Markov Chains and Stationary Distributions

David Mandel

February 4, 2016

A collection of facts to show that any initial distribution will converge to a stationary distribution for irreducible, aperiodic, homogeneous Markov chains with a full set of linearly independent eigenvectors.

Definition Let A be an $n \times n$ square matrix. A is *irreducible* if for every pair of indices $i, j = 1, \ldots, n$ there exists an $m \in \mathbb{N}$ such that $(A^m)_{ij} \neq 0$.

Remark In the context of Markov chains, a Markov chain is said to be irreducible if the associated transition matrix is irreducible. Also in this context, a Markov chain is called irreducible if all its states *communicate*, which means exactly the definition for irreducible.

Definition Let A be a non-negative $n \times n$ square matrix. The *period of index* i, $i = 1, \ldots, n$, is the GCD of all $m \in \mathbb{N}$ such that $(A^m)_{ii} > 0$.

Remark If A is irreducible the period of each index is the same; hence we may speak of the *period of* A in such a case.

Remark If the period of A is 1, then A is called *aperiodic*.

Theorem 0.1 *(Perron-Frobenius) Let A be an irreducible, non-negative $n \times n$ matrix with period α and spectral radius $\rho(A) = r$. Then*

1. *The number r is a unique eigenvalue of A (it is a simple root of the characteristic equation of A).*
2. *A has a left eigenvector z with associated eigenvalue r, and z has all positive entries.*
3. *A has exactly α complex eigenvalues with modulus r and each is a simple root of the characteristic polynomial of A.*

Proposition 0.2 *A row-stochastic square matrix has a largest eigenvalue of one.*

Proof Let A be an $n \times n$ row-stochastic matrix; i.e., $\sum_{j=1}^{n} a_{ij} = 1$ for all $i = 1, \ldots, n$. Since

$$A\mathbf{1} = \mathbf{1},$$

the vector $\mathbf{1} \in \mathbb{R}^n$ is a right eigenvector with eigenvalue 1. Hence 1 is an eigenvalue of A. To see this is also the largest eigenvalue, let $z \in \mathbb{C}^n$ be an eigenvector of A with associated eigenvalue $\lambda \in \mathbb{C}$. That is,

$$Az = \lambda z.$$

Now let k be such that $|z_i| \leq |z_k|$ for all $i = 1, \ldots, n$. The kth entry of the equation above is

$$\sum_{j=1}^{n} a_{kj}\, z_j = \lambda z_k.$$

Hence

$$|\lambda z_k| = |\lambda| \cdot |z_k| = \left| \sum_{j=1}^{n} a_{kj}\, z_j \right| \leq \sum_{j=1}^{n} a_{kj}\, |z_j| \leq \sum_{j=1}^{n} a_{kj}\, |z_k| = |z_k|.$$

Therefore $|\lambda| \leq 1$. ∎

Remark If Π is the transition matrix for a Markov chain then Π is row-stochastic, hence it has a largest right eigenvalue of 1. If Π is irreducible and aperiodic, then by P-F theorem the eigenvalue of 1 is unique and all other eigenvalues have moduli strictly less than 1.

Proposition 0.3 *The right eigenvectors of A^T are the (transpose of the) left eigenvectors of A, and the corresponding eigenvalues are the same.*

Proof Let (λ, z) be an eigenpair of A^T. That is, $A^T z = \lambda z$. Then $z^T A = \lambda z^T$. So (λ, z^T) is a left eigenpair of A. ∎

Proposition 0.4 *A matrix and its transpose have the same set of eigenvalues.*

Proof Let A be a square matrix and note $(A - \lambda I)^T = A^T - \lambda I$ since the identity matrix I is symmetric. Thus since $det(B) = det(B^T)$ for any square matrix B,

$$det(A - \lambda I) = det((A - \lambda I)^T) = det(A^T - \lambda I),$$

hence A and A^T have the same characteristic polynomials and therefore the same set of eigenvalues. ∎

Remark The two propositions above mean that we can inspect the eigenvalues of a transition matrix Π and these will be the same as the left eigenvalues of Π. Furthermore, we may compute the eigenvectors for Π^T and those will be the left eigenvectors of Π.

Remark If Π is the transition matrix for a Markov chain then Π and Π^T have the same set of eigenvalues. We mentioned above that Π has a largest eigenvalue of 1, and hence Π^T has a largest eigenvalue of 1 as well. That is, there is an eigenvector $z \in \mathbb{R}^n$ such that $\Pi^T z = z$, which is true iff

$$z^T \Pi = z^T.$$

If Π is irreducible then z has strictly positive entries by P-F theorem. Since we can normalize z so the entries sum to 1, we know that any irreducible Markov chain has a stationary distribution.

If Π is irreducible and aperiodic then this stationary distribution is unique, by P-F theorem. More specifically, all other eigenvalues of Π are strictly less than 1 in modulus, so there is only one eigenvector z such that $z^T \Pi = z^T$, where z^T is the unique stationary distribution. Thus we state the following important result.

Theorem 0.5 *An irreducible, aperiodic, homogeneous Markov chain on a finite state space has a unique stationary distribution. Furthermore, if Π is diagonalizable, i.e., Π has n linearly independent eigenvectors, then the marginal distribution will converge to this unique stationary distribution as time tends to infinity regardless of the initial distribution.*

Before we prove this, note the following lemmas.

Lemma 0.6 *Let A be an $n \times n$ matrix. Then A is diagonalizable iff A has n linearly independent eigenvectors.*

In fact, $A = PDP^{-1}$ with D a diagonal matrix iff the columns of P are n linearly independent eigenvectors of A, and in this case the diagonal elements of D are the eigenvalues of A that correspond, respectively, to the eigenvectors in P.

Proof Let P be any $n \times n$ matrix with columns $z_1, \ldots, z_n$ and let D be an $n \times n$ diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Then

$$AP = A \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} = \begin{bmatrix} Az_1 & Az_2 & \cdots & Az_n \end{bmatrix},$$

and

$$PD = P \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 z_1 & \lambda_2 z_2 & \cdots & \lambda_n z_n \end{bmatrix}.$$

Assume A is diagonalizable and $A = PDP^{-1}$. Then $AP = PD$, which from above gives

$$\begin{bmatrix} Az_1 & Az_2 & \cdots & Az_n \end{bmatrix} = \begin{bmatrix} \lambda_1 z_1 & \lambda_2 z_2 & \cdots & \lambda_n z_n \end{bmatrix},$$

or

$$Az_1 = \lambda_1 z_1, Az_2 = \lambda_2 z_2, \ldots, Az_n = \lambda_n z_n.$$

Since P is invertible its columns are linearly independent. Since these columns are nonzero (otherwise they wouldn't be linearly independent), the above relations show that (λ_i, z_i) are eigenpairs for $i = 1, \ldots, n$. So, a diagonalizable matrix has n linearly independent eigenvectors, where the columns of P are these eigenvectors and the diagonal of D are the eigenvalues.

Now assume A has n linearly independent eigenvectors $z_1, \ldots, z_n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. Construct a matrix $P = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix}$ and a diagonal matrix $D = diag(\lambda_1, \ldots, \lambda_n)$. Then from above we see that $AP = PD$, which is true without the eigenvectors being linearly independent. Since the eigenvectors are linearly independent, P is invertible and so $A = PDP^{-1}$. ∎

Lemma 0.7 *If A has n linearly independent eigenvectors, then so does A^T.*

Proof Since A has n linearly independent eigenvectors, A may be diagonalized as $A = PDP^{-1}$, where the columns of P are the linearly independent eigenvectors of A and the diagonal elements of D are the eigenvalues of A. Then $A^T = (P^{-1})^T DP^T$. By the above lemma, the columns of $(P^{-1})^T$ are the n linearly independent eigenvectors of A^T. ∎

Remark The decomposition $A^T = (P^{-1})^T DP^T$ is another was of showing A and A_T have the same set of eigenvalues, but relies on the fact that A has a full set of linearly independent eigenvectors.

Proof We have already argued the existence of a unique stationary distribution. For the convergence, note if Π has n linearly independent eigenvectors $z_1, \ldots, z_n$ then any element of $\mathbb{R}^n$ may be written as a linear combination of eigenvectors. In particular, any initial probability mass function q_0 may be written

$$q_0 = \sum_{i=1}^{n} c_i z_i,$$

where $c_i \in \mathbb{R}$, $i = 1, \ldots, n$. The pmf at time $t = k$ is given by

$$q_k = \Pi q_{k-1}.$$

Note we can write q_k as

$$\begin{aligned} q_k &= \Pi q_{k-1} \\ &= \Pi^k q_0 \\ &= \Pi^k \sum_{i=1}^{m} c_i z_i \\ &= \sum_{i=1}^{m} c_i \Pi^k z_i \\ &= \sum_{i=1}^{m} c_i \lambda_i^k z_i. \end{aligned}$$

Since Π is irreducible and aperiodic there is only one eigenvalue, say λ_1, with modulus 1, and all other have modulus strictly less than one. Hence

$$\lim_{k \to \infty} q_k = \sum_{i=1}^{m} c_i \lambda_i^k z_i = c_1 z_1.$$

Thus $q := c_1 z_1$ is the unique stationary distribution that any initial distribution converges to. In particular since q is a pmf, the scaling factor c_1 is simply the normalizing constant making the sum of the entries in z_1 to be 1. ∎