Stat 453: Markov Chains

Matt Cocci

September 15, 2013

1 Introduction

One main application of Markov Chains is in **multi-state transition models**, which describe random movements of a *subject* through various *states*.

2 Markov Chain Definition

Some of assumptions to make things more tractable:

i. Discrete time.

ii. Finite number of states that the subject can be in.

iii. History independence: the distribution for time $n+1$ only considers, at most, the time n and the state at time n. All prior states and additional information are superfluous. This defining characteristic is called the *Markov Property*.

The only condition that's really necessary to have a legitimate Markov Chain is condition (iii), which expresses the Markov Property. Conditions (i) and (ii) are simplifications we could dispose of, if we wish to consider continuous and infinite Markov Chains.

Definition $\{X_t\}$ is a *non-homogeneous Markov Chain* when $\{X_t\}$ is an infinite sequence of random variables $X_0, X_1, \dots$ which satisfy the following properties

i. X_n denotes the *state number* of a subject at time t.

ii. Each X_t is a discrete-type random variable over n possible discrete values.

iii. The *transition probabilities* are history independent when it comes to pre-t states, but the transition probabilities may vary with time, t:

$$\begin{aligned} Q_{ij}^{(t)} &= P(X_{t+1} = j \mid X_t = i,\ X_{t-1} = k_{t-1}, \dots,\ X_0 = k_0) \\ &= P(X_{t+1} = j \mid X_t = i) \end{aligned}$$

So $Q_{ij}^{(t)}$ is the probability of moving from state i to state j at time t.

2.1 Homogeneous vs. Non-Homogeneous Markov Chains

If the transition probabilities, $Q_{ij}^{(t)}$, do not depend on t, then they are denoted by Q_{ij} and we have a *homogeneous Markov Chain*.

For this type of Markov Chain, only state or position, *not* time, factors into the distribution for the next period. This differs from the definition above, in that we explicity allow $Q_{ij}^{(t_1)} \neq Q_{ij}^{(t_2)}$ for $t_1 \neq t_2$.

2.2 Order of a Markov Chain

We can generalize a bit and allow the Markov Chain to depend upon more than one previously observed state. In particular, we define an *order-n Markov Chain* to be a Markov Chain that depends upon the previous n values.

Above, we defined an order-1 Markov chain. If we want to consider higher-order Markov Chains, we'll have to do a bit of work when it comes time to put the probabilities in the transition matrices that we next define. Namely, we will have to convert them to order-1.

3 Transition Matrices

Hopefully, the notation above clearly suggested that it will be convenient to place the many transition probabilities into a *transition matrix*, whose i, j entry is the transition probability for moving from state i to state j. So if there are n states, define

$$\mathbf{Q}^{(t)} = \left(Q_{ij}^{(t)}\right)_{ij} \quad \text{where } \mathbf{Q}^{(t)} \text{ is } n \times n$$

And just a word about the funky notation you see above: it says "the matrix $\mathbf{Q}^{(t)}$ is a matrix populated with the values/probabilities $Q_{ij}^{(t)}$ at the ij entry."

Now in a transition matrix, all the elements in a given row will add up to 1, in which case we call $\mathbf{Q}^{(t)}$ a *stochastic matrix*.[1] If the elements of each column added up to 1 as well, we would call the matrix *doubly-stochastic*.

You should think about the row $Q_{(i\cdot)}^{(t)}$ as the conditional distribution of a Markov Chain, X_{t+1}, given $X_t = i$.

3.1 Longer Term Transition Probabilities and Matrices

Often, we'll want to look further ahead than just the next period, which is all that our current formulation allows. To do so, we'll define the *longer term transition probabilities* as

$$Q_{ij}^{(t,k)} = P(X_{t+k} = j \mid X_t = i)$$

where $Q_{ij}^{(t,k)}$ is the probability of being in state j in k periods given that you are currently at time t and in state i.[2] The above sections dealt with the special case where $k = 1$.

Quite naturally, the *longer term transition matrix* is defined

$$\mathbf{Q^{(t,k)}} = \left(Q_{ij}^{(t,k)}\right)_{ij}$$

Theorem In non-homogeneous Markov Chains, the longer-term probability $Q_{ij}^{(t,k)}$ can be computed as the (i, j)-entry of the matrix

$$\mathbf{Q^{(t,k)}} = \mathbf{Q^{(t)}} \times \mathbf{Q^{(t+1)}} \times \cdots \times \mathbf{Q^{(t+k-1)}}$$

And for a homogeneous Markov Chain, this matrix is just $\mathbf{Q}^{(\cdot,\mathbf{k})} = (\mathbf{Q})^k$, dropping the t in the superscript since the transition probabilities do not change over time.

[1] We will use "transition matrix" and "stochastic matrix interchangeably.

[2] It's important to note that the transition probability $Q_{ij}^{(t,k)}$ does **not** care what happens in between time t and time $t + k$. It's just the probability that you were in state i at time t and may be in state j at time $t + k$. In fact, it's even entirely possible that you were *already* in state j sometime between time t and $t + k$! So if we're considering $Q_{ii}^{(t,k)}$, this is not the probability of staying of in state i from time t to $t + k$. You can drift away and then come back.

4 Chapman-Kolmogorov Equations (Homogeneous Case)

Theorem If we restrict ourselves to the homogeneous case, the *Chapman-Kolmogorov Equations* tell us that

$$Q_{ij}^{(\cdot,\,k+\ell)} = \sum_{s=1}^{n} Q_{is}^{(\cdot,\,k)} Q_{sj}^{(\cdot,\,\ell)}$$

This equation captures the idea of "interposing" some intermediate time and state between now (time t, state i) and the future (time $t+k+\ell$, and state j). The logic goes as follows:

1. Before you get to time $t+k+\ell$ and state j, you have to stop at time $t+k$.

2. At that time $t+k$, you'll be in any one of the states, call it s, where s could equal $1, \cdots, n$ (where n is the number of states.

3. So if you sum over the probability of moving as follows:

$$\text{state } i \text{ at } t \quad \Rightarrow \quad \text{state } s \text{ at } t+k \quad \Rightarrow \quad \text{state } j \text{ at } t+k+\ell$$

   (where s ranges over all possible intermediary states), you'll get the desired probability.

Matrix Representation: In matrix form, this result can be written more compactly as

$$\mathbf{Q}^{(\cdot,\,k+\ell)} = \mathbf{Q}^{(\cdot,\,k)} \times \mathbf{Q}^{(\cdot,\,\ell)}$$

restricting ourselves to the homogeneous case. Things are a little tougher if we want to consider the non-homogeneous case, as we'll have to multiply more matrices together and keep track of subscripts.

5 Marginal Distributions

Suppose we have:

1. A homogeneous Markov Chain, $\{X_t\}$ with a transition matrix, $\mathbf{Q}^{(\cdot)}$
2. A marginal distribution for X_t, denoted by row-matrix $\psi^{(t)}$ The ith entry of $\psi^{(t)}$ is

$$\psi_i^{(t)} = P\{X_t = i\}$$

First, suppose we want to get the marginal distribution of X_{t+1}. This we can do by using the law of total probability:

$$\begin{aligned} P\{X_{t+1} = j\} &= \sum_{i=1}^{n} P\{X_{t+1} = j \mid X_t = i\} \cdot P\{X_t = i\} \\ &= \sum_{i=1}^{n} Q_{ij}^{(\cdot)} \cdot \psi_i^{(t)} \\ \Rightarrow \qquad \psi^{(t+1)} &= \psi^{(t)} \mathbf{Q}^{(\cdot)} \end{aligned}$$

More generally, and still restricting to the homogeneous case, this implies that for any arbitrary m,

$$\psi^{(t+m)} = \psi^{(t)} \mathbf{Q}^{(\cdot,\, m)}$$

Note : All of this math works whether we are *certain* of the initial distribution, X_t, or whether we aren't quite sure.

- If we are *certain* of the initial state i, X_t, or if we want to assume an initial state i, $\psi^{(t)}$ will be a vector that is all zeros *except* for ith position, which will have a 1.
- If we want to think about things more probabilistically or if we are uncertain of the state X_t, then $\psi^{(t)}$ will be a vector that sums to one and has, in each position $i = 1, \ldots, n$, a probability of being in state i at time t.

6 Stationary Distributions

In this section, we dispense with the non-homegeneous case, assuming $\mathbf{Q}^{(t)}$ is homegeneous over time. Therefore, we drop the superscript altogether and simply write $\mathbf{Q}$.

6.1 Definition

A distribution ψ^* is called *stationary* or *invariant* if $\psi^* = \psi^* \mathbf{Q}$, which implies that $\psi^* = \psi^* (\mathbf{Q})^m$ for all m too.[3] Mathematically, a stationary distribution is a fixed point if we think of $\mathbf{Q}$ as a map:

$$\begin{aligned} P : \mathbb{R}^n &\to \mathbb{R}^n \\ \psi &\to \psi P \end{aligned}$$

At least one such distribution exists for each stochastic matrix, $\mathbf{Q}$.[4]

Importantly, if X_0 is a stationary distribution, then X_t will have this same distribution for all t. As a result stationary distributions have a natural interpretation of *stochastic steady states.*

6.2 Solving for Stationary Distributions

We saw above that a stationary distribution, ψ^*, must solve

$$\psi = \psi \mathbf{Q} \quad \Leftrightarrow \quad \psi \left(I_n - \mathbf{Q} \right) = 0 \tag{1}$$

But note that this does not require ψ^* to be a probability distribution since the zero vector happens to solve Equation 1. So we want to impose the additional constraint

$$\begin{aligned} \sum_{i=1}^{n} \psi_i = 1 \quad &\Leftrightarrow \quad \psi b = 1 \qquad \text{where } b_i \in \mathbb{R}^n \text{ and } b_i = 1 \text{ for } i = 1, \ldots, n. \\ &\Leftrightarrow \quad \psi B = b \qquad \text{where } B \text{ is an } n \times n \text{ matrix of 1s} \end{aligned} \tag{2}$$

By adding together the two conditions, Equations 1 and 2, we see that ψ must satisfy

$$\begin{aligned} & \psi \left(I_n - \mathbf{Q} + B \right) = b \\ \Leftrightarrow \qquad & \left(I_n - \mathbf{Q} + B \right)^T \psi^T = b^T \end{aligned}$$

In this way, we can solve for the stationary distribution by inverting the matrix to get:

$$\psi^* = \left[\left(I_n - \mathbf{Q} + B \right)^T \right]^{-1} b^T \tag{3}$$

[3] Take note that a stationary *distribution* differs from a stationary *process*. The former is the limiting distribution of a Markov Chain, or the marginal distribution of a stationary process. The latter is a stochastic process whose joint probability distribution does not change when shifted through time or space. Finally, a "stationary process" is *not* the same as a "process with a stationary distribution."

[4] Use Brouwer's fixed point theorem.

6.3 Uniform Ergodicity

Definition A stochastic matrix, $\mathbf{Q}$, is called uniformly ergodic if there is a positive integer m such that all elements of $(\mathbf{Q})^m$ are *strictly* positive.

Uniqueness of the Stationary Distribution Note that there may in fact be many stationary distributions for the stochastic matrix, $\mathbf{Q}$ (as in the case of the identity matrix). But one sufficient condition for uniqueness is *uniform ergodicity.*

Convergence to Steady-State If uniform ergodicity holds, we also get the result that for any non-negative row vector, ψ, summing to one (so a proper distribution)

$$\psi \mathbf{Q}^t \to \psi^* \quad \text{as } t \to \infty$$

where ψ^* is the unique stationary distribution. So regardless of the distribution of X_0, the distribution of X_t converges to ψ^*.

Time in States To get another import interpretation and result, assume $\{X_t\}$ is a Markov Chain with stochastic matrix $\mathbf{Q}$. Also assume that $\mathbf{Q}$ is uniformly ergodic with stationary distribution ψ^*. Then

$$\lim_{s \to \infty} \frac{1}{s} \sum_{t=1}^{s} 1\{X_t = j\} \to \psi_j^* \qquad \forall j \in \{1, \dots, n\} \tag{4}$$

This tells use that the fraction of time the Markov Chain $\{X_t\}$ spends in state j converges to ψ_j^* as time goes to infinity. Therefore, if we consider many Markov chains with the same stochastic matrix, $\mathbf{Q}$, the long-run cross-sectional averages for a population will equal time-series averages for individual chains.