Stat 453: Markov Chains

Matt Cocci

August 4, 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 the assumptions made 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 M is a **non-homogeneous Markov Chain** when M is an infinite sequence of random variables $M_0, M_1, \dots$ which satisfy the following properties

i. M_n denotes the *state number* of a subject at time n.

ii. Each M_n is a discrete-type random variable over r values.

iii. The **transition probabilities** are history independent:

$$\begin{aligned} Q_n^{(i,j)} &= P(M_{n+1} = j \mid M_n = i, M_{n-1} = k_{n-1}, \dots, M_0 = k_0) \\ &= P(M_{n+1} = j \mid M_n = i) \end{aligned}$$

2.1 Homogeneous vs. Non-Homogeneous Markov Chains

If the transition probabilities, $Q_n^{(i,j)}$, do not depend on n, then they are denoted by $Q^{(i,j)}$ and we have a **homogeneous Markov Chain**. This type of Markov Chain considers only state or position, not time, when forming the distribution for the next period. This differes from the definition above, in that we explicity allow $Q_n^{(i,j)} \neq Q_m^{(i,j)}$ for $n \neq m$.

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. Now, 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 oder-1.

3 Transition Matrices

It's 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:

$$\mathbf{Q_n} = \left(Q_n^{(i,j)}\right)_{i\times j}$$

In a transition matrix, all the elements in a given row will add up to 1, in which case we call $\mathbf{Q_n}$ a **stochastic matrix**. If the elements of each column added up to 1 as well, we would call the matrix **doubly-stochastic**, but this is not generally something that we'll deal with.

3.1 Longer Term Transition 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

$${}_kQ_n^{(i,j)} = P(M_{n+k} = j \mid M_n = i)$$

where the **longer term transition matrix** is defined

$$\mathbf{{}_kQ_n} = \left({}_kQ_n^{(i,j)}\right)_{i\times j}$$

Theorem In non-homogeneous Markov Cahins, the longer-term probability ${}_kQ_n^{(i,j)}$ can be computed as the (i, j)-entry of the matrix

$$\mathbf{{}_kQ_n} = \mathbf{Q_n} \times \mathbf{Q_{n+1}} \times \cdots \times \mathbf{Q_{n+k-1}}$$

And for a homogeneous Markov Chain, this matrix is just $\mathbf{{}_kQ} = \mathbf{Q^k}$.

Interpretation It's important to note that the transition probability ${}_kQ_n^{(i,j)}$ does not care what happens in between time n and time $n+k$. It's just the probability that you were in state i at time n and may be in state j at time $n+k$. It's entirely possibly that you were *already* in state j between time n and $n+k$. In the special case of ${}_kQ_n^{(i,i)}$, this is *not* the probability that you remain in state i the whole time. Rather it includes the probability and the probability that you drift away from i and return at time $n+k$.

3.2 Chapman-Kolmogorov Equations (Homogeneous Case)

Theorem The **Chapman-Kolmogorov Equations** tell us that

$$_{k+\ell}Q^{(i,j)} = \sum_{s=1}^{n} {}_kQ^{(i,s)}{}_\ell Q^{(s,j)}$$

if we restrict ourselves to the homogeneous case. In matrix form, this result can be written as

$$_{\mathbf{k}+\ell}\mathbf{Q} = {}_{\mathbf{k}}\mathbf{Q} \times {}_\ell\mathbf{Q}.$$

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

3.3 Finding Probabilities

Let's define a set of values that describes the initial distribution of for the possible states at time 0:

$$\alpha_i^0 = P(M_0 = i)$$
$$\Sigma\, \alpha_i^0 = 1$$

Then we can find a posterior distribution for time k—which will give us another set of values—by computing

$$P(M_k = j) = \sum_{i=1}^{r} P(M_k = j \mid M_0 = i)P(M_0 = i)$$

$$\alpha_j^k = \sum_{i=1}^{r} {}_kQ^{(i,j)}\alpha_i$$

In vector and matrix notation, we get (for one timestep) the following completely equivalent statements:

$$_{\mathbf{1}}\mathbf{Q}^T \cdot \vec{\alpha}^0 = \vec{\alpha}^1$$
$$\left(\vec{\alpha}^0\right)^T \cdot {}_{\mathbf{1}}\mathbf{Q} = \left(\vec{\alpha}^1\right)^T$$

4 Classifications of States

In order to decide whether there is a limiting matrix for a given Markov Chain, we introduce some terminology to allow us to answer that question.

4.1 Definitions

Here are a few important concepts:

1. State j is **accessible** from state i if ${}_nQ^{(i,j)} > -$ for some n.

2. If i and j are accessible to *each other*, we say that i and j **communicate**.

3. The notion of "communicate" actually forms an equivalence class, so we say that two states which communicate are in the same **class**.

4. A Markov Chain with only one class is said to be **irreducible**. In that case, all possible, non-trivial states communicate with each other.

5. If we let f_i represent the probability of starting in i and returning to i at some later point, we say that
   a) i is a **recurrent state** if $f_i = 1$. In a finite, irreducible chain, all states are recurrent.
   b) i is a **transient state** if $f_i < 1$. In a finite state Markov Chain, it's impossible for all states to be transient.

6. If we let T be the time it takes to return to some recurrent state "i", then
   a) If ET is finite, we call our Markov Chain **positive recurrent**. In a finite state Markov Chain, then all recurrent states are positive.
   b) If ET is infinite, we call our Markov Chain **null recurrent**. We won't really deal with this, since we'll mostly consider finite state Markov Chains, which are all positive recurrent, as just stated.

7. A state i is said to be **periodic** with period d if ${}_kQ^{(i,j)} = 0$ whenever k is not divisible by d—where d is the largest integer with this property. A period 1 state is called **aperiodic**, or **non-periodic**. Periodicity is a class property.

8. An **ergodic state** is non-periodic, positive recurrent. Note that this is a condition for stationarity.

4.2 Big Result

Theorem For a *irreducible*, *ergodic* Markov Chain, there exists a limiting distribution

$$\lim_{k\to\infty} {}_kQ^{(i,j)} = \pi_j$$

where π_j is a unique solution to the following system of equations, given in both traditional notation and in matrix notation

$$\begin{cases} \pi_j = \sum_{i=1}^{r} \pi_i Q^{(i,j)} & \mathbf{\Pi}^T = \mathbf{\Pi}^T \cdot \mathbf{Q} \\ 1 = \sum_{i=1}^{r} \pi_i & \Sigma\, \pi_i = 1 \end{cases}$$

Note that the limit is independent of the starting point i. In a sense, it represents the long-run probabilities of being in either state.

Example Here's an example for a 2×2 matrix:

$$\begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix}$$

Solving this with the following equations

$$\pi_0 = \alpha\pi_0 + \beta\pi_1$$

$$1 = \pi_0 + \pi_0$$

will give the following result

$$\pi_0 = \frac{\beta}{1-\alpha+\beta}, \quad \pi_1 = \frac{1-\alpha}{1-\alpha+\beta}$$

4.3 Gambler's Ruin

Here, the probability of moving right on any given play is p and the probability of moving left is q. We can only move left or right a finite amount of steps. Namely, we stop at N on the right and 0 on the left.

We'll want to know P_i, which is the probability that our wealth reaches N before it reaches 0, given that we started at position i—an integer in the inteval $(0, N)$. To do so, we set up the recursive equation

$$P_i = p \cdot P_{i+1} + q \cdot P_{i-1}$$

which says that the probability of winning starting with i can be broken up into two separate cases:

1. You win with probability p, in which case your chances of winning are P_{i+1}.
2. You lose with probability q, in which case, your chances of winning are P_{i-1}.

Solving, the following expression

$$p \cdot P_i + q \cdot P_i = P_i = p \cdot P_{i+1} + q \cdot P_{i-1}$$

$$\Rightarrow P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1})$$

we can get a recursive definition by subbing in for i:

$$P_2 - P_1 = \frac{q}{p}(P_1 - 0), \qquad P_3 - P_2 = \frac{q}{p}\left(\frac{q}{p}P_1\right), \ldots$$

$$\Rightarrow P_{i+1} - P_i = \left(\frac{q}{p}\right)^i P_1, \quad 1 \le i \le N-1$$

Then, we can add up all the terms in this telescoping sum to get get lots of stuff to cancel and become:

$$P_i - P_1 = P_1 \left[\sum_{i=1}^{N-1} \left(\frac{q}{p} \right)^i \right], \qquad or \qquad P_i = P_1 \left[\sum_{i=1}^{N-1} \left(\frac{q}{p} \right)^i \right]$$

which can be rewritten, using the fact that $P_N = 1$ along with some other simplifications, to give

$$P_i = \begin{cases} \frac{i}{N} & p = q = \frac{1}{2} \\ \frac{1-(q/p)^i}{1-(q/p)^N} & p \neq q \end{cases}$$

Intuition When $N \to \infty$ like in the case of a casino, for all practical purposes, you can't win. You can only win if $p > q$.

4.4 Checking the Efficacy of a Drug

Suppose there are two drugs, each with cure rate P_1 and P_2 respectively. But these are unknown population values, and we wish to establish a test to see if $P_1 > P_2$ or vice versa. So we define, where j indexes the trial or observation pairs

$$X_j = \begin{cases} 1 & \textit{cured by Drug } 1 \\ 0 & \textit{not cured} \end{cases} \qquad Y_j = \begin{cases} 1 & \textit{cured by Drug } 2 \\ 0 & \textit{not cured} \end{cases}$$

Then, we define the discrepancy between the two total cures

$$M = (X_1 + X_2 + \cdots + X_n) - (Y_1 + Y_2 + \cdots + Y_n)$$

We will then decide if a drug is better be testing

$$M > T_M \Rightarrow P_1 > P_2$$

$$M < T_{-M} \Rightarrow P_1 < P_2$$

where T_M and T_{-M} are some threshold values.

To determine whether the test is good—i.e. that it doesn't give false signals—let's compute p, the probability that the sum M moves up, given that it moves *either* up or down, while q is the analagous value for down.

$$p = \frac{P_1(1-P_2)}{P_1(1-P_2) + (1-P_1)P_2}, \qquad q = 1 - p$$

Now we can easily find the probability that the test will assert that $P_2 > P_1$—it's a problem completely analagous to the Gambler's ruin problem. So if we take $i = M$ and $N = 2M$ in the Gambler's Ruin problem, we get that

$$P(Test\ says\ P_2 > P_1) = 1 - \frac{1-(q/p)^M}{1-(q/p)^{2M}}$$

Thus we can choose *hypothetical* values for P_1 and P_2, find the values p and q, then compute the probability that the test will render a certain result, given our choice of M. This allows us to calibrate the test given our desired values for Type I and Type II errors.

5 Cash Flows and Actuarial Present Values

The big result is the following equation, which gives the actuarial present value of cash flows at transitions. Let ${}_{\ell+1}C^{(i,j)}$ denote the cash flow at time $\ell+1$ if the subject is in state i at time ℓ and state j at time $\ell+1$. Next, let ${}_k v_n$ denote the value at time n of one unit to be paid k periods in the future at time $n+k$.

$$APV_{s@n}(\mathbf{C^{(i,j)}} = \sum_{k=0}^{\infty} \left[{}_k Q_n^{(i,j)} Q_{n+k}^{(i,j)}\right] \left[{}_{n+k+1}C^{(i,j)}\right] \left[{}_{k+1}v_n\right]$$

6 Miscellaneous

Definition We can **exponentially smooth** probabilities when we work up our transition probabilities matrix:

$$Q_n = \alpha T_n + (1-\alpha) Q_{n-1}.$$