

Markov Chains, part I

December 8, 2010

1 Introduction

A Markov Chain is a sequence of random variables X_0, X_1, \dots , where each $X_i \in \mathcal{S}$, such that

$$\mathbb{P}(X_{i+1} = s_{i+1} \mid X_i = s_i, X_{i-1} = s_{i-1}, \dots, X_0 = s_0) = \mathbb{P}(X_{i+1} = s_{i+1} \mid X_i = s_i);$$

that is, the value of the next random variable is dependent at most on the value of the previous random variable.

The set \mathcal{S} here is what we call the “state space”, and it can be either continuous or discrete (or a mix); however, in our discussions we will take \mathcal{S} to be discrete, and in fact we will always take

$$\mathcal{S} = \{1, 2, \dots, N\}.$$

Since X_{t+1} only depends on X_t , it makes sense to define “transition probabilities”

$$P_{i,j} := \mathbb{P}(X_{t+1} = j \mid X_t = i),$$

which completely determine the dynamics of the Markov chain... well, almost: we need to either be given X_0 , or we to choose its value according to some distribution on the state space. In the theory of Hidden Markov Models, one has a set of probabilities π_1, \dots, π_N , $\pi_1 + \dots + \pi_N = 1$, such that $\mathbb{P}(X_0 = i) = \pi_i$; however, in some other applications, such as in the Gambler’s Ruin Problem discussed in another note, we start with the value for X_0 .

Ok, so how could we generate a sequence X_0, X_1, \dots , given X_0 and given the $P_{i,j}$ ’s? Well, suppose $X_0 = i$. Then, we choose X_1 at random from $\{1, 2, \dots, N\}$, where $\mathbb{P}(X_1 = j) = P_{i,j}$. Next, we select X_2 at random according to the distribution $\mathbb{P}(X_2 = k) = P_{j,k}$. We then continue the process.

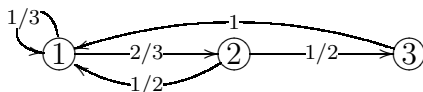
1.1 Graphical representation

Sometimes, a more convenient way to represent a Markov chain is to use a *transition diagram*, which is a graph on N vertices that represent the states. The edges are directed, and each corresponds to a transition probability $P_{i,j}$; however, not all the N^2 edges are necessarily in the graph – when an edge is missing, it means that the corresponding $P_{i,j}$ has value 0.

Here is an example: suppose that $N = 3$, and suppose

$$P_{1,1} = 1/3, P_{1,2} = 2/3, P_{2,1} = 1/2, P_{2,3} = 1/2, P_{3,1} = 1.$$

Then, the corresponding transition diagram looks like this



1.2 Matrix representation, and population distributions

It is also convenient to collect together the $P_{i,j}$'s into an $N \times N$ matrix; and, I will do this here a little bit backwards from how you might see it presented in other books, for reasons that will become clear later on: form the matrix P whose (j,i) entry is $P_{i,j}$; so, the i th column of the matrix represents all the transition probabilities *out of* node i , while the j th row represents all transition probabilities *into* node j . For example, the matrix corresponding to the example in the previous subsection is

$$P = \begin{bmatrix} P_{1,1} & P_{2,1} & P_{3,1} \\ P_{1,2} & P_{2,2} & P_{3,2} \\ P_{1,3} & P_{2,3} & P_{3,3} \end{bmatrix} = \begin{bmatrix} 1/3 & 1/2 & 1 \\ 2/3 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix}.$$

Notice that the sum of entries down a column is 1.

Now we will reinterpret this matrix in terms of population distributions: suppose that the states $1, \dots, N$ represent populations – say state i represents “country i ”. Associated to each of these populations, we let $p_i(t)$ denote the fraction of some total population residing in country i at time t . In transitioning from a population makeup at time t to time $t+1$, some fraction of the population in each state will be sent to the other states; the fraction

of the population at state i at time t sent to state j at time $t + 1$ will be $P_{i,j}$. So, the fraction of the total population in state j at time $t + 1$ will be

$$\sum_{i=1}^N p_i(t) P_{i,j}.$$

It turns out that this can be very compactly represented by a matrix product: letting P be the matrix as described above, we have that

$$P \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_N(t) \end{bmatrix} = \begin{bmatrix} p_1(t+1) \\ p_2(t+1) \\ \vdots \\ p_N(t+1) \end{bmatrix}.$$

This brings us now to a few natural questions, along with their answers:

- Is there a choice for $p_i(t)$ such that the population is “stable under the transformation P ”? That is to say, is there choice for $p_i(t)$ that is an eigenvector for eigenvalue $\lambda = 1$? **Answer:** Yes! In fact, we will give a nice construction.
- Is it always the case that iterating the above matrix equation, producing $\vec{p}(t), \vec{p}(t + 1), \vec{p}(t + 2), \dots$, we reach equilibrium regardless of the choice of starting distribution $\vec{p}(t)$? **Answer:** No!
- Is there more than one equilibrium distribution? **Answer:** There could be, as we shall see.
- How quickly does the process converge to equilibrium, assuming it converges at all? **Answer:** Turns out that it depends on the second largest (in magnitude) eigenvalue.
- Is there a simple general condition guaranteeing that the equilibrium distribution is unique, and that we always converge to it no matter what our starting distribution $\vec{p}(0)$ happens to be? **Answer:** Yes! And it goes by the name “Perron-Frobenius Theorem”.

2 Convergence to equilibrium

The first thing we will see is that the matrix P has $\lambda = 1$ as an eigenvalue: since the transpose of P , denoted P^t always has row sum equal to 1, it follows that

$$P^t \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

So, 1 is an eigenvalue for P . Since P and P^t have the same eigenvalues, it follows that $\lambda = 1$ must also be an eigenvalue for P .

Next, we want to pick out a vector $\vec{p}(0)$ that is an eigenvector for $\lambda = 1$: basically, we begin by selecting $\vec{q}(0)$ to be $(1/N, 1/N, \dots, 1/N)$, and then compute $\vec{q}(1), \vec{q}(2), \dots$; that is, we compute the sequence

$$\begin{bmatrix} 1/N \\ 1/N \\ 1/N \\ \vdots \\ 1/N \end{bmatrix}, P \begin{bmatrix} 1/N \\ 1/N \\ 1/N \\ \vdots \\ 1/N \end{bmatrix}, P^2 \begin{bmatrix} 1/N \\ 1/N \\ 1/N \\ \vdots \\ 1/N \end{bmatrix}, \dots$$

Next, we compute the average of the first several of these; that is, we let

$$\vec{r}_0 := \vec{q}(0), \vec{r}_1 := \frac{\vec{q}(0) + \vec{q}(1)}{2}, \dots, \vec{r}_k := \frac{\vec{q}(0) + \vec{q}(1) + \dots + \vec{q}(k)}{k+1}, \dots$$

And now it turns out that the terms in this sequence behave more and more like eigenvectors the further out one goes. To see this, notice that

$$P\vec{r}_k = P \frac{\vec{q}(0) + \vec{q}(1) + \dots + \vec{q}(k)}{k+1} = \frac{\vec{q}(1) + \vec{q}(2) + \dots + \vec{q}(k+1)}{k+1},$$

which is pretty close to our expression for \vec{r}_k ; in fact, the difference between \vec{r}_k and $P\vec{r}_k$ is just the vector

$$\frac{\vec{q}(k+1) - \vec{q}(0)}{k+1}. \tag{1}$$

Now, how big are the coordinates of this vector? Well, we know that since P is a transition matrix, if the sum of entries is \vec{q} is 1, then the same will be

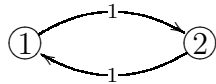
true of $P\vec{q}$, which means that coordinates of $\vec{q}(k+1)$ and $\vec{q}(0)$ must all be ≤ 1 . It follows, then, that this vector (1) has all coordinates bounded from above by $1/(k+1)$ in absolute value.

What this means is that if the sequence $\vec{r}_0, \vec{r}_1, \dots$ converges to some vector \vec{r} , then \vec{r} must be an eigenvector.

Well, it's not obvious that this sequence of vectors should actually converge; but that's ok, because all we *really* need is that there is a subsequence of these vectors that converges. And, it is a standard fact of Analysis that every bounded, infinite sequence of vectors has a convergent subsequence – in our case, converging to an eigenvector of P for $\lambda = 1$.

3 The Perron-Frobenius Theorem

Although we now know that there always *exists* an equilibrium distribution, it doesn't follow that every initial population $\vec{p}(0)$ necessarily converges to it when we iteratively multiply it by P on the left. A good example here is represented by the following transition diagram:



The transition matrix corresponding to this graph is

$$P := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Starting with initial distribution $(1, 0)$, repeated iteration leads us to the 2-cycle

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \dots$$

Furthermore, the equilibrium distribution, if it even exists, need not be unique. One example of this is just a Markov Chain having two states, where each state transitions only to itself, and with probability 1. In this case, the transition matrix is just the identity matrix, which of course has *every* non-zero vector as an eigenvector.

The following theorem will imply general conditions guaranteeing convergence to a unique equilibrium:

Theorem 1 (Perron-Frobenius) *Suppose that A is an $N \times N$ matrix with non-negative entries such that some positive integer power of A (i.e. A^k , where $k \in \mathbb{Z}_+$) has all positive entries. Then, we must have that*

(i) *A has a positive real number eigenvalue λ that is larger in magnitude than all other eigenvalues of A .*

(ii) *That eigenvalue λ is simple, meaning that the characteristic polynomial of A has λ as a root with multiplicity 1.*

(iii) *There is an eigenvector of A corresponding to eigenvalue λ whose entries are all strictly positive; furthermore, in light of the fact that λ has multiplicity 1, all other eigenvectors corresponding to this eigenvalue λ must be a scalar multiple of it.*

Before we give the proof of this very important theorem, we give a corollary:

Corollary 1 *Suppose that P is a transition matrix such that P^k has all positive entries for some $k \geq 1$. Then, $\lambda = 1$ is the largest eigenvalue of P in absolute value, and it is a simple root (multiplicity 1) of the characteristic polynomial of A . Furthermore, given any initial population vector $\vec{p}(0)$, whose sum of entries is 1, we have that $\vec{p}(0), \vec{p}(1), \vec{p}(2), \dots$ converges to the unique equilibrium distribution \vec{p} .*

3.1 Proof of the corollary

First, we note that if α is an eigenvalue of P , then $\alpha \leq 1$; for, if \vec{v} is its corresponding eigenvector, then from the fact that P^n is also a transition matrix (sum down the columns is 1, entries are all between 0 and 1), we have that $P^n \vec{v} = \alpha^n \vec{v}$ is a bounded vector for all $n = 1, 2, \dots$ – so, α *couldn't* satisfy $|\alpha| > 1$, since that would lead to unbounded growth.

Of course, $\lambda = 1$ is an eigenvalue of P , and as long as P^k is a positive matrix for some k , Perron-Frobenius tells us that all other eigenvalues must have magnitude strictly smaller than 1. Now we find the Jordan Canonical Form decomposition of the matrix P , which has the form

$$P = M^{-1} \begin{bmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_k \end{bmatrix} M,$$

where the B_i represent square sub-matrices of the form

$$B_i = [\alpha] \text{ or } \begin{bmatrix} \alpha & 1 \\ 0 & \alpha \end{bmatrix} \text{ or } \begin{bmatrix} \alpha & 1 & 0 \\ 0 & \alpha & 1 \\ 0 & 0 & \alpha \end{bmatrix} \text{ or } \dots,$$

where α is one of the eigenvalues of P . We reserve B_1 to be the 1×1 block corresponding to eigenvalue 1; from the simplicity of $\lambda = 1$, as given to us by Perron-Frobenius, we have that this is the only block corresponding to 1 and that all the other eigenvalues are of strictly smaller magnitude. From this, and a little work, one can show that B_i^n converges to a 0 block matrix when $i = 2, 3, \dots, k$; and so,

$$\lim_{n \rightarrow \infty} P^n = M^{-1} \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} M.$$

When you work this matrix out, what you find is that each row is a scalar multiple of the first row. Since the sum down each column must equal 1, this means that every entry in that first row must be equal; in other words, this limit matrix has the form

$$\begin{bmatrix} c_1 & c_1 & c_1 & \dots & c_1 \\ c_2 & c_2 & c_3 & \dots & c_2 \\ c_3 & c_3 & c_3 & \dots & c_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_N & c_N & c_N & \dots & c_N \end{bmatrix}.$$

From this it follows that for any vectors \vec{v} whose sum of entries is 1,

$$\lim_{n \rightarrow \infty} P^n \vec{v} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix}.$$

It follows that we always settle down to the same distribution – the equilibrium distribution, regardless of what \vec{v} happens to be, so long as its entries sum to 1. This basically completes the proof of the corollary.

3.2 Proof of Perron-Frobenius

We will use the notation $B > 0$ to indicate that all the entries of a matrix B are positive, and will use the notation $B \geq 0$ to indicate that all the entries are at least 0.

We first note that if $A^k > 0$ then since $A \geq 0$, we must have $A^{k+1}, A_{k+2}, \dots > 0$ as well.

Now suppose, for the time being, we could prove the Perron-Frobenius theorem for when B is a positive matrix, instead of the more general case for when B^k is a positive matrix; in particular, suppose we can show that the largest eigenvalue of B is always a positive real number. We then want to conclude that Perron-Frobenius holds in the general case (where a matrix A need not be a positive matrix, but that $A^k > 0$ for some $k \geq 1$): well, first note that if α is that largest eigenvalue of A , then α^{k+j} must be the largest eigenvalue of A^{k+j} , $j = 0, 1, 2, \dots$, in magnitude. By Perron-Frobenius in the case of a positive matrix B , since $B = A^{k+j}$ is a positive matrix we conclude that α^{k+j} is a positive real number for each $j = 0, 1, 2, \dots$. The only way this could happen, even just for both $j = 0$ and $j = 1$, is if α itself is a positive real number.

So, to prove Perron-Frobenius in general, then, it suffices to prove it just for the case where A is a positive matrix, and that is the case upon which we shall now focus our attention.

Now we will show that A even *has* a positive eigenvalue: consider the collection of all pairs (α, \vec{v}) , where $\alpha > 0$ and \vec{v} is a vector with all positive entries that is not identically 0, such that $A\vec{v} \geq \alpha\vec{v}$. That is to say, if we let $\vec{v} = (v_1, \dots, v_N)$, then

$$A \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = \alpha \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}. \quad (2)$$

Among all such pairs, let (λ, \vec{w}) be one for which $\alpha = \lambda > 0$ is maximal. Now, if the vector of ε_i 's corresponding for $\alpha = \lambda$ is $\vec{0}$, then \vec{w} is an eigenvector corresponding to $\lambda > 0$; on the other hand, if at least one of those ε_i 's is non-zero (and hence positive), then letting $\vec{y} = (y_1, \dots, y_N)$ be the vector $A\vec{w}$,

and applying A to both sides of (2) in the case $\alpha = \lambda$ and $\vec{v} = \vec{w}$, we get

$$A\vec{y} = \lambda\vec{y} + A \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}.$$

Now, this last vector $A\vec{\varepsilon}$ clearly has all positive entries, since at least one ε_i is non-zero and since A itself is a positive matrix. What this means is that we can we could have increased λ slightly to λ' , and found a vector \vec{z} such that $A\vec{z} \geq \lambda'\vec{z}$. In other words, $\lambda > 0$ must not have been maximal, which is a contradiction. So, we are forced to conclude that all the ε_i 's are 0 and that \vec{w} is an eigenvector.

Next, we show that all other eigenvalues of A are at most as large as λ : suppose that $\lambda' \neq \lambda$ is any other eigenvalue of A , and suppose that $\vec{x} = (x_1, \dots, x_N)$ is its associated eigenvector. It fairly obvious from the triangle inequality that since A has all positive entries, each entry of $A|\vec{x}| = A(|x_1|, \dots, |x_N|)$ is at least as large (in absolute value) as the corresponding entry of $A\vec{x} = \lambda'\vec{x}$; and so, if we let $\lambda'' > 0$ denote the largest positive real number such that $A|\vec{x}| \geq \lambda''|\vec{x}|$, we must have that

$$|\lambda'| \leq \lambda'' \leq \lambda.$$

We now show that all other eigenvalues are strictly smaller in magnitude than λ : basically, if we had that $|\lambda'| = \lambda'' = \lambda$, then if $A_{i,j}$ denotes the i, j entry of A , we will have that

$$\left| \sum_{j=1}^N A_{i,j}x_j \right| = |\lambda| \cdot |x_i| = \sum_{j=1}^N A_{i,j}|x_j|.$$

And the only way that that could happen is if all the x_j 's “point in the same direction”, by which I mean that there exists a complex number $\omega \neq 0$ such that

$$\text{For all } j = 1, 2, \dots, N, \quad x_j = \omega u_j, \quad u_j \in \mathbb{R}_+.$$

Since \vec{u} is a positive vector, in order for

$$A\vec{u} = A(\omega^{-1}\vec{x}) = \omega^{-1}\lambda'\vec{x} = \lambda'\vec{u}$$

we must have $\lambda' > 0$, and then λ' either equals λ or is strictly smaller than it.

Lastly we show that this largest eigenvalue λ has multiplicity 1: there are two ways that λ can fail to have multiplicity 1. Either there are two or more linearly independent eigenvectors corresponding to λ ; or else there is a single Jordan Block in the Jordan Canonical Form decomposition of A that has two or more rows and columns.

Suppose first that A has two linearly independent eigenvectors \vec{v} and \vec{w} corresponding to λ , one of which – say it is \vec{v} – is a vector with all positive entries (which we know exists from the arguments above). Then, it is clear that any linear combination of \vec{v} and \vec{w} is also an eigenvector; and, by choosing coefficients δ_1 and δ_2 appropriately, it is easy to see that we can create a non-zero, non-negative vector $\delta_1\vec{v} + \delta_2\vec{w}$ with the property that one of its coordinates is 0. But now if we apply A (which has all positive entries) to this vector, we will produce a vector with all positive entries; hence, that $\delta_1\vec{v} + \delta_2\vec{w}$ could not have been an eigenvector, because that 0 entry should have remained 0 upon multiplying through by A .

Finally, suppose that there is only one Jordan Block associated with λ , and that it is 2×2 , or 3×3 , etc. Consider the matrix $B = \lambda^{-1}A$. The eigenvalues of this matrix are basically just λ^{-1} times the eigenvalues of A ; and, that corresponding Jordan Block we aim to eliminate will correspond to an eigenvalue $1 = \lambda^{-1}\lambda$ for B . The analogue of that Jordan Block for B will thus be

$$\begin{bmatrix} 1 & 1/\lambda \\ 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & 1/\lambda & 0 \\ 0 & 1 & 1/\lambda \\ 0 & 0 & 1 \end{bmatrix} \text{ or } \dots,$$

Taking high powers of this block will lead to unbounded growth; for example

$$\begin{bmatrix} 1 & 1/\lambda \\ 0 & 1 \end{bmatrix}^n = \begin{bmatrix} 1 & n/\lambda \\ 0 & 1 \end{bmatrix}. \quad (3)$$

All the other Jordan Blocks that arise for when B is put into Jordan Canonical Form will correspond to eigenvalues of magnitude smaller than 1; and so, when we compute high powers of them, they will converge to 0. It follows,

then, that

$$\lim_{n \rightarrow \infty} B^n = M^{-1} \begin{bmatrix} C^n & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} M,$$

where C^n has the form (3), or corresponding 3×3 , 4×4 , ... analogue.

Now, this middle $N \times N$ matrix sandwiched between M^{-1} and M can be interpreted geometrically as just the transformation B^n with respect to a basis determined by the matrix M ; and we see that using this new basis we must have a vector that grows in length the more times we multiply it by B . Clearly this means that as $n \rightarrow \infty$ the largest entry of the matrix B^n has unbounded growth. We now show that this is a contradiction: since the largest eigenvalue (in magnitude) of B is 1, and since B is a positive matrix, we know from the first part of Perron-Frobenius that there exists an eigenvector \vec{q} with all positive entries such that $B\vec{q} = \vec{q}$. We must also have, then, that $B^n\vec{q} = \vec{q}$ for all $n = 1, 2, 3, \dots$ as well. This, however, is impossible, since if the i th row of B^n is unbounded for infinitely many n we will have that the i th coordinate of $B^n\vec{q}$ should likewise be unbounded (multiply $B^n\vec{q}$ and check this for yourself). This completes the proof of Perron-Frobenius, since we have now basically eliminated those 2×2 or 3×3 , etc. blocks.