Stochastic Matrices

The following 3×3 matrix defines a discrete time Markov process with three states:

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$$

where P_{ij} is the probability of going from $j \to i$ in one step. A *stochastic matrix* satisfies the following conditions:

$$\forall_{i,j} P_{ij} \geq 0$$

and

$$\forall_j \sum_{i=1}^{M} P_{ij} = 1.$$

Example

The following 3×3 matrix defines a discrete time Markov process with three states:

$$\mathbf{P} = \begin{bmatrix} 0.90 & 0.01 & 0.09 \\ 0.01 & 0.90 & 0.01 \\ 0.09 & 0.09 & 0.90 \end{bmatrix}$$

where $P_{23} = 0.01$ is the probability of going from $3 \rightarrow 2$ in one step. You can verify that

$$\forall_{i,j} P_{ij} \geq 0$$

and

$$\forall_j \sum_{i=1}^{3} P_{ij} = 1.$$

Example (contd.)

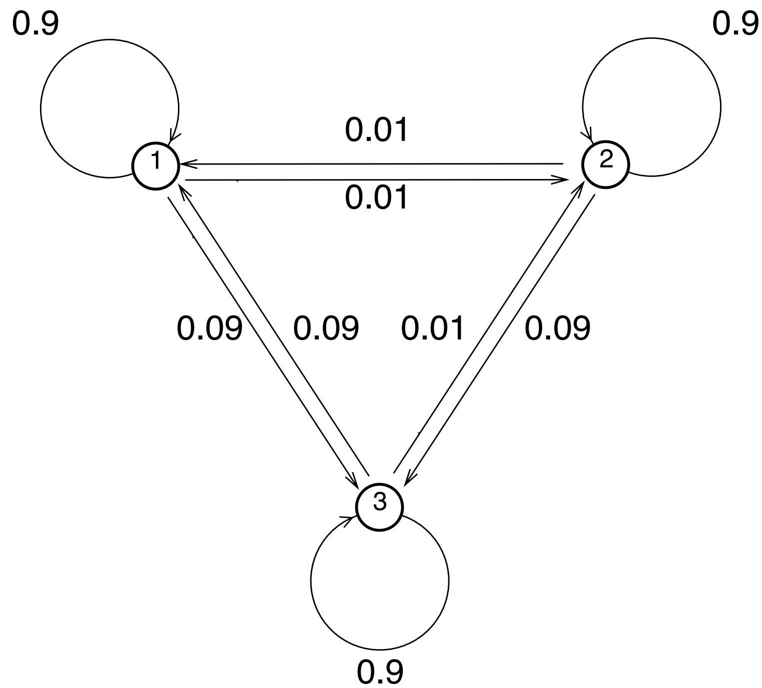

Figure 1: Three-state Markov process.

Single Step Transition Probabilities

$$
\begin{aligned}
\mathbf{x}^{(1)} &= \mathbf{P}\mathbf{x}^{(0)} \\
\mathbf{x}^{(2)} &= \mathbf{P}\mathbf{x}^{(1)} \\
&\vdots \\
\mathbf{x}^{(t+1)} &= \mathbf{P}\mathbf{x}^{(t)}
\end{aligned}
$$

$$
\underbrace{\begin{bmatrix} 0.641 \\ 0.188 \\ 0.171 \end{bmatrix}}_{\mathbf{x}^{(t+1)}} = \underbrace{\begin{bmatrix} 0.90 & 0.01 & 0.09 \\ 0.01 & 0.90 & 0.01 \\ 0.09 & 0.09 & 0.90 \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} 0.7 \\ 0.2 \\ 0.1 \end{bmatrix}}_{\mathbf{x}^{(t)}}
$$

$$
x_i^{(t+1)} = \sum_{j=1}^{M} P_{ij}\, x_j^{(t)}
$$

n-step Transition Probabilities

Observe that $\mathbf{x}^{(3)}$ can be written as follows:

$$\begin{aligned}\mathbf{x}^{(3)} &= \mathbf{P}\mathbf{x}^{(2)} \\ &= \mathbf{P}\left(\mathbf{P}\mathbf{x}^{(1)}\right) \\ &= \mathbf{P}\left(\mathbf{P}\left(\mathbf{P}\mathbf{x}^{(0)}\right)\right) \\ &= \mathbf{P}^3\mathbf{x}^{(0)}.\end{aligned}$$

n-step Transition Probabilities (contd.)

Similar logic leads us to an expression for $\mathbf{x}^{(n)}$:

$$\begin{aligned} \mathbf{x}^{(n)} &= \underbrace{\mathbf{P}\left(\mathbf{P}\ldots\left(\mathbf{P}\mathbf{x}^{(0)}\right)\right)}_{n} \\ &= \mathbf{P}^{n}\mathbf{x}^{(0)}. \end{aligned}$$

An n-step transition probability matrix can be defined in terms of a single step matrix and a $(n-1)$-step matrix:

$$\left(\mathbf{P}^{n}\right)_{ij} = \sum_{k=1}^{M} P_{ik}\left(\mathbf{P}^{n-1}\right)_{kj}.$$

Analysis of Two State Markov Process

$$\mathbf{P} = \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix}$$

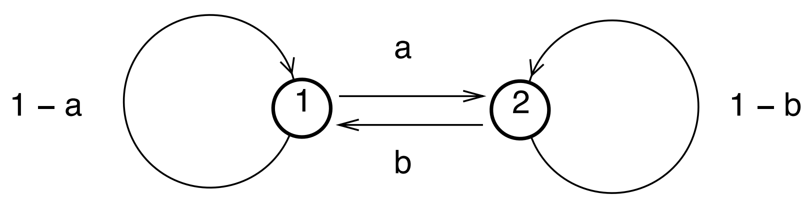

Figure 2: Two-state Markov process.

Analysis of Two State Markov Process (contd.)

It is readily shown by induction that the n-step transition probabilities for the two state Markov process are given by the following formula:

$$\mathbf{P}^n = \frac{1}{a+b}\begin{bmatrix} b & b \\ a & a \end{bmatrix} + \frac{(1-a-b)^n}{a+b}\begin{bmatrix} a & -b \\ -a & b \end{bmatrix}.$$

For conciseness, we introduce the following abbreviations:

$$\mathbf{P}_1 = \begin{bmatrix} b & b \\ a & a \end{bmatrix}$$

and

$$\mathbf{P}_2 = \begin{bmatrix} a & -b \\ -a & b \end{bmatrix}.$$

Analysis of Two State Markov Process (contd.)

The following identities will also help:

- Identity 1

$$\mathbf{P}_1\mathbf{P} = \begin{bmatrix} b & b \\ a & a \end{bmatrix} \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix} = \begin{bmatrix} b & b \\ a & a \end{bmatrix} = \mathbf{P}_1$$

- Identity 2

$$\begin{aligned} \mathbf{P}_2\mathbf{P} &= \begin{bmatrix} a & -b \\ -a & b \end{bmatrix} \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix} \\ &= \begin{bmatrix} a-a^2-ab & -b+ab+b^2 \\ a^2-a+ab & -ab+b-b^2 \end{bmatrix} \\ &= (1-a-b)\mathbf{P}_2 \end{aligned}$$

Analysis of Two State Markov Process (contd.)

To do a proof by induction, we need to prove the basis step and the induction step. First the basis step:

$$\begin{aligned}
\mathbf{P}^1 &= \frac{1}{a+b}\begin{bmatrix} b & b \\ a & a \end{bmatrix} + \frac{(1-a-b)}{a+b}\begin{bmatrix} a & -b \\ -a & b \end{bmatrix} \\
&= \frac{1}{a+b}\begin{bmatrix} b+a-a^2-ab & b-b+ab+b^2 \\ a-a+a^2+ab & a+b-ab-b^2 \end{bmatrix} \\
&= \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix} \\
&= \mathbf{P}.
\end{aligned}$$

Analysis of Two State Markov Process (contd.)

Now we prove the induction step:

$$
\begin{aligned}
\mathbf{P}^n\mathbf{P} &= \left(\frac{1}{(a+b)}\mathbf{P}_1 + \frac{(1-a-b)^n}{(a+b)}\mathbf{P}_2\right)\mathbf{P} \\
&= \frac{1}{(a+b)}\mathbf{P}_1\mathbf{P} + \frac{(1-a-b)^n}{(a+b)}\mathbf{P}_2\mathbf{P} \\
&= \frac{1}{(a+b)}\mathbf{P}_1 + \frac{(1-a-b)^{n+1}}{(a+b)}\mathbf{P}_2 \\
&= \mathbf{P}^{n+1}.
\end{aligned}
$$

Limiting Distribution

$$\mathbf{P} = \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix}$$

$$\mathbf{P}^n = \frac{1}{a+b}\begin{bmatrix} b & b \\ a & a \end{bmatrix} + \frac{(1-a-b)^n}{a+b}\begin{bmatrix} a & -b \\ -a & b \end{bmatrix}$$

Note that $|1-a-b| < 1$ when $0 < a < 1$ and $0 < b < 1$. Thus, $|1-a-b|^n \to 0$ as $n \to \infty$.

Limiting Distribution (contd.)

It follows that:

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \frac{b}{a+b} & \frac{b}{a+b} \\ \frac{a}{a+b} & \frac{a}{a+b} \end{bmatrix}.$$

It is easy to show that $[b/(a+b), a/(a+b)]^{\mathrm{T}}$ is an eigenvector with eigenvalue one of $\mathbf{P}$:

$$\begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix} \begin{bmatrix} \frac{b}{a+b} \\ \frac{a}{a+b} \end{bmatrix} = \begin{bmatrix} \frac{b}{a+b} - \frac{ab}{a+b} + \frac{ab}{a+b} \\ \frac{ab}{a+b} + \frac{a}{a+b} - \frac{ab}{a+b} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{b}{a+b} \\ \frac{a}{a+b} \end{bmatrix}.$$

Spectral Theorem (reprise)

$$
\begin{aligned}
\mathbf{P}^n &= \mathbf{X}\Lambda^n\,\mathbf{Y}^{\mathrm{T}} \\
&= \lambda_1^n \mathbf{x}_1 \mathbf{y}_1^{\mathrm{T}} + \lambda_2^n \mathbf{x}_2 \mathbf{y}_2^{\mathrm{T}} \\
&= \begin{bmatrix} \frac{b}{a+b} \\ \frac{a}{a+b} \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} + (1-a-b)^n \begin{bmatrix} -1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{-a}{a+b} & \frac{b}{a+b} \end{bmatrix} \\
&= \frac{1}{a+b} \begin{bmatrix} b & b \\ a & a \end{bmatrix} + \frac{(1-a-b)^n}{a+b} \begin{bmatrix} a & -b \\ -a & b \end{bmatrix}.
\end{aligned}
$$

Existence of Limiting Distribution

In order to understand when a Markov process will have a limiting distribution and when it will not we will

- Prove that a stochastic matrix has no eigenvalue with magnitude greater than one.
- Prove that a stochastic matrix always has at least one eigenvalue equal to one.
- Identify those conditions in which this eigenvalue will be the unique eigenvalue of unit magnitude.

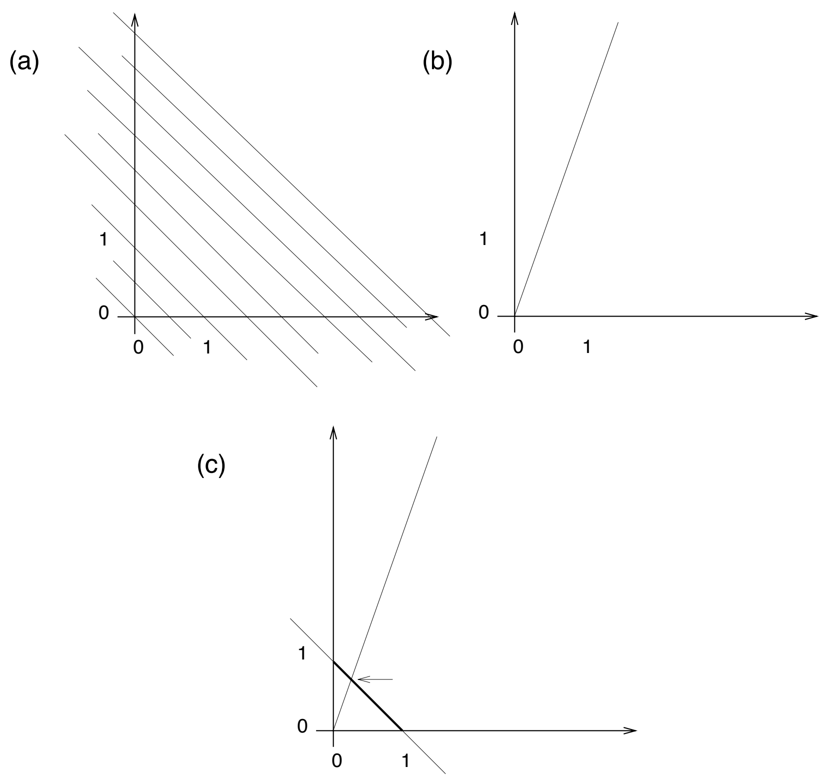

Figure 3: (a) $\begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix}$ maps pts. in lines of constant 1−norm to pts. in same line. These lines have slope equal to -1. (b) $\begin{bmatrix} b & b \\ a & a \end{bmatrix}$ maps points to line of slope $\frac{a}{b}$ through origin. (c) $\begin{bmatrix} \frac{b}{a+b} & \frac{b}{a+b} \\ \frac{a}{a+b} & \frac{a}{a+b} \end{bmatrix}$ maps distributions (thick segment) to point $\begin{bmatrix} \frac{b}{a+b} \\ \frac{a}{a+b} \end{bmatrix}$.

Spectral Radius

- The *spectral radius*, $\rho(\mathbf{A})$, of a matrix $\mathbf{A}$ is defined as the magnitude of its largest eigenvalue.
- The $1-$norm of a vector $\mathbf{x}$ is defined as follows:

$$||\mathbf{x}||_1 = \sum_i |x_i|.$$

- The $1-$norm of a matrix $\mathbf{A}$ is defined as follows:

$$||\mathbf{A}||_1 = \max_{||\mathbf{x}||_1=1} ||\mathbf{A}\mathbf{x}||_1.$$

- The *Gelfand spectral radius theorem* states that

$$\rho(\mathbf{A}) = \lim_{n\to\infty} ||\mathbf{A}^n||_1^{\frac{1}{n}}.$$

Spectral Radius (contd.)

Lemma 3.1 Let $\mathbf{P}$ be stochastic. Then

$$||\mathbf{P}||_1 = 1.$$

Proof:

$$\begin{aligned}
y_i &= \sum_j P_{ij} x_j \\
\sum_i y_i &= \sum_i \sum_j P_{ij} x_j \\
\sum_i y_i &= \sum_j \sum_i P_{ij} x_j \\
\sum_i y_i &= \sum_j x_j \sum_i P_{ij} \\
\sum_i y_i &= \sum_j x_j.
\end{aligned}$$

It follows that

$$||\mathbf{Px}||_1 = ||\mathbf{x}||_1.$$

Spectral Radius (contd.)

Consequently,

$$
\begin{aligned}
||\mathbf{P}||_1 &= \max_{||\mathbf{x}||_1=1} ||\mathbf{P}\mathbf{x}||_1 \\
||\mathbf{P}||_1 &= \max_{||\mathbf{x}||_1=1} ||\mathbf{x}||_1 \\
||\mathbf{P}||_1 &= 1.
\end{aligned}
$$

Spectral Radius (contd.)

Lemma 3.2 The product of two stochastic matrices is stochastic.

Proof: Let $\mathbf{P}$ and $\mathbf{Q}$ be stochastic, then

$$\begin{aligned}
(\mathbf{PQ})_{ij} &= \sum_k P_{ik} Q_{kj} \\
\sum_i (\mathbf{PQ})_{ij} &= \sum_i \sum_k P_{ik} Q_{kj} \\
&= \sum_k \sum_i P_{ik} Q_{kj} \\
&= \sum_k Q_{kj} \sum_i P_{ik} \\
&= \sum_k Q_{kj} \\
&= 1.
\end{aligned}$$

Spectral Radius (contd.)

Theorem 3 The spectral radius, $\rho(\mathbf{P})$, of a stochastic matrix, $\mathbf{P}$, is one.

Proof: It is straightforward to show by induction on n and Lemma 3.2 that $\mathbf{P}^n$ is stochastic for all integers, $n > 0$. It follows, by Lemma 3.1, that

$$||\mathbf{P}^n||_1 = 1$$

for all integers, $n > 0$. Consequently,

$$\rho(\mathbf{P}) = \lim_{n\to\infty} ||\mathbf{P}^n||_1^{\frac{1}{n}} = 1$$

by the Gelfand spectral radius theorem.

Existence of Limiting Distribution (contd.)

We just showed that a stochastic matrix cannot have an eigenvalue with magnitude greater than one. We will now show that every stochastic matrix has at least one eigenvalue equal to one.

Existence of $\lambda = 1$

Let $\mathbf{P}$ be a stochastic matrix. Since $\sum_{i=1}^{M} P_{ij} = 1$ and $\sum_{i=1}^{M} I_{ij} = 1$, it follows that:

$$
\begin{aligned}
0 &= \sum_{i=1}^{M} P_{ij} - \sum_{i=1}^{M} I_{ij} \\
&= \sum_{i=1}^{M} \left(P_{ij} - I_{ij}\right).
\end{aligned}
$$

Consequently, the rows of $\mathbf{P} - \mathbf{I}$ are not linearly independent. Consequently, $\mathbf{P} - \mathbf{I}$ is singular:

$$\det(\mathbf{P} - \mathbf{I}) = 0.$$

Existence of $\lambda = 1$ (contd.)

Recall that $\mathbf{x}$ is an *eigenvector* of $\mathbf{P}$ with *eigenvalue*, λ, iff:

$$\lambda \mathbf{x} = \mathbf{P}\mathbf{x}.$$

The eigenvalues of $\mathbf{P}$ are the roots of the *characteristic polynomial*:

$$\det(\mathbf{P} - \lambda \mathbf{I}) = 0.$$

Since $\det(\mathbf{P} - \mathbf{I}) = 0$, it follows that $\lambda = 1$ is an eigenvalue of $\mathbf{P}$.

Existence of Limiting Distribution (contd.)

We just showed that

- A stochastic matrix cannot have an eigenvalue λ with magnitude greater than one.
- Every stochastic matrix has at least one eigenvalue λ_1 equal to one.

We will now identify those conditions in which this eigenvalue will be the unique eigenvalue of unit magnitude.

Uniqueness of $|\lambda| = 1$

A matrix, **P**, is *positive* if and only if for all i and j it is true that $P_{ij} > 0$. In 1907, Perron proved that every positive matrix has a positive eigenvalue, λ_1, with larger magnitude than the remaining eigenvalues. If **P** is positive and of size $M \times M$ then:

$$\lambda_1 > |\lambda_i| \text{ for } 1 < i \leq M.$$

Irreducibility

Two states, i and j in a Markov process *communicate* iff 1) i can be reached from j with non-zero probability:

$$\sum_{n=1}^{N_1} (\mathbf{P}^n)_{ij} > 0$$

and 2) j can be reached from i with non-zero probability:

$$\sum_{n=1}^{N_2} (\mathbf{P}^n)_{ji} > 0$$

for some sufficiently large N_1 and N_2. If every state communicates with every other state, then the Markov process is *irreducible*.

Aperiodicity

A state i in a Markov process is *aperiodic* if for all sufficiently large N, there is a non-zero probability of returning to i in N steps:

$$\left(\mathbf{P}^N\right)_{ii} > 0.$$

If a state is aperiodic, then every state it communicates with is also aperiodic. If a Markov process is irreducible, then all states are either periodic or aperiodic.

Positive Stochastic Matrices

Theorem 4 If $\mathbf{P}$ is irreducible and aperiodic then $\mathbf{P}^N$ is positive for some sufficiently large N:

$$\forall_{i,j} \left(\mathbf{P}^N\right)_{ij} > 0.$$

Proof: Let $\mathbf{P}$ be irreducible and aperiodic and let

$$N_{ij} = \min_{\left(\mathbf{P}^N\right)_{ij} > 0} (N).$$

We observe that N_{ij} is guaranteed to exist for all i and j by *irreducibility.*

Positive Stochastic Matrices (contd.)

Now let

$$\begin{aligned} N &= M + \max_{i,j}(N_{ij}) \\ N - N_{ij} &= M + \max_{i,j}(N_{ij}) - N_{ij} \\ &\geq M. \end{aligned}$$

where M satisfying

$$\forall_i \left(\mathbf{P}^M\right)_{ii} > 0$$

is guaranteed to exist by *aperiodicity*.

Positive Stochastic Matrices (contd.)

Now let

$$\left(\mathbf{P}^N\right)_{ij} = \left(\mathbf{P}^{N_{ij}+N-N_{ij}}\right)_{ij}$$

which is just

$$\begin{aligned}\left(\mathbf{P}^N\right)_{ij} &= \sum_k \left(\mathbf{P}^{N_{ij}}\right)_{ik}\left(\mathbf{P}^{N-N_{ij}}\right)_{kj} \\ &\geq \left(\mathbf{P}^{N_{ij}}\right)_{ij}\left(\mathbf{P}^{N-N_{ij}}\right)_{jj}\end{aligned}$$

We observe that $\left(\mathbf{P}^{N_{ij}}\right)_{ij} > 0$ by definition of N_{ij}. Since $N - N_{ij} \geq M$, it follows that $\left(\mathbf{P}^{N-N_{ij}}\right)_{jj} > 0$. We therefore see that

$$\begin{aligned}\left(\mathbf{P}^N\right)_{ij} &\geq \left(\mathbf{P}^{N_{ij}}\right)_{ij}\left(\mathbf{P}^{N-N_{ij}}\right)_{jj} \\ &> 0.\end{aligned}$$

Uniqueness of $|\lambda| = 1$ (contd.)

When a Markov process is irreducible and aperiodic, then $\mathbf{P}^N$ for some sufficiently large N will be a positive matrix and its unique positive eigenvalue of largest magnitude, λ_1, equals one:

$$\mathbf{x}_1 = \mathbf{P}^N \mathbf{x}_1.$$

Since one is the unique positive eigenvalue of largest magnitude of $\mathbf{P}^N$, it follows that one is also the unique positive eigenvalue of largest magnitude of $\mathbf{P}$.

Periodic and Irreducible Markov Process

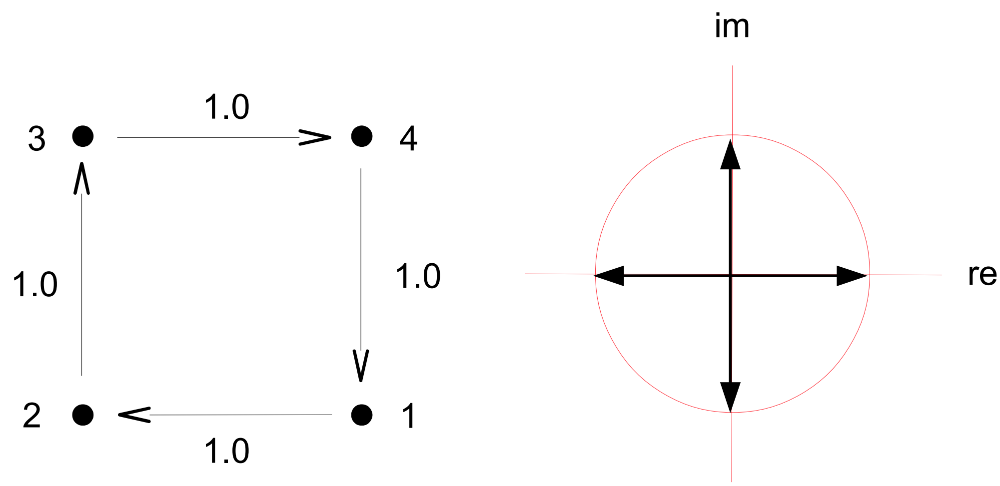

Figure 4: A Markov process with four states and period four (left). The transition matrix has four distinct eigenvalues on the unit circle in the complex plane (right). There is no limiting distribution.

Aperiodic and Irreducible Markov Process

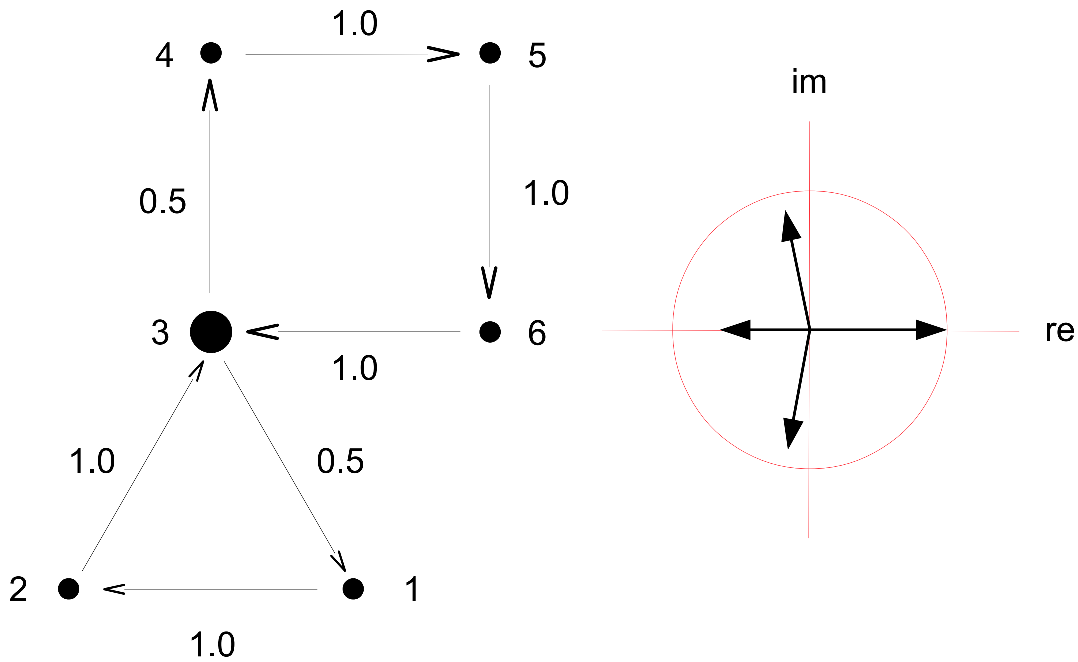

Figure 5: An aperiodic Markov process with six states (left). The transition matrix has two eigenvalues of magnitude zero, one eigenvalue of unit magnitude, and three eigenvalues with magnitude less than one (right). Because the rank of $\lim_{n\to\infty}\mathbf{P}^n = 1$, there is a unique limiting distribution. In the limit, the process visits state 3 twice as often as the other states (which are visited with equiprobability).

Limiting Distributions

Let $\mathbf{x}^{(0)}$ be an initial distribution. We can write $\mathbf{x}^{(0)}$ as a linear combination of the eigenvectors of $\mathbf{P}$:

$$\mathbf{x}^{(0)} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + ... + c_M\mathbf{x}_M.$$

Can $\mathbf{x}^{(1)} = \mathbf{P}\mathbf{x}^{(0)}$ also be written as a linear combination of eigenvectors?

$$\mathbf{x}^{(1)} = \mathbf{P}c_1\mathbf{x}_1 + \mathbf{P}c_2\mathbf{x}_2 + ... + \mathbf{P}c_M\mathbf{x}_M$$

Limiting Distributions (contd.)

Since $\mathbf{P}\mathbf{x}_i = \lambda_i \mathbf{x}_i$ it follows that:

$\mathbf{x}^{(1)} = \lambda_1 c_1 \mathbf{x}_1 + \lambda_2 c_2 \mathbf{x}_2 + \ldots + \lambda_M c_M \mathbf{x}_M.$

Furthermore, $\mathbf{P}^n \mathbf{x}_i = \lambda_i^n \mathbf{x}_i$. It follows that:

$\mathbf{x}^{(n)} = \lambda_1^n c_1 \mathbf{x}_1 + \lambda_2^n c_2 \mathbf{x}_2 + \ldots + \lambda_M^n c_M \mathbf{x}_M.$

Since $\lambda_1 = 1$ and $|\lambda_i| < 1$ for all i, in the limit as $\mathbf{n} \to \infty$:

$$\lim_{n\to\infty} \mathbf{x}^{(n)} = c_1 \mathbf{x}_1.$$

Observe that $\mathbf{x}_1$ is independent of $\mathbf{x}^{(0)}$ and that c_1 must equal one.

Genetic Drift

- There are $2N$ individuals of which j possess gene variant, A, and the remaining $2N - j$ possess gene variant, B.
- Let $\{0, ..., 2N\}$ be the states of a Markov process modeling the number of individuals in successive generations who possess gene variant, A.
- An individual inherits his gene variant either from his father (50% probability) or his mother (50% probability).

Genetic Drift (contd.)

- The probability of any individual inheriting A is $p_A = j/2N$ and inheriting B is $p_B = 1 - p_A$.
- The probability that exactly k individuals will possess A in the next generation, given that j individuals possess it in the current generation can be modeled by the binomial distribution:

$$P_{kj} = \binom{2N}{k} p_A^k (1 - p_A)^{2N-k}$$

Genetic Drift (contd.)

- Gene variant A becomes extinct if $j = 0$.
- Given an initial population with exactly j individuals possessing gene variant A, what is the probability that gene variant A will become extinct?

Adding Mutation

- Gene variant, A, mutates to gene variant, B, with probability, $p_{A \to B}$.
- Gene variant, B, mutates to gene variant, A, with probability, $p_{B \to A}$.
- The probability that an individual inherits A (and it doesn't mutate to B) or that he inherits B (and it mutates to A) is:

$$p_A = \frac{j}{2N}(1 - p_{A \to B}) + \left(1 - \frac{j}{2N}\right) p_{B \to A}$$