} = -\log|\nu|$ and the entropy growth rate h in the ensemble of random Markov matrices as a function of dimension d. For details on the statistics see the caption of figure 4.

The latter is tested by calculating the statistical correlation between the reciprocal correlation decay time $\tau_\mathrm{c}^{-1} = -\log|\nu|$ and the entropy growth rate h over the ensemble of random Markov matrices and is given by

$$\mathrm{Corr}(d) = \frac{\langle(\tau_\mathrm{c}^{-1} - \langle\tau_\mathrm{c}^{-1}\rangle)\,(h - \langle h\rangle)\rangle}{\sqrt{\langle(\tau_\mathrm{c}^{-1} - \langle\tau_\mathrm{c}^{-1}\rangle)^2\rangle\langle(h - \langle h\rangle)^2\rangle}}.$$

J. Stat. Mech. (2009) P07005

The correlation Corr(d) as a function of dimension d is presented in figure 6 and we see that it slowly decreases with increasing d. Currently it is not possible to determine the dependence of the correlation Corr(d) on dimension more precisely.

5. Conclusions

We define the ensemble of random Markov matrices, present its basic properties and point out a few of the potential physical, technical and mathematical applications. We analyse the statistical properties of the stationary distribution $\pi = (\pi_i)_{i=1}^d$ corresponding to a typical element of the ensemble, and study the distribution of the entropy growth rate h over the ensemble, where we find a good agreement with analytical predictions stating that π_i is a Gaussian variable and h is asymptotically equal to $\log(e^{\gamma-1}d)$ in the limit of large dimensions $d \to \infty$. Further we analyse the second-largest eigenvalue ν of the Markov matrices, which is connected to the correlation decay in the Markov chains. We show numerically and provide a heuristic proof that on average, over the ensemble, the second-largest eigenvalue decreases with increasing dimension d as $|\nu| \sim d^{-1/2}$. Additionally we calculate the correlation between the correlation decay rate and the entropy growth rate and find that it decreases with increasing dimension of the Markov matrices.

We believe that the current results enrich the understanding of Markov processes in the limit of large state spaces and all applications which can be described by Markov processes.

Acknowledgments

The author would like to thank Karol Życzkowski and Tomaž Prosen for very interesting discussions and useful comments. The financial support by the Slovenian Research Agency, grant Z1-0875-1554-08, is gratefully acknowledged.

References

[1] Meyn S and Tweedie R L, 1995 *Markov Chains and Stochastic Stability* (Berlin: Springer)
[2] Meyer C D, 2000 *Matrix Analysis and Applied Linear Algebra* (Philadelphia, PA: Society for Industrial and Applied Mathematics)
[3] Christiansen F and Politi A, *Symbolic encoding in symplectic maps*, 1996 *Nonlinearity* **9** 1623
[4] Casati G and ans Prosen T, *Triangle map: a model of quantum chaos*, 2000 *Phys. Rev. Lett.* **85** 4261
[5] Horvat M *et al*, *On ergodic and mixing properties of the triangle map*, 2009 *Physica* D **238** 395
[6] Chafaï D, *The Dirichlet Markov Ensemble*, 2007 arXiv:0709.4678 [math.PR]
[7] Goldberg G, Okunev P, Neumann M and Schneider H, *Distribution of subdominant eigenvalues of random matrices*, 2000 *Method. Comput. Appl. Probab.* **2** 137
[8] Goldberg G and Neumann M, *Distribution of subdominant eigenvalues of matrices with random rows*, 2002 *SIAM J. Matrix Anal. Appl.* **24** 747
[9] Bruzda W, Cappellini V, Sommers H-J and Życzkowski K, *Random quantum operations*, 2009 *Phys. Lett.* A **373** 320
[10] Feller W, 1968 *An Introduction to Probability Theory and its Applications* 3rd edn, vol 1 (New York: Wiley)
[11] Slomczynski W, *Subadditivity of entropy for stochastic matrices*, 2002 *Open Syst. Inf. Dyn.* **9** 201
[12] Girko V L, *Circular law*, 1984 *Theory Probab. Appl.* **29** 694
[13] Tao T and Van V, *Random matrices: the circular Law*, 2007 arXiv:0708.2895v5 [math.PR]
[14] Kotz S and Nadarajah S, 2002 *Extreme Value Distributions: Theory and Applications* (Singapore: World Scientific)
[15] Resnik S I, 1987 *Extreme Values, Regular Variation, and Point Processes* (New York: Springer)

[16] Ott E, 1993 *Chaos in Dynamical Systems* (Cambridge: Cambridge University Press)
[17] Ruelle D, *Resonances of chaotic dynamical systems*, 1986 *Phys. Rev. Lett.* **56** 405
[18] Eckmann J-P and Ruelle D, *Ergodic theory of chaos and strange attractors*, 1985 *Rev. Mod. Phys.* **57** 617

J. Stat. Mech. (2009) P07005