Markov Chain Notes

1 Discrete Time Markov Chains

The following is a brief summary of key concepts and properties of discrete time Markov Chains, much of it adapted from Resnik (2005)[1].

1.1 Equivalence Classes among states & Asymptotic Distributions.

If it is possible to reach state j from state i, we say *j is* ***accessible*** *from i*, denoted as $i \to j$.

If $i \to j$ and $j \to i$, we say that states i and j ***communicate***, which we denote as $i \leftrightarrow j$. This is an equivalence relation (i.e., it is reflexive, symmetric, and transitive) and thus defines equivalence classes among states of a Markov Chain.

If a transition matrix $\mathbf{P}$ is ***irreducible*** (i.e., $i \leftrightarrow j$ for all i,j in the state space), then it has a left eigenvalue $\lambda_* = 1$ (which may or may not be unique!) and the associated left eigenvector $\mathbf{v}$ is positive-valued. Thus, when scaled appropriately, that eigenvector is a ***stationary distribution*** of the Markov Chain, since $\mathbf{v}\,\mathbf{P} = \mathbf{v}$

Convergence Condition: If $\mathbf{P}$ is *irreducible* and ***aperiodic*** (i.e., $\mathbf{P}^n$ converges to some limit as $n \to \infty$) then initial distributions converge to that unique stationary distribution. An nxn transition matrix $\underline{\mathbf{P}\text{ is irreducible and aperiodic } \textit{iff} \text{ all entries in } \mathbf{P}^{n^2-2n+2} \text{ are positive,}}$ $\underline{\text{i.e., } \textit{iff } \mathbf{P} \text{ is } \textit{power-positive}.}$

Note: Finite dimensional transition matrices have at most 1 stationary distribution, and exactly one if they are aperiodic. If $\mathbf{P}^n \to \mathbf{Q}$, the rows of $\mathbf{Q}$ are identical and are the unique stationary distribution.

1.2 Classification of States: *Recurrence* and *Transience.*

Some questions require analyses that decompose the state space of a Markov Chain into *recurrent* and *transient* states. To do this we use the following definitions:

State j is ***recurrent*** if the markov chain returns to that state in a finite number of steps with probability 1. Otherwise, the state is called ***transient***.

[1] Resnik, S. L. 2005. *Adventures in Stochastic Processes.* Ch. 2

Proposition: The state space S of a Markov Chain can be partitioned into a set of **transient** states (T), and closed disjoint classes of **recurrent** states (C_i), so that $S = T \cup C_1 \cup C_2 ...$

Procedure to identify transient and recurrent states.

First, we define a collection of closed sets (*closed* here means starting in a set, the Markov Chain can only transition among events in that set). Second, we identify equivalence classes (of states that communicate with each other) within each closed set. The equivalence classes that are closed contain the recurrent states. The non-closed classes contain transient states.

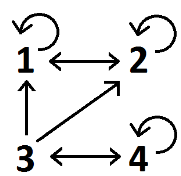
More specifically,

1. Choose a state i and find all states accessible from i, all states accessible from those states, etc. This closed set of states that can (eventually) be reached from state i will be denoted as $cl(i)$. This is the smallest closed set containing i. Find a state k not in $cl(i)$ and repeat for all remaining states.
2. Next, identify the number of equivalence classes within each closed set. Some closed sets are a single equivalence class (i.e., $i \leftrightarrow j$ for all i,j in that equivalence class). Those closed equivalence classes contain recurrent states. Closed sets may also contain more than 1 equivalence class. Non-closed equivalence classes (i.e. states that all communicate with one another, but can also access other states and thus never return) are the transient states.

Example:

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

Directed Graph representation.



The only closed equivalence class is {1,2}. Those are the recurrent states. It is possible to start at 3 or 4, and (via $3 \to 2$) never return. Thus, {3,4} are the transient states.

1.3 Absorption Probabilities and Expected Hitting Times.

Once states are classified as transient and recurrent, the transition matrix $\mathbf{P}$ can be rearranged so that the transient states make up the first rows (columns) of the matrix, and the recurrent states make up the remaining rows (columns). This puts the matrix into the form

$$\mathbf{P} = \begin{bmatrix} Q & R \\ 0 & P_{rec} \end{bmatrix}.$$

The *fundamental matrix*

$$(I - Q)^{-1}$$

can be used to calculate absorption times and probabilities as detailed below.

Absorption Probabilities: It is sometimes desirable to know the probability of a certain state being reached before a set of other states. This is often accomplished by modifying the transition probabilities so that set of target states become *absorbing states* (i.e., once at state j the chain remains there w.p. 1) which are a special kind of recurrent state. We then can frame the question as "What is the probability that the first recurrent state reached is the j^{th} recurrent state, given that the Markov Chain started at the i^{th} transient state?"

The probability of reaching the j^{th} absorbing state, starting from the i^{th} transient state is calculated from the matrix U where u_{ij} is the desired probability, and U is given by

$$U = (I - Q)^{-1} R.$$

Example: Using the example transition matrix from exercise 2 above, what is the probability of hitting 1 before 2, given that we start in state 3?

To answer this, let us first rearrange the given transition matrix $\mathbf{P}$ as described above, and call the result $\mathbf{P}'$ (rows are labeled to clarify the rearrangement).

$$\mathbf{P} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad \mathbf{P}' = \begin{array}{c} 3 \\ 4 \\ 1 \\ 2 \end{array} \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

This implies $Q = \begin{bmatrix} 0 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ and $R = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & 0 \end{bmatrix}$.

Now we can rephrase our question: *what is the probability of hitting 1 (the first recurrent state) first, given that we start in state 3 (the first transient state)?* The answer to our question is therefore given by u_{11}, which is obtained from

$$U = (I - Q)^{-1} R = \begin{bmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{4}{3} & \frac{8}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

These results are perhaps not unexpected: Inspecting the directed graph and transition probabilities given on the previous page, we see that the only way to exit the set of transients is via 3, and that transitioning to 1 is twice as likely as 2.

Expected Absorption Times. Let τ be the number of transitions it takes from starting at transient state i until hitting the first recurrent state. Then we can define a function g and compute the expected cumulative values starting from i until reaching a recurrent state as

$$w_i = \mathbf{E}\left(\sum_{n=0}^{\tau-1} g(X_n) \right).$$

If T is the set of transients, it can be shown that for the vector of values $\mathbf{g} = [g(i), i \in T]'$ that the vector expected values $w = [w_i, i \in T]'$ is given by

$$w = (I - Q)^{-1} \mathbf{g}.$$

In the special case that $g = 1$, this yields the expected time spent in the transient states.

Example: Continuing with the above example, we see that the expected times until leaving the transient state (starting from either state 3 or 4) are given by

$$w = \begin{bmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{4}{3} & \frac{8}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

1.4 Estimating transition probabilities from data

Given data $x = (x_1, \ldots, x_N)$ that are a sequence of N states representing one realization of a trajectory of a discrete time Markov Chain model, the MLE for the transition probability p_{ij} is the fairly straightforward proportion of times transitions from state i go to j. That is,

$$\widehat{p_{ij}} = \frac{N_{ij}(x)}{N_i(x)} \tag{1}$$

where $N_{ij}(x)$ is the number of transitions from i to j in the time series x, and $N_i(x)$ is the number of transitions from i to any state.

For more on fitting discrete time Markov Chains to data, see the `markovchain` package in `R`.