Notes on Finite State Sapce Markov Processes

December 14, 2018

Abstract

These are some notes on Discrete State Sapce Markov Processes.

0.1 Stochastic Processes with a Finite State Space

Let $\mathcal{S}$ be a countable set; either finite or infinite. A *discrete time stochastic process* (X_j) *with state space* $\mathcal{S}$ is a function associating to each $j \in \mathbb{N}$, or some other subset of the integers, a random variable X_j on some probability space (Ω, P) where each X_j has values in $\mathcal{S}$. A *continuous time stochastic process* (X_t) *with state space* $\mathcal{S}$ is a function associating to each $t \in [0, \infty)$, or some other connected subset of the real numbers, a random variable X_t on some probability space (Ω, P) where eaxh X_t has values in $\mathcal{S}$. The Poisson process is an example of a continuous time stachastic process with values in the non-negative integers.

Any sequence of random varibles $\{X_j\}_{j\in\mathbb{N}}$, with values in a countable set $\mathcal{S}$ may be regarded as a stochastic process, and we have spent considerable time studying the case in which the X_j's are independent and identicallly distributed. We now move beyond this special situtation.

For each (random) outcome $\omega \in \Omega$, a discrete time strochastic process (X_j) gives us a random sequence $j \mapsto X_j(\omega)$, while a continuous time stochastic process (X_t) gives us a random path $t \mapsto X_t(\omega)$. These random functions are called the *sample paths* of the stochastic process, and they are the main object of study in the theory of stochastic processes.

0.1 EXAMPLE. *The* Ehrenfest process *is a stochastic process arsing as follows: For $N \in \mathbb{N}$, consider a set of N numbered balls that are distributed among two urns, urn A and urn B. At each step of the process, a number $k \in \{1, \ldots, N\}$ is selected uniformly at random. The kth ball is then taken out of the urn that it is in, and it is put into the other urn, Let X_j denote the number of balls in urn A at the jth step, The state space of this stochastic process is $\{0, 1, \ldots, N\}$.*

0.2 Markov chains

0.2 DEFINITION (Markov chain)**.** A discrete time stochastic process (X_j) with discrete state space $\mathcal{S}$ indexed by $j \in \{0, 1, 2, \dots\}$ is a *Markov chain* in case for each $n \in \mathbb{N}$, and all $(x_0, \dots, x_{n+1}) \in \mathcal{S}^{n+1}$,

$$P(X_{n+1} = x_{n+1} | X_n = x_n, X)_{n-1} = x_{n-1}, \dots, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n) \ .$$

A Markov chain has *stationalry transitions* in case for all $i, j \in \mathcal{S}$,

$$P(X_{n+1} = j | X_n = i)$$

is independent of n.

That is, a discrete time stochastic process (X_j) with values in $\mathcal{S}$ is a Markov chain if, given the value of X_n, X_{n+1} is independent of $\{X_0, \dots, X_{n-1}\}$. Of course, a sequence of independent random variables is Markov, but the Markov property is much more general, as we shall see in the examples that follow.

0.3 EXAMPLE. *The* Ehrenfest process *is a Markov chain with stationary transitions and state sapce $\mathcal{S} = \{0, 1, \dots, N\}$. To see this, suppose that $X_{n-1} = j$. Then X_n will be either $j - 1$ (if $j \neq 0$) or $j + 1$ (if $j \neq N$) depending on whether the next ball selected is in urn A or urn B. Given that $X_{n-1} = j$, the probability that the selected ball is in urn A is $\frac{j}{N}$ and hence $P(X_n = j - 1 | X_n = j) = \frac{j}{N}$. Likewise, the probability that the selected ball is in urn B is then $\frac{N-j}{N}$ and hence $P(X_n = j - 1 | X_n = j) = \frac{N-j}{N}$. The process is Markov, because the statsitics of the value of X_{n+1} depend only on the value of X_n, and not on any othe rest of the previous histroy of the process. The transitions are stationaly becasue the transition probabilities do not depend on n.*

Let us focus first on the case of a Markov chain with stationary transitions and a finite states-pace $\mathcal{S}$. We identify $\mathcal{S}$ with $\{1, \dots, M\}$ for some $M \in \mathbb{N}$. Then the $M \times M$ matrix P with entries

$$P_{i,j} = P(X_1 = j | X_0 = i)$$

is called the *transition matrix* of the Markov chain (X_j). The *initial distribution* is the vector $\boldsymbol{\pi}_0 \in [0, 1]^M$ with entries

$$(\boldsymbol{\pi}_0)_k = P(X_0 = k)$$

. Given $\boldsymbol{\pi}_0$ and the transition matrix P, we can compute the probability of any event of the form

$$\{X_0 = x_0, X_0 = x_1, \dots, X_n = x_n\} \ ,$$

and therefore, any event at all. All of the information needed to answer any question about the sample paths is contained in $\boldsymbol{\pi}_0$ and P.

To see this, use the chain rule, and the Markov property to wrtie

$$\begin{aligned} P(X_n = x_n, \ldots, X_0 = x_0) &= P(X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) P(X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \\ &= P_{x_{n-1},x_n} P(X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \\ &= P(X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) P_{x_{n-1},x_n} \ , \end{aligned}$$

where in the last step we simply moved the number P_{x_{n-1},x_n} to the right for convenience.

The same reasoning then gives

$$P(X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = P(X_{n-2} = x_{n-2}, \ldots, X_0 = x_0) P_{n-2,x_{n-1}} \ ,$$

and then, combining formulas

$$P(X_n = x_n, \ldots, X_0 = x_0) = P(X_{n-2} = x_{n-2}, \ldots, X_0 = x_0) P_{n-2,x_{n-1}} P_{x_{n-1},x_n} \ .$$

Continuing in the same way, we ventually arrive at

$$\begin{aligned} P(X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) &= P(X_0 = x_0) P_{x_0,x_1} P_{x_1,x_2} \cdots P_{x_{n-1},x_n} \\ &= (\boldsymbol{\pi}_0)_{x_0} P_{x_0,x_1} P_{x_1,x_2} \cdots P_{x_{n-1},x_n} \ . \end{aligned} \tag{0.1}$$

We can now give a useful fromula for $P(X_n = j)$.

0.4 THEOREM. *Let (X_j) be a Markov chain with stationary transitions and state sapce $\mathcal{S}$, initial distribution $\boldsymbol{\pi}_0$, and transition matrix P. Then for all $j \in \mathcal{S}$ and all $n \in N$,*

$$P(X_n = j) = (\boldsymbol{\pi}_0 P^n)_j \ , \tag{0.2}$$

where the right side is the jth entry of the vector $\boldsymbol{\pi} P^n$ obtained by multiplying the row vector $\boldsymbol{\pi}$ on the right by P^n, the nth matrix power of P.

Proof. To apply the previous formula, write x_n in place of j. Then summing over all possible values of $X_1, \ldots, X_{n-1}$,

$$\begin{aligned} P(X_n = x_n, X_0 = x_0) &= \sum_{x_1, \ldots x_{n-1}=1}^{M} P(X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \\ &= \sum_{x_1, \ldots x_{n-1}=1}^{M} (\boldsymbol{\pi}_0)_{x_0} P_{x_0,x_1} P_{x_1,x_2} \cdots P_{x_{n-1},x_n} \ . \end{aligned}$$

Therefore,

$$\begin{aligned} P(X_n = x_n | X_0 = x_0) &= \frac{P(X_n = x_n, X_0 = x_0)}{P(X_0 = x_0)} \\ &= \sum_{x_1, \ldots x_{n-1}=1}^{M} P_{x_0,x_1} P_{x_1,x_2} \cdots P_{x_{n-1},x_n} \ , \end{aligned}$$

which is nothing other than the x_0, x_n entry of P^n, the nth matrix power of the transition matrx P. Again by the chain rule,

$$P(X_n = x_n) = \sum_{x_0=1}^{M} P(X_n = x_n | X_0 = x_0) P(X_0 = x_0) = \sum_{x_0=1}^{M} (\boldsymbol{\pi}_0)_{x_0} (P^n)_{x_0,x_n} \ .$$

In other words, for each $j \in \mathcal{S}$, $P(X_n = j)$ is the jth entry of $\boldsymbol{\pi} P^n$, where we regard $\boldsymbol{\pi}_0$ as a row vector (a $1 \times M$ matrix) so we can multiply it on the right by the $M \times M$ matrix P^n. □

For each $n \in \mathbb{N}$, define the probability vector $\boldsymbol{\pi}_n$ by

$$(\boldsymbol{\pi}_n)_j = P(X_n = j) \ ,$$

and regard it as a row vector. Then we can restate the conclusion of Theroem 0.4 is

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 P^n \ . \tag{0.3}$$

It follows that for all $m, n \in \mathbb{N}$, $\boldsymbol{\pi}_{m+n} = \boldsymbol{\pi}_0 P^{m+n} = (\boldsymbol{\pi}_0 P^m) P^n = \boldsymbol{\pi}_m P^n$. That is,

$$\boldsymbol{\pi}_{m+n} = \boldsymbol{\pi}_m P^n \ . \tag{0.4}$$

0.3 Properties of the transition matrix

Let P be the transition matrix of a Markov chain with stationary transtions and state space $\mathcal{S} = \{1, \ldots, M\}$. Then since for each n,

$$P_{i,j} = P(X_n = j | X_{n-1} = i) \ ,$$

and since $1 = P(X_n \in \{1, \ldots, M\})$,

$$1 = \sum_{j=1}^{M} P(X_n = j | X_{n-1} = i) = \sum_{j=1}^{M} P_{i,j} \ ,$$

the sum of the entries in each row is 1, and of course since each entry is a probability, each entry is non-negative. Such a matrix is called a *row stochastic Matrix.*

0.5 LEMMA. *For any vector* $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_M \end{pmatrix} \in \mathbb{R}^M$*, and every* $M \times M$ *row stochastic matrix* P*,*

$$\max_{1 \le i \le M} \{(P\mathbf{v})_i\} \le \max_{1 \le i \le M} \{v_i\} \quad \text{and} \quad \min_{1 \le i \le M} \{(P\mathbf{v})_i\} \ge \min_{1 \le i \le M} \{v_i\}$$

Proof. This is an immediade conseqeunce of the fact that each $(P\mathbf{v})_i$ is a weigthed avaerage of the entries $\{v_1, \ldots, v_M\}$ of $\mathbf{v}$ because an average is always no smaller than the minimum, and no greater than the maximum, no matter what the weights are. □

Conversely, given an $M \times M$ row stochastic matrix P and a probability row vector $\boldsymbol{\pi}$, we can construct a Markov chain with stationary transitions, initial distribution $\boldsymbol{\pi}$ and transition matrix P by using (0.1) to define probabilities of events of the form $P(X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0)$. By the discussion that led to this formula, this specification of the probabilities makes (X_j) a Markov chain with stationary transitions.

0.6 LEMMA. *An $M \times M$ matrix P is row stochastic if and only if it has non-negative entries and $\mathbf{v}_1$, the vetor that has 1 in each entry, is an eigenvectors of P with eigenvalue 1.*

Proof. For each i,

$$(P\mathbf{v}_1)_i = \sum_{j=1}^{M} P_{i,j}(\mathbf{v}_1)_j = \sum_{j=1}^{M} P_{i,j} \ .$$

Therefore, $P\mathbf{v}_1 = \mathbf{v}_1$ if and only if for each i, then $(P\mathbf{v}_1)i = 1$ and hence if and only if for each i $\sum_{j=1}^{M} P_{i,j} = 1$ □

0.7 COROLLARY. *Let P and Q be any two $M \times M$ row stochastic matrices. Then the product PQ is row stochastic. In particular, P^n is row stochastic for all $n \in \mathbb{N}$.*

Proof. Since each entry of P and of Q is no negative so is $\sum_{K=1}^{M} P_{i,k}Q_{k,j}$ Moreover $(PQ)\mathbf{v}_1 = P(Q\mathbf{v}_1) = P\mathbf{v}_1 = \mathbf{v}_1$, so that $\mathbf{v}_1$ is an eigenvector of PQ with eiqgenvalue 1. □

0.8 LEMMA. *Let P be any $M \times M$ row stochastic matrix. If λ is any other eigenvalue of P, then $|\lambda| \leq 1$.*

Suppose moreover that for some n, $P^n_{i,j} > 0$ for all i, j, and let $\mathbf{v}_1$ be the vector in $\mathbb{R}^M$ all of whose entries are 1. Then $\mathbf{v}_1$ spans the eigenspace for the eigenvalue 1, and all other egienvalues λ of P satisfy $|\lambda| < 1$.

Proof. Let $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_M \end{pmatrix}$ be any eigenvector: $P\mathbf{v} = \lambda\mathbf{v}$. Suppose that the i_0 entry of $\mathbf{v}$ is largest in absolute value. We may normalize $\mathbf{v}$ so that $v_{i_0} = 1$, and then $|v_j| \leq 1$ for all j. Then

$$\lambda = (P\mathbf{v})_{i_o} = \sum_{j=1}^{M} P_{i_0,j}v_j \ ,$$

and hence

$$|\lambda| \leq \sum_{j=1}^{M} P_{i_0,j}|v_j| \leq \sum_{j=1}^{M} P_{i_0,j}1 = 1 \ .$$

This completes the proof of the first part. Now suppose that for some n, $P^n_{i,j} > 0$ for all i, j. Again, let $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_M \end{pmatrix}$ be any eigenvector of P, and let λ be the corresponding eigenvalue.

Again we normalize $\mathbf{v}$ so that $v_{i_0} = 1$ and $|v_j| \leq 1$ for al j. Then $\mathbf{v}$ is an eigenvector of P^n with eigenvalue λ^n, so that, as above,

$$\lambda^n = (P^n \mathbf{v})_{i_o} = \sum_{j=1}^{M} P^n_{i_0,j} v_j \ ,$$

and hence

$$|\lambda^n| = \left| \sum_{j=1}^{M} P^n_{i_0,j} v_j \right| \leq \sum_{j=1}^{M} P^n_{i_0,j} |v_j| \leq \sum_{j=1}^{M} P_{i_0,j} 1 = 1 \ .$$

There is equality in the first inequality if and only if $v_j = |v_j|$ for each j, and then there is equality in the second inequality if and only if $|v_j| = 1$ for each j. Hence $|\lambda^n| = 1$ if and only if $v_j = 1$ for each j. But in this case, $\mathbf{v} = \mathbf{v}_1$. Hence if $|\lambda^n| = 1$, $\mathbf{v} = \mathbf{v}_1$, and then $\lambda = 1$. Otherwise, $|\lambda^n| < 1$, and hence $|\lambda| < 1$. □

0.9 LEMMA. *Let P be any $M \times M$ row stochastic matrix. Suppose that for some n, $P^n_{i,j} > 0$ for all i, j, Then there is a row probability vector $\boldsymbol{\pi}_\infty$ such that*

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \boldsymbol{\pi}_\infty \\ \vdots \\ \boldsymbol{\pi}_\infty \end{bmatrix} \ ,$$

where the matrix on the right is the $M \times M$ matrix, each of whose rows is $\boldsymbol{\pi}_\infty$. The row vector $\boldsymbol{\pi}_\infty$ is a left eigenvector of P with eigenvalue 1; that is

$$\boldsymbol{\pi}_\infty P = \boldsymbol{\pi}_\infty,$$

and $\boldsymbol{\pi}_\infty$ spans the corresponding eigenspace.

Proof. We suppose first that P is diagonalizable. Then there exists a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$ of $\mathbb{C}^M$ consisting of eigenvectors of P: $P\mathbf{v}_j = \lambda_j \mathbf{v}_j$ for $j = 1, \ldots, M$. Without loss of generality, we may suppose that $\lambda_1 = 1$ and the $\mathbf{v}_1$ is the vector each of whose entries is 1. Let $V = [\mathbf{v}_1, \ldots, \mathbf{v}_M]$, the $M \times M$ matrix whose jth column is $\mathbf{v}_j$. Let Λ be the $M \times M$ diagonal matrix whose jth diagoanal entry is λ_j. Then $PV = V\Lambda$, and since the columns of V are linearly independent, V is invertible and $P = V\Lambda V^{-1}$. It follows that

$$P^n = V\Lambda^n V^{-1} \ . \tag{0.5}$$

Now since $\lambda_1 = 1$ and $|\lambda_j| < 1$ for all $j > 1$,

$$\lim_{n\to\infty} \Lambda^n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} = \Lambda_\infty \ ,$$

where the matrix on the right has a 1 in the upper left corner, and all of the other entries are zero. It follws that

$$\lim_{n\to\infty} P^n = V\Lambda_\infty V^{-1} \ . \tag{0.6}$$

Since Λ_∞ is a rank one matrix, so is $\lim_{n\to\infty} P^n \lim_{n\to\infty} P^n$, and all of its rows are proportional. Moreover, for each n, since P^n is row stochastic, each row of P^n is a probability vector, and this must be true in the limit. But probability vectors that are proportional are equal, and so all rows of the limit are the *same* probability vector, which we shall call $\boldsymbol{\pi}_\infty$.

Since $\lim_{n\to\infty} P^n = \lim_{n\to\infty} P^{n+1} = (\lim_{n\to\infty} P^n)P$, we have that

$$\begin{bmatrix} \boldsymbol{\pi}_\infty \\ \vdots \\ \boldsymbol{\pi}_\infty \end{bmatrix} P = \begin{bmatrix} \boldsymbol{\pi}_\infty \\ \vdots \\ \boldsymbol{\pi}_\infty \end{bmatrix} \quad \text{which is the same as} \quad \begin{bmatrix} \boldsymbol{\pi}_\infty P \\ \vdots \\ \boldsymbol{\pi}_\infty P \end{bmatrix} = \begin{bmatrix} \boldsymbol{\pi}_\infty \\ \vdots \\ \boldsymbol{\pi}_\infty \end{bmatrix} \ ,$$

showing the $\boldsymbol{\pi}_\infty P = \boldsymbol{\pi}_\infty$. Now let $\mathbf{x} = (x_1, \ldots, x_M)$ be any row vector in $\mathbb{C}^M$ that is a left eigenvector of P with eigenvalue 1. Then for all n, $\mathbf{x} = \mathbf{x}P^n$, and hence

$$\mathbf{x} = \mathbf{x} \begin{bmatrix} \boldsymbol{\pi}_\infty \\ \vdots \\ \boldsymbol{\pi}_\infty \end{bmatrix} = \left(\sum_{i=1}^{M} x_i \right) \boldsymbol{\pi}_\infty \ .$$

This shows that $\mathbf{x}$ is a multiple of $\boldsymbol{\pi}_\infty$; hence $\boldsymbol{\pi}_\infty$ spans the left eigenspace with eigenvalue 1.

If P is not diagonalizable, one must use generalized eigenvectors: For every $M \times M$ matrix P, there is a basis of $\mathbb{C}^M$ consisting of generalized eigenvectors of P. The first thing to be shown is that there is no generalized eigenvector of P with eigenvalue 1 apart from multiples of the eigenvector $\mathbf{v}_1$. Suppose this is false: Then there is some non-zero vector $\mathbf{v}$ such that $(P-I)^2\mathbf{v} = 0$ but $(P-I)\mathbf{v} \neq 0$. This means that $\mathbf{w} := (P-I)\mathbf{v}$ is an eienvector of P with eigenvalue 1, hence it is a multiple of $\mathbf{v}_1$. Multiplying $\mathbf{v}$ by a constant, we may assume that $(P-I)\mathbf{v} = \mathbf{v}_1$. That is, $P\mathbf{v} = \mathbf{v} + \mathbf{v}_1$. Iterating, we see that

$$P^n\mathbf{v} = \mathbf{v} + n\mathbf{v}_1 \ .$$

But since P^n is a row stochastic matrix, by Lemma 0.5, the entries on the left are bounded unifirmly in n, but those on ghe right are not. This contradiction shows that there is no generalized eigenvector of P with eigenvalue 1.

Hence in the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$ of P consisirting of generalized eigenvectors, we have that $P\mathbf{v}_j - \lambda_j \mathbf{v}j$ with $|\lambda_j| < 1$ for all $j > 1$. It is then a simple matter to see that (0.6) is still valid. □

0.10 EXAMPLE. *Consider the Ehrenfest process for $N = 3$. If we identify $\mathcal{S}$ with $\{1, 2, 3, 4\}$ by letting j denote the state in which there are $j-1$ balls in urn A, then: Then the transition matrix P is the 4×4 matrix*

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{bmatrix} \ .$$

However, it can be adavantageous to consider other orderings of the states. Since X_j changes from even to odd or odd to even at each step, if $X_0 \in \{0,2\}$, then $X_2 \in \{0,2\}$, while if $X_0 \in \{1,3\}$, then $X_2 \in \{1,3\}$. This will give P^2 a block structure if we number our states accoridng to these groups.

Therefore, let state 1 denote 0 balls in urn A, state 2 denote 2 balls in urn A, state 3 denote 1 ball in urn A, and state 4 denote 3 balls in urn A. That is, relative to our first ordering, we have swapped the roles of states 2 and 3. This swaps the second and third columns and rows of the matrix P, so that now

$$P = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & C \\ D & 0 \end{bmatrix} \quad \text{where} \quad C = \frac{1}{3}\begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad D = \frac{1}{3}\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \ . \ .$$

For this ordering of the states, we compute

$$P^2 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{2}{9} & \frac{7}{9} & 0 & 0 \\ 0 & 0 & \frac{7}{9} & \frac{2}{9} \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \quad \text{where} \quad A = \frac{1}{9}\begin{bmatrix} 3 & 6 \\ 2 & 7 \end{bmatrix} \quad \text{and} \quad B = \frac{1}{9}\begin{bmatrix} 7 & 2 \\ 6 & 3 \end{bmatrix} \ . \ .$$

It follows that for $m \in \mathbb{N}$,

$$P^{2m} = \begin{bmatrix} A^m & 0 \\ 0 & B^m \end{bmatrix} \qquad \text{and} \qquad P^{2m+1} = \begin{bmatrix} 0 & C \\ D & 0 \end{bmatrix}\begin{bmatrix} A^m & 0 \\ 0 & B^m \end{bmatrix} = \begin{bmatrix} 0 & CB^{2m} \\ DA^{2m} & 0 \end{bmatrix} \ .$$

In particular, it is never the case that $P^n_{i,j} > 0$ for all i, j, no matter what n, is. This reflects the fact that for each ω, $X_n(\omega)$ aleternates between even and odd as n increases.

However, A and B are both rwo stochastic matrices, and even for their first power, all entries are strinctly positive. Hence all of Lemma 0.9 applies to them.

It is easy to compute A^m by diaginalizing A: We compute that the characteristic polynomial of A is $t^2 - \frac{10}{9}t + \frac{1}{9}$, and hence the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = \frac{1}{9}$. Corresponding eigenvectors are $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$. Then with $V = [\mathbf{v}_1, \mathbf{v}_2] = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}$,

$$A = V\begin{bmatrix} 1 & 0 \\ 0 & 9^{-1} \end{bmatrix}V^{-1} \qquad \text{so that} \qquad A^m = V\begin{bmatrix} 1 & 0 \\ 0 & 9^{-m} \end{bmatrix}V^{-1}$$

Therefore,

$$\lim_{m\to\infty} A^m = V\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}V^{-1} = \frac{1}{4}\begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\pi}_A \\ \boldsymbol{\pi}_A \end{bmatrix} \quad \text{where} \quad \boldsymbol{\pi}_A = \frac{1}{4}(1,3) \ .$$

In the same way, one finds that

$$\lim_{m\to\infty} B^m = \frac{1}{4}\begin{bmatrix} 3 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\pi}_B \\ \boldsymbol{\pi}_B \end{bmatrix} \quad \text{where} \quad \boldsymbol{\pi}_B = \frac{1}{4}(3,1) \ .$$

The diagonalization of A and B illustrates how the proof of Lemma 0.9 works, but to compute the limits $\lim_{m\to\infty} A^m$ and $\lim_{m\to\infty} B^m$, we did not need to carry out the full diagonalization: All we needed to do was to compute the left eigenvectors of A and B with eigenvalue 1.

We now do this for A: Let $\boldsymbol{\pi} = (a, b)$ be such that $\boldsymbol{\pi} = \boldsymbol{\pi}A$, or what is the same, $A^T\boldsymbol{\pi}^T = \boldsymbol{\pi}^t$. To find $\boldsymbol{\pi}^T$, We find all solutions of $(A^T - I)\mathbf{x} = 0$. Since

$$(A^T - I) = \frac{1}{9}\begin{bmatrix} -6 & 2 \\ 6 & -2 \end{bmatrix} ,$$

every solutiuon $\mathbf{x}$ of $(A^T - I)\mathbf{x} = 0$ orthogonal to $\begin{pmatrix} -6 \\ 2 \end{pmatrix}$, and the only such vector whose entries sum to 1 is $\frac{1}{4}\begin{pmatrix} 1 \\ 3 \end{pmatrix}$. Therefore,

$$\boldsymbol{\pi}_A = \frac{1}{4}(1, 3) \quad \text{and} \quad \lim_{m\to\infty} A^m = \begin{bmatrix} \boldsymbol{\pi}_A \\ \boldsymbol{\pi}_A \end{bmatrix} .$$

Similar computations may be made for B. Avoiding the full diagonalization was not so important i nthis 2×2 example, but it becomes mushc more importanta as the size of the row stochastic matries we are working with becomes larger.

In any case, we have found that for large m,

$$P^{2m} \approx \frac{1}{4}\begin{bmatrix} 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 3 & 1 \end{bmatrix} \qquad \text{and then} \qquad P^{2m+1} = PP^{2m} \approx \frac{1}{4}\begin{bmatrix} 0 & 0 & 3 & 1 \\ 0 & 0 & 3 & 1 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{bmatrix} .$$

Now suppose $\boldsymbol{\pi}_0 = (a, b, c, d)$. Then $P(X_{2m} = j)$ may be determined by computing

$$\boldsymbol{\pi}_{2m} = \boldsymbol{\pi}_0 P^{2m} \approx (a + b, 3(a + b), 3(c + d), c + d)) ,$$

where the approximation is exact in the limit $m \to \infty$. Recalling the ordering of the states we have that, for large m,

$$\begin{aligned} P(X_{2m} = 0) &\approx P(X_0 = 0) + P(X_0 = 2) \\ P(X_{2m} = 2) &\approx 3P(X_0 = 0) + 3P(X_0 = 2) \\ P(X_{2m} = 3) &\approx 3P(X_0 = 1) + 3P(X_0 = 3) \\ P(X_{2m} = 4) &\approx P(X_0 = 1) + P(X_0 = 3) \end{aligned}$$

0.11 EXAMPLE. *We have seen that the transition matrix P Ehrenfest process does not satisfy the condition that for some n, $P^n_{i,j} > 0$ for all i, j. However, with a small change, we arrive at a process for which this condition is satsified:*

At each step, instead of choosing a number in $\{1, 2, 3\}$, and then moving the ball with that number to the other urn, we choose, uniformly at random a number in $\{0, 1, 2, 3\}$. If the number chosen is zero, no change is made. If the number chosen is in $\{1, 2, 3\}$, we proceed as before.

*Before we made the change, one could go from any even number of balls in urn A (*0 *or* 2*) to any other even number in two steps. One could also go from any odd number of balls in urn A (*1 *or* 3*) to any other eodd number in two steps. However, if* $i := X_0(\omega)$ *is even, then after any even number* n *of steps,* $X_n(\omega)$ *will be even, so that for* j *odd,* $P^n_{i,j} = 0$*, while after any odd number* n *of steps,* $X_n(\omega)$ *will be odd, so that for* j *even,* $P^n_{i,j} = 0$*.*

In the modified version, we have the option of staying in place, and so after 3 *steps one can go from any state to any other: If you have already gone from* i *to* j *in one or two steps, you don't have to leave on the next step, and can wait around if you have arrived early. So for the modified process,* $P^3_{i,j} > 0$ *for all* i, j*.*

There is no longer any reason to group the states in any special way so we number our state in the simpliest way: We identify $\mathcal{S}$ *with* $\{1, 2, 3, 4\}$ *by letting* j *denote the state in which there are* $j - 1$ *balls in urn A. Then one easily works out that the transition matrix is*

$$P = \frac{1}{4}\begin{bmatrix} 1 & 3 & 0 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 1 \end{bmatrix} .$$

To find $\boldsymbol{\pi}_\infty$*, we have to solve the equation* $(P^T - I)\mathbf{x} = 0$*. We compute*

$$P^T - I = \frac{1}{4}\begin{bmatrix} -3 & 1 & 0 & 0 \\ 3 & -3 & 2 & 0 \\ 0 & 2 & -3 & 3 \\ 0 & 0 & 1 & -3 \end{bmatrix} .$$

Using Gaussian elimination, we find that the equation $(P^T - I)\mathbf{x} = 0$ *is equivalent to the equation* $U\mathbf{x} = 0$ *where*

$$U = \frac{1}{4}\begin{bmatrix} -3 & 1 & 0 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} .$$

By back substitution, we find that every solution of $U\mathbf{x} = 0$ *is a multiple of* $\begin{pmatrix} 1 \\ 3 \\ 3 \\ 1 \end{pmatrix}$*. There is only one multiple for which hte entries sum to* 1*, and hence we have that*

$$\boldsymbol{\pi}_\infty = \frac{1}{8}(1, 3, 3, 1) .$$

0.4 The rate of approach to equilibrium

Throughout the rest of these notes, we consider a Markov chain (X_j) in $\mathcal{S} = \{1, \ldots, M\}$ with stationary transitions governed by the row stochastic matrix P, and we suppose that:

$$\text{For some } m \in \mathbb{N},\ P^m_{i,j} > \delta \text{ for all } i, j\ . \tag{0.7}$$

We have seen in the previous section that

$$\lim_{n\to\infty} P^n = P_\infty = \begin{bmatrix} \boldsymbol{\pi}_\infty \\ \vdots \\ \boldsymbol{\pi}_\infty \end{bmatrix} =: P_\infty$$

where $\boldsymbol{\pi}_\infty$ is the unique probability vector satisfying $\boldsymbol{\pi}_\infty = \boldsymbol{\pi}_\infty P$, and that, as a consequence, for any initial probability vector $\boldsymbol{\pi}_0$, $\lim_{n\to\infty} \boldsymbol{\pi}_0 P^n = \boldsymbol{\pi}_\infty$.

In this section, we shall show that these two limits are achieved exponentially fast, and shall explain some of the implications of this. Let f be any real valued function on $\{1,\dots,M\}$. We may identify the function f with the vector $\mathbf{f} = \begin{pmatrix} f(1) \\ \vdots \\ f(M) \end{pmatrix}$. Then

$$\mathrm{E}(f(X_t)) = \sum_{j=1}^{N} (\boldsymbol{\pi}_n)_j f(j) = \boldsymbol{\pi}_n \cdot \mathbf{f}$$

Since $\lim_{n\to\infty} \boldsymbol{\pi}_n = \boldsymbol{\pi}_\infty$, we define

$$\langle f \rangle_\infty = \sum_{j=1}^{M} (\boldsymbol{\pi}_\infty)_j f(j) = \boldsymbol{\pi}_\infty \cdot \mathbf{f} \tag{0.8}$$

The main theorems in this section are the following:

0.12 THEOREM. *Suppose that the Markov chain (X_j) with state space $\{1,\dots,M\}$ satisfies* (0.7). *Then there are constants $C < \infty$ and $r < 1$, depending only on the values of m and δ in* (0.7), *such that for all n,*

$$|\mathrm{E}(f(X_n)) - \langle f \rangle_\infty| \le \left(\max_{1\le j\le M} \{|f(j)|\} \right) C r^n \ . \tag{0.9}$$

In particular, if f is such that $\langle f \rangle_\infty = 0$, then $\mathrm{E}(f(X_n))$ converges to zero exponentially fast, no matter what the initial distribution $\boldsymbol{\pi}_0$ may be. In particular, for all i,

$$\sum_{j=1}^{M} |(\boldsymbol{\pi}_n)_j - (\boldsymbol{\pi}_\infty)_j| \le C r^n \ , \tag{0.10}$$

From this theorem, we shall deduce an interesting theorem on *average occupation times*. For any function f on $\{1,\dots,M\}$, define

$$\overline{f(X)}_N := \frac{1}{N} \sum_{j=1}^{N} f(X_j) \ .$$

The random variable $\overline{f(X)}_N$ is the weighted average value of f, when we weight the average by the fraction of the steps the Markov chain spends in each sate in $\{1,\dots,M\}$. Here is a particularly interesting choice for f. For each $i \in \{1,\dots,M\}$ define

$$\mathbf{1}_i(j) = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases} .$$

Then the random variable $\overline{\mathbf{1}_i(X)}_N$ is the fraction of the steps 1 through N that the Markov chain spends in state i.

0.13 THEOREM. *Suppose that the Markov chain (X_j) with state space $\{1,\dots,M\}$ satisfies (0.7). Then there is a constant $K < \infty$ depending only on the values of m and δ in (0.7), such that for all $N \in \mathbb{N}$, and all $\epsilon > 0$,*

$$P(|\overline{\mathbf{1}_i(X)}_N - (\boldsymbol{\pi}_\infty)_i| > \epsilon) \leq \frac{K}{N\epsilon^2} . \tag{0.11}$$

That is, with high probability, for large N, the fraction of the first N steps of the Markov chain that are spent in state i is very close to $(\boldsymbol{\pi}_\infty)_i$, no matter what the initial distribution $\boldsymbol{\pi}_0$ may be. We shall give a complete proof of Theorem 0.13, which is quite simple. Working harder, but still using the ideas that go into the proof we give, together with some more ideas that we will not go into, one can show that the rate of convergence in (0.13) is also exponential.

However, Theorem 0.13, simple as it may be, is enough to give important insight into the meaning of the vector $\boldsymbol{\pi}_\infty$: For large values of N, and any $i \in \{1,\dots,M\}$ it is very likely that the fraction of the first N steps the Markov chain is in state i is very close to $(\boldsymbol{\pi}_\infty)_i$.

0.14 EXAMPLE. *A jogging enthusiast owns 5 pairs of running shoes. Each pair is kept at either the front doors or the back door of the jogger's house. Every day the jogger tosses a fair coin and goes to the front door if the toss is "heads" and goes to the back door if the toss is "tails". The jogger puts on a pair of running shoes if there is one at the chosen door, and goes jogging. If there are no pair of shoes at the chosen door, the jogger goes barefoot.*

At the end of the jogger's run, they flip a fair coin again and enter home by the front door if the toss is "heads" and by the back door if the toss is "tails". They take off the running shoes and leave them at the chosen door. The jogger follows this routine for several years. What is the fraction of the days in these several years that the jogger ran barefoot?

Let X_n be the number of pairs of running shoes at the front door before the run on day n. Then (X_n) is a stochastic process, and in fact it is a Markov process: Given that $X_n = j$, X_{n+1} will be one of $j-1$ (if $j \neq 0$), j, or $j+1$ (if $j \neq 5$),and the probabilities of the various possible transitions can be computed knowing only that $X_n = j$.

Suppose $X_n = 0$. If the jogger goes to the front door before the run, there are no shoes, they run barefoot, and after returning home, there are still no shoes at the front does. If the jogger goes to the back door before the run, there are 5 pairs of shoes. They pick one and go jogging. If they

return to the back door, they leave the shoes there, and still there are no shoes at the front door. But if they go to the front door upon returning, they leave a pair of shoes there. Therefore:

$$P(X_{n+1} = 0|X_n = 0) = \frac{3}{4} \quad \text{and} \quad P(X_{n+1} = 1|X_n = 0) = \frac{1}{4} \ .$$

all other transition probabilities, given $X_n = 0$*, are* 0*.*

Suppose $X_n = 1$*. If the jogger goes to the front door before the run, they put on the pair of shoes there, and after returning home, if the go to the front door, they leave the shoes there, and again there is exactly one pair at the front door. But if the return to the back door, thy leave the shoes there, and then there are no shoes at the front door. If they go to the back door before running, they put on a pair of shoes, and run. If they return to the back door, they leave the shoes there, and the result is that there is still one pair at the front door. But if they return to the front doors there will now be two pairs at the front door, Therefore,*

$$P(X_{n+1} = 0|X_n = 1) = P(X_{n+1} = 2|X_n = 1) = \frac{1}{4} \quad \text{and} \quad P(X_{n+1} = 1|X_n = 1) = \frac{1}{2} \ .$$

all other transition probabilities, given $X_n = 1$ *are* 0*.*

In the same way, we find:

$$P(X_{n+1} = 1|X_n = 2) = P(X_{n+1} = 3|X_n = 2) = \frac{1}{4} \quad \text{and} \quad P(X_{n+2} = 2|X_n = 2) = \frac{1}{2} \ .$$

all other transition probabilities, given $X_n = 2$ *are* 0*.*

$$P(X_{n+1} = 2|X_n = 2) = P(X_{n+1} = 4|X_n = 3) = \frac{1}{4} \quad \text{and} \quad P(X_{n+1} = 3|X_n = 3) = \frac{1}{2} \ .$$

all other transition probabilities, given $X_n = 3$ *are* 0*.*

$$P(X_{n+1} = 0|X_n = 4) = P(X_{n+1} = 5|X_n = 4) = \frac{1}{4} \quad \text{and} \quad P(X_{n+1} = 4|X_n = 4) = \frac{1}{2} \ .$$

all other transition probabilities, given $X_n = 4$ *are* 0*.*

$$P(X_{n+1} = 5|X_n = 5) = P(X_{n+1}\frac{3}{4} \quad \text{and} \quad P(X_{n+1} = 4|X_n = 5) = \frac{1}{4} \ .$$

all other transition probabilities, given $X_n = 5$ *are* 0*.*

The transition matrix is then the 6×6 *matrix*

$$P = \frac{1}{4}\begin{bmatrix} 3 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{bmatrix} .$$

To find $\boldsymbol{\pi}_\infty$, we find all solutions of the equation $(P^T - I)\mathbf{x} = 0$, since we want the left eigenvectors of P with eigenvalue 1, so we solve for the right eigenvectors of P^T with eigenvalue 1.

$$4(P^T - I) = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

By row reduction, this is equivalent to

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\text{`} & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} ,$$

and no evidently $x_0 = x_1 = x_2 = x_3 = x_4 = x_5$. That is, the eigenvectors are all multiples of $\mathbf{v}_1$, the vector each of whose entries is 1. Normalizing to male this a probability vector,

$$\boldsymbol{\pi}_\infty = \frac{1}{6}(1, 1, 1, 1, 1, 1) \ .$$

Actually we could do this without any computation: The matrix P is symmetric, and clearly you can get from any state to any other state in exactly 5, staying put if you arrive early – there is positive probability not to move from each state. So $P^5_{i,j} > 0$ for all i, j. Hence $\mathbf{v}_1$ spans the eigenspace of P for eigenvalue 1. But P is symmetric, so the same is true of P^T.

The runner goes barefoot only if there are no shoes at the front door, (probability $\frac{1}{6}$) and they go to the front door at the start of the run (probability $\frac{1}{2}$), or if there are no shoes at the back door (probability $\frac{1}{6}$) and they go to the back door at the start of the run (probability $\frac{1}{2}$). So the fraction of the time they run barefoot is

$$\frac{1}{2}\frac{1}{6} + \frac{1}{2}\frac{1}{6} = \frac{1}{6} \ .$$

0.5 Proof of Theorem 0.13

We now prove Theorem 0.13, assuming Theorem 0.12, which we prove next.

0.15 LEMMA. *Suppose the Markov chain (X_j) is such that (0.10) is satisfied. Let f and g be two real valued functions on $\{1, \ldots, M\}$ such that*

$$\max\{|f(j)| \ ; \ 1 \le j \le M\} \le 1 \quad \text{and} \quad \max\{|g(j)| \ ; \ 1 \le j \le M\} \le 1 \ .$$

Suppose also that $\langle g \rangle_\infty = 0$. Then

$$|\mathrm{E}(f(X_m)g(X_{m+n})| \le Cr^n \ .$$

Proof. $\mathrm{E}(f(X_m)g(X_{m+n}) = \sum_{j=1}^{M} f(j)\mathrm{E}(g(X_{m+n}|X_m = j)(\boldsymbol{\pi}_m)_j$. Let $\boldsymbol{\delta}_j$ denote the probability vector whose jth entry is 1. By the Markov property, $\mathrm{E}(g(X_{m+n}|X_m = j) = \boldsymbol{\delta}_j P^n \cdot \mathbf{g}$. and hence

$$\mathrm{E}(f(X_m)g(X_{m+n}) = \sum_{j=1}^{M} f(j)\left(\boldsymbol{\delta}_j P^n \cdot \mathbf{g}\right)(\boldsymbol{\pi}_m)_j \ . \tag{0.12}$$

where $\mathbf{g}$ is the vector whose k entry is $f(k)$. But for each j, since $\langle g\rangle_\infty = \boldsymbol{\pi}_\infty \cdot \mathbf{g} = 0$,

$$\boldsymbol{\delta}_j P^n \cdot \mathbf{g} = \boldsymbol{\delta}_j P^n \cdot \mathbf{g} - \boldsymbol{\pi}_\infty \cdot \mathbf{g} = (\boldsymbol{\delta}_j P^n - \boldsymbol{\pi}_\infty) \cdot \mathbf{g} \ .$$

Therefore,

$$|\boldsymbol{\delta}_j P^n \cdot \mathbf{g}| = |(\boldsymbol{\delta}_j P^n - \boldsymbol{\pi}_\infty) \cdot \mathbf{g}| \le \|\boldsymbol{\delta}_j P^n - \boldsymbol{\pi}_\infty\|_1 \le Cr^n \ ,$$

using the fact that $|g(j)| \le 1$ for all j. Going back to (0.12) and using the fact that $|f(j)| \le 1$ for all j,

$$|\mathrm{E}(f(X_m)g(X_{m+n})| \le Cr^n \sum_{j=1}^{M} (\boldsymbol{\pi}_m)_j = Cr^n \ .$$

□

Proof of Theorem 0.12. We compute, using Lemma 0.15 in the first inequality,

$$\begin{aligned}
\mathrm{E}((\overline{f(X)}_N)^2) &= \frac{1}{N^2}\sum_{i,j=1}^{N} \mathrm{E}(f(X_i)f(X_j)) \\
&= \frac{1}{N^2}\sum_{i,j=1}^{N} \mathrm{E}(f(X_i)f(X_j)) \\
&= \frac{2}{N^2}\sum_{i=1}^{N}\left(\sum_{j=i}^{N} \mathrm{E}(f(X_i)f(X_j))\right) \\
&\le \frac{2}{N^2}\sum_{i=1}^{N}\left(\sum_{j=i}^{N} Cr^{j-1}\right) \\
&\le \frac{2}{N^2}\sum_{i=1}^{N}\left(\sum_{j=i}^{\infty} Cr^{j-1}\right) \\
&\le \frac{2}{N^2}\sum_{i=1}^{N}\left(C\frac{1}{1-r}\right) = \frac{2C}{N(1-r)}
\end{aligned}$$

Therefore, for any $\epsilon > 0$, by Markov's inequality,

$$P(|\overline{f(X)}_N| > \epsilon) \le \frac{K}{N\epsilon^2} \ . \tag{0.13}$$

□

0.6 Proof of Theorem 0.12

In class, we gave a proof by means of "coupling". Here we give a proof based in analysis fo the eigenvectors and eigenvalues of the transition matrix P.

First of all, to discuss convergence, we need to specify the notions of distance that we shall use to measure how close we have come to the limit. The word "close" needs to be given a precise, quantitative mathematical meaning. In other words, we need to specify the *metrics* that we shall use. We need two of these: one metric for vectors and one metric for matrices.

For any vector $\mathbf{v}$ in $\mathbb{C}^M$, define

$$\|\mathbf{v}\|_1 = \sum_{j=2}^{M} |v_j| \ . \tag{0.14}$$

It is easy to see that for all $\mathbf{v}, \mathbf{w} \in \mathbb{C}^N$, $\|\mathbf{v}+\mathbf{w}\|_1 \le \|\mathbf{v}\|_1 + \|\mathbf{w}\|_1$ simply because $|v_j + w_j| \le |v_j| + |w_j|$ for each j. Therefore, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^N$,

$$\|\mathbf{u}-\mathbf{w}\|_1 = \|(\mathbf{v}-\mathbf{w}) + (\mathbf{v}-\mathbf{w})\|_1 \le \|\mathbf{v}-\mathbf{w}\|_1 + \|\mathbf{v}-\mathbf{w}\|_1 \ . \tag{0.15}$$

Define $d_1(\mathbf{v}, \mathbf{w}) := \|\mathbf{v}-\mathbf{w}\|_1$. Then (0.15) can be rewritten as

$$d_1(u, w) \le d_1(u, v) + d_1(v, w) \ . \tag{0.16}$$

In other words, the function $(\mathbf{v}, \mathbf{w}) \mapsto d_1(\mathbf{v}, \mathbf{w})$ satisfies the triangle inequality, and defines a metric on $\mathbb{C}^M$, the other conditions being obviously satisfied. The function $\mathbf{v} \mapsto \|\mathbf{v}\|_1$ has one more simple, but important, property:

0.16 LEMMA. *For all $\mathbf{v} \in \mathbb{C}^M$ and all $\lambda \in \mathbb{C}$,*

$$\|\lambda \mathbf{v}\|_1 = |\lambda| \|\mathbf{v}\|_1 \ .$$

Proof. This follows immediately from the fact that for each j, $|\lambda v_j| = |\lambda||v_j|$. □

We now define a metric on the set of $M \times M$ matrices that is intimately related to the metric we have just defined on C^M. For any $M \times M$ matrix A, we define

$$\|A\|_{1,1} = \max\{ \|A\mathbf{v}\|_1 \ : \ \|\mathbf{v}\|_1 = 1 \} \ .$$

The maximum is attained since $\{\mathbf{v} \ : \ \|\mathbf{v}\|_1 = 1\}$ is a closed and bounded subset of $\mathbb{C}^M$, and $\mathbf{v} \mapsto \|A\|_{1,1}$ is continuous on $\mathbb{C}^M$.

For any two $M \times M$ matrices A and B, let $\mathbf{v}_0$ be such $\|A+B\|_{1,1} = \|(A+B)\mathbf{v}_0\|_1$. Then

$$\|(A+B)\mathbf{v}_0\|_1 = \|A\mathbf{v}_0 + B\mathbf{v}_0\|_1 \le \|A\mathbf{v}_0\|_1 + \|B\mathbf{v}_0\|_1 \ .$$

Since $\mathbf{v}_0$ satisfies the condition $\|\mathbf{v}\|_1 \le 1$, but it may not be the maximizer for either A or B just because it is the maximizer for $A+B$, $\|A\mathbf{v}_0\|_1 \le \|A\|_{1,1}$ and $\|B\mathbf{v}_0\|_1 \le \|B\|_{1,1}$. Hence

$$\|A+B\|_{1,1} \le \|A\|_{1,1} + \|B\|_{1,1} \ .$$

Then just as for vectors, it follows that if we define

$$d_{1,1}(A,B) = \|A - B\|_{1,1} \ ,$$

this function satisfies the triangle inequality and defines a metric. This metric is compatible with matrix multiplication in the following sense:

0.17 LEMMA. *Let A and B be any two $M \times M$ matrices, Then*

$$\|AB\|_{1,1} \leq \|A\|_{1,1}\|B\|_{1,1} \ ,$$

so that, in particular, for any three $M \times M$ matrices A, B and C,

$$d_{1,1}(CA,CB) \leq \|C\|_{1,1} d_{1,1}(A,B) \qquad \text{and} \qquad d_{1,1}(AC,BC) \leq \|C\|_{1,1} d_{1,1}(A,B) \ .$$

Proof. Let $\mathbf{v}$ satisfy $\|\mathbf{v}\|_1 \leq 1$. Then by the definition of $\|\cdot\|_{1,1}$, $\|AB\mathbf{v}\|_1 = \|A(B\mathbf{v})\|_1$. Then as long as $B\mathbf{v} \neq 0$, define

$$\mathbf{w} = \frac{1}{\|B\mathbf{v}\|_1} B\mathbf{v} \ .$$

By Lemma 0.16, $\|\mathbf{w}\|_1 = 1$ and then by definition

$$\|AB\mathbf{v}\|_1 = \|A(B\mathbf{v})\|_1 = \|B\mathbf{v}\|_1 \|A\mathbf{w}\|_1 \leq \|B\|_{1,1}\|A\|_{1,1} \ .$$

□

Now we apply this to study the rate of convergence to equilibrium for Markov chains. Suppose first that P is diagonalizable. Then from (0.5) and (0.6):

$$P^n - P_\infty = V(\Lambda^n - \Lambda_\infty)V^{-1} = V(\Lambda - \Lambda_\infty)^n V^{-1} \ ,$$

using the specific structure of Λ and Λ_∞ to obtain the last equality. Taking the transpose, and defining $Q = P^T$, $Q_\infty = P_\infty^T$ and $W = V^T$, we can rewrite this is

$$Q^n - Q_\infty = W^{-1}(\Lambda - \Lambda_\infty)^n W \ ,$$

Let

$$r := \max\{|\lambda_2|, \ldots, |\lambda_N}\}$$

be the second largest number among the absolute values of the eigenvalues of P. We have seen that $r < 1$, and then all of the diagonal entries of $\Lambda - \Lambda_\infty$ are bounded in magnitude by r, so that $\|\Lambda - \Lambda_\infty\|_{1,1} \leq r$, and then by Lemma 0.17 and a simple induction, for all n,

$$\|(\Lambda - \Lambda_\infty)^n\|_{1,1} \leq r^n \ .$$

By Lemma 0.17 once more,

$$\|Q^n - Q_\infty\|_{1,1} \leq (\|W\|_{1,1}\|W^{-1}\|_{1,1}) r^n \ .$$

That is, letting C denote the constant $C := \|W\|_{1,1}\|W^{-1}\|_{1,1}$

$$\|Q^n - Q_\infty\|_{1,1} \leq Cr^n \ .$$

Now let $\boldsymbol{\pi}_0$ be any initial probability vector, so that $\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 P^n$ is the probability vector specifying the distribution of X_n, Then

$$\boldsymbol{\pi}_n^T = Q^n \boldsymbol{\pi}_0^T \qquad \text{and} \quad Q_\infty \boldsymbol{\pi}_0^T = \boldsymbol{\pi}_\infty^T \ ,$$

where the second equality holds because each column of Q_∞ is $\boldsymbol{\pi}_\infty^T$, and $\boldsymbol{\pi}_0$ is a probability vector, so that $Q_\infty \boldsymbol{\pi}_0^T$ is a weighted average of the columns of Q_∞. Therefore,

$$\|\boldsymbol{\pi}_n^T - \boldsymbol{\pi}_\infty^T\|_1 = \|Q^n \boldsymbol{\pi}_0^T - Q_\infty \boldsymbol{\pi}_0^T\|_1 \leq \|Q^n - Q_\infty\|_{1,1}\|\boldsymbol{\pi}_0\|_1 = \|Q^n - Q_\infty\|_{1,1} \leq Cr^n \ .$$

That is,

$$\sum_{j=1}^{M} |(\boldsymbol{\pi}_n)_j - (\boldsymbol{\pi}_\infty)_j| \leq Cr^n \ , \tag{0.17}$$

showing that the probability vectors $\boldsymbol{\pi}_n$ converge exponentially fast to $\boldsymbol{\pi}_\infty$ in the metric d_1. Notice that (0.10) is much stronger that the statement that $|(\boldsymbol{\pi}_n)_j - (\boldsymbol{\pi}_\infty)_j| \leq Cr^n$ for each j, especially if M is very large.

A similar result may be proved using generalized eigenvectors when P is not diagonalizable, but since it may be hard to find the eigenvalue that has the second largest magnitude, it may be hard to compute r. Fortunately, there is another way, relying more on probabilistic reasoning than on linear algebra, but still using the same metrics.