Finite State Markov Chains

Finite state space with states $i=1, \ldots, N$.

Probability that the state at period $t+1$ is $j$, conditional on the state at period $t$ being $i$, is independent of $t$ and is independent of the value of the state in previous periods:

$$
\Pr \left\{ X_{t+1} = j \mid X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0 \right\} = \Pr \left\{ X_{t+1} = j \mid X_t = i \right\} = p_{ij}
$$

(Does this involve a loss of generality?)

Let $P$ be the matrix of these probabilities:
Here $P$ is a “stochastic matrix”: each of its rows corresponds to a probability distribution, i.e. the elements are non-negative and sum to 1.

With this notation, probability distributions over states are row vectors: i.e. if the distribution over states at time $t$ is given by $\lambda$, the distribution at $t + 1$ is $\lambda P$.

For example, if at time $t$ the state is $i = 2$, the distribution over states $t + 1$ is
\[ [0,1,0,\ldots,0] P = \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,N} \\ P_{2,1} & P_{2,2} & & P_{2,N} \\ P_{N,1} & P_{2,N} & & P_{N,N} \end{bmatrix} = [P_{2,1}, P_{2,2}, \ldots, P_{2,N}]. \]

Now compute the probabilities of states 2 period ahead:

\[
\Pr \{ X_{t+2} = j \mid X_t = i \} = \sum_k p_{kj} \Pr \{ X_{t+1} = k \mid X_t = i \} = \sum_k p_{kj} p_{ik} = (P^2)_{ij}
\]

In general \( \Pr \{ X_{t+\tau} = j \mid X_t = i \} = (P^\tau)_{ij} \):

Looking \( \tau \) periods ahead corresponds to exponentiating the matrix \( P \) \( \tau \) times.
State $j$ is *accessible from state* $i$ if for some integer $n \geq 0$, $P^n_{ij} > 0$:

There is a positive probability of getting from $i$ to $j$ in a finite number of steps. (Note that we allow $n=0$ with $P^0 = I$ so that that every state is accessible from itself.)

State $i$ *communicates with* $j$ if $j$ is *accessible from state* $i$ and $i$ is *accessible from state* $j$.

“Communicate with” is an equivalence relation (it is symmetric, transitive and reflexive- we allow $n=0$ in the definition of accessible to get reflexivity.)

It may be possible to start in one class $C$ and move to some other class $D$; but then its not possible to move from $D$ back to $C$ (or they would be in the same equivalence class.)
A Markov process is *irreducible* if the whole state space belongs to the same class, i.e. if all states communicate with one another.

A state $i$ is *absorbing* if $p_{i,i} = 1$.

A class $C$ is *closed* if \{ $i \in C$ and $j$ is accessible from state $i$ \} implies $j \in C$.

Thus $i$ is absorbing if \{ $i$ \} is closed.

A state $i$ is *recurrent* if $\Pr(X_t = i \text{ for infinitely many } t \mid X_0 = i) = 1$: Starting in state $i$, we expect to return to it infinitely many times.

A state $i$ is *transient* if $\Pr(X_t = i \text{ for infinitely many } t \mid X_0 = i) = 0$: Even if we start in state $i$, we eventually never come back.
Theorem: In a finite state Markov chain:
(a) Every state is either transient or recurrent.
(b) If $i$ communicates with $j$ and $i$ is recurrent, then $j$ is recurrent (so the same is true for transience.)
(c) Every recurrent class is closed, and every closed class is recurrent.
(for a proof see e.g. Norris *Markov Chains* pp. 24-27)

So it’s pretty easy to find the recurrent classes with a finite state space: just look for the closed classes.

Invariant distributions: Probability distributions $\mu$ such that $\mu P = \mu$: the analog of an equilibrium state for a Markov chain.

This says that the invariant distribution is an eigenvector with eigenvalue 1. So we can find the invariant distribution by doing linear algebra.

Every stochastic matrix has at least one invariant distribution. (how would you prove this?)
The invariant distribution need not be unique: consider \( P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \)

Every irreducible stochastic matrix has a unique invariant distribution.

But this unique invariant distribution needn’t correspond to what we’d think of as a long-run equilibrium: Consider \( P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \)

This process is irreducible; it is a simple two-cycle.

Average time in each state converges to .5, but (.5,.5) isn’t a fully accurate description of the long run outcome:
Note that if the initial state is state 1, the distribution of states at any odd-numbered time period is $(1,0)$ and not $(.5,.5)$. So it is not true that $[1, 0] P^k \rightarrow [.5, .5]$; in this sense $[.5,.5]$ may not be a very accurate description of the long term behavior of the system.

Here is a condition to rule this out:

We will say that a stochastic matrix is *regular* or *ergodic* if it is irreducible and in addition there is an integer $k \geq 1$ such that $P^k \gg 0$, i.e. every element of $P^k$ is positive.

So any strictly positive (every element positive) stochastic matrix is regular.

Also a matrix is regular if it is irreducible and there is a state $i$ with $p_{ii} = \Pr(X_{t+1} = i \mid X_t = i) > 0$; this means the chain is “aperiodic.”

This one is easier to check.
**Theorem** (Perron-Frobenius) Suppose $P$ is a regular stochastic matrix. Then $P$ has a unique invariant distribution, i.e. there is a unique probability distribution $\mu^*$ with $\mu^* P = \mu^*$. The eigenvalue 1 is simple (a unique eigenvector has this value) and dominant (all other eigenvalues are smaller in absolute value), so for any initial distribution $\mu(0)$ we have

$$\mu(t) = \mu(0) P^k \to \mu^*.$$ 

In addition, the rate of convergence is pinned down by the second-largest eigenvalue: Order the eigenvalues so that $1 > |\lambda_2| \geq |\lambda_3| \geq \ldots$.

Then $\|\mu(t) - \mu^*\| \leq C |\lambda_2|^t$

(see e.g. Karlin and Taylor *A First Course in Stochastic Processes*, pp. 543-547.)
Now reconsider the matrix \( P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \) which is irreducible but periodic and so not regular.

Note that it has two eigenvectors with eigenvalue whose absolute value is 1: \([1/2, 1/2]\) has eigenvalue 1, \([1/2, -1/2]\) has eigenvalue -1.

(Only one of these eigenvectors is an invariant distribution, the other is invariant but isn’t a distribution.)

\( Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \) is aperiodic but it isn’t regular because it isn’t irreducible.
Now consider adding “small shocks” to the system to obtain

\[ Q' = \begin{bmatrix} 1 - \epsilon & \epsilon \\ \epsilon & 1 - \epsilon \end{bmatrix} \]

This matrix is regular. What is its unique invariant distribution? Eigenvalues?

\[ \lambda^2 - 2\lambda(1 - \epsilon) + (1 - \epsilon)^2 - \epsilon^2 = 0 \]

\[ \lambda = \frac{2(1 - \epsilon) \pm 2\sqrt{(1 - \epsilon)^2 - (1 - \epsilon)^2 + \epsilon^2}}{2} = 1 - \epsilon \pm \epsilon \]

So eigenvalues are 1 and \(1 - 2\epsilon\).

Perron-Frobenius says that if \(\epsilon\) is very small it may take a long time to converge to the invariant distribution. This makes sense, since we can see that if the system starts in state 1, it is likely to be in state 1 for quite a while.
Theorem (Ergodic Theorem) Suppose $P$ is irreducible, and let $\lambda$ be its unique invariant distribution. Then the fraction of time the process spends in each state $i$ converges to $\lambda_i$. That is, $\Pr\left(\frac{1}{s} \sum_{t=0}^{s-1} I(X_t = i) \rightarrow \lambda_i \right) = 1$ irrespective of the initial condition $X_0$.

(This is a consequence of the strong law of large numbers for Markov processes. For a proof see e.g. Norris *Markov Chains* pp. 53-55.)

The ergodic theorem says that the invariant distribution also describes the time-average behavior of the system.