Matrix-vector representation

Chains (think ensemble) evolve according to:
\[ p(z) = \sum_{z'} p(z') T(z', z) \]

Matrix vector representation:
\[ p = Tp' \]

And, after \( n \) iterations after a starting point:
\[ p^{(n)} = T^N p^{(0)} \]

Aside on stochastic Matrices

- A right (row) stochastic matrix has non-negative entries, and its rows sum to one.
- A left (column) stochastic matrix has non-negative entries, and its columns sum to one.
- A doubly stochastic matrix has both properties.

Matrix representation

A single transition is given by
\[ p = Tp' \]

Note what happens for stationary state:
\[ p^* = Tp^* \]

So, \( p^* \) is an eigenvector with eigenvalue one.

And, intuitively, if things converge, \( p^* = T^m p^{(0)} \)

Aside on stochastic Matrices

- \( T \) is a left (column) stochastic matrix.
  - If you are right handed, take the transpose
- The column vector, \( p \), also has non-negative elements, that sum to one (sometimes this is called a stochastic vector).
- Fun facts that we should do on the board
  - The product of a stochastic matrix and vector is a stochastic vector.
  - The product of two stochastic matrices is a stochastic matrix.
Aside on (stochastic) Matrix powers

Consider the eigenvalue decomposition of \( T, \ T = E \Lambda E^{-1} \)

\[ T^N = E \Lambda^N E^{-1} \]

Since \( T^N \) cannot grow without bound, the eigenvalues are inside \([-1,1]\).

In fact, for our situation, the second biggest absolute value of the eigenvalues is less than one (not so easy to prove), which means the biggest one is 1.

Aside on (stochastic) Matrix powers

Recall that we are studying \( E \Lambda^{-1} p \)

We have \( \Lambda^{-1} = \begin{pmatrix} e_i^T \\ 0 \\ \vdots \\ 0 \end{pmatrix} \)

So, \( \Lambda^{-1} p = \begin{pmatrix} e_i^T \cdot p \\ 0 \\ \vdots \\ 0 \end{pmatrix} = ? \)

Aside on (stochastic) Matrix powers

We have \( T^N = E \Lambda^N E^{-1} \)

\[ \Lambda = \begin{pmatrix} 1 \\ \lambda_2 \\ \cdots \\ \lambda_k \end{pmatrix} \quad \text{and} \quad \Lambda^{-1} = \begin{pmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix} \]

\[ \Lambda^{-1} = \begin{pmatrix} e_i^T \\ 0 \\ \vdots \\ 0 \end{pmatrix} \]

Aside on (stochastic) Matrix powers

Write \( p \) in terms of the eigen basis

\[ p = \sum_i a_i e_i \]

\[ e_i^T p = \sum_i a_i e_i^T e_i = a_i \]

and, \( \Lambda^{-1} p = \begin{pmatrix} e_i^T p \\ a_i \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} a_i \\ 0 \\ 0 \\ 0 \end{pmatrix} \)
Aside on (stochastic) Matrix powers

Recall that we are studying \( E \Lambda^{-1} \mathbf{p} \)

\[
\Lambda^{-1} \mathbf{p} = \begin{pmatrix} a_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}
\]

So, \( E \Lambda^{-1} \mathbf{p} = a_1 \mathbf{e}_1 \)

---

Demo

- According to the previous, if \( T \) is a stochastic matrix, then:

\[
\mathbf{p}^* \equiv T^n \mathbf{p}
\]

(No matter what \( \mathbf{p} \)! They all will give the same answer).

Also, \( \mathbf{p}^* \parallel \mathbf{e}^{(1)} \)

---

Aside on (stochastic) Matrix powers

So,

\[
E \Lambda^{-1} \mathbf{p} = e_1 \left( e_1^T \cdot \mathbf{p} \right) \parallel e_1 \parallel \mathbf{p}^*
\]

In summary, \( \mathbf{p}^* \parallel e_1 \) together with \( \mathbf{p}^* \) stochastic means that \( E \Lambda^{-1} \mathbf{p} = \mathbf{p}^* \)

This is true, no matter what the initial \( \mathbf{p} \) is.

So, glossing over details, we have convergence to equilibrium.

---

Justification relies on Perron Frobenius theorem

Let \( A = (a_{ij}) \) be an \( n \times n \) positive matrix, \( a_{ii} > 0 \) for \( 1 \leq i \leq n \). Then the following statements hold:

1. There is a positive real number \( \lambda \), called the Perron root or the Perron-Frobenius eigenvalue, such that \( \lambda \) is an eigenvalue of \( A \) and any other eigenvalue \( \lambda \) (possibly, complex) is strictly smaller than \( \lambda \) in absolute value. \( \lambda \) \( \lambda \) is real and positive. This is known as the Perron-Frobenius theorem.

2. The Perron-Frobenius eigenvalue is simple, i.e., it is a simple root of the characteristic polynomial of \( A \). Consequently, the eigenspace associated to \( \lambda \) is one-dimensional. (This is true for the left eigenspace, i.e., left eigenvectors associated to \( \lambda \) are the same as right eigenvectors associated to \( \lambda \).

3. There exists an eigenvector \( w = (w_1, \ldots, w_n) \) of \( A \) with eigenvalue \( \lambda \) such that all components of \( w \) are positive: \( w_i > 0 \) for all \( 1 \leq i \leq n \). \( w \) is called a positive eigenvector of \( A \). (This is true for the right eigenvector associated to \( \lambda \).

4. If \( A \) is a nonnegative \( n \times n \) matrix, then it has a real positive eigenvalue \( \lambda \) (in fact, the Perron-Frobenius theorem guarantees that \( \lambda \) is an eigenvalue of \( A \)) and a corresponding positive eigenvector \( w \) (known as the Perron vector).

5. \( \lambda \) is the unique largest eigenvalue of \( A \).

6. The Perron-Frobenius theorem states that if \( A \) is a non-negative \( n \times n \) matrix with \( A_{ii} > 0 \) for all \( i \), then \( A \) has a unique positive eigenvector \( \mathbf{v} \) (known as the Perron vector).

7. A Min-max formula: If \( A \) is a non-negative \( n \times n \) matrix with \( A_{ii} > 0 \) for all \( i \), then \( \lambda \) is the largest eigenvalue of \( A \), and \( \mathbf{v} \) is the corresponding eigenvector.

8. The Perron-Frobenius theorem states that for any non-negative \( n \times n \) matrix \( A \), the following inequalities hold:

\[
\min_{i} \sum_{j} a_{ij} \leq \lambda \leq \max_{i} \sum_{j} a_{ij}.
\]

From Wikipedia
Main points about P-F

- The maximal eigenvalue is strictly maximal (item 1).
- The corresponding eigenvector is “simple” (item 2)
- It has all positive (or negative) components (item 3).
- There is no other eigenvector that can be made non-negative.
- The maximal eigenvalue of a stochastic matrix has absolute value 1 (item 8 applied to stochastic matrix).

Aside on (stochastic) Matrix powers

Summary

\( p^* = TP^* \) is an eigenvector with eigenvalue one.

We have written it as \( p^* \| e^i \) because \( e^i \) is the eigenvector normalized to norm 1 (standard form).

Intuitively (perhaps), \( T \) will reduce any component of \( p \) orthogonal to \( p^* \), and \( T^N \) will kill off such components as \( N \to \infty \).

Algebraic proof

Neal ’93 provides an algebraic proof which does not rely on spectral theory.

(A question on the final studies this further for those that are interested).