Markov chains and the Frobenius-Perron theorem

1. The Perron-Frobenius theory of nonnegative matrices (based on notes by Mike Boyle)

A matrix of real numbers is called <u>nonnegative</u> if all the entries are nonnegative. It is called <u>positive</u> if all the entries are <u>positive</u>. A matrix is called <u>primitive</u> if it is square, nonnegative, and some power of the matrix is positive.

Theorem 1 (Perron theorem) If A is a primitive matrix, then A has a unique eigenvalue λ of greatest absolute value. λ is real, positive, of algebraic multiplicity 1, and has an associated eigenvalue all of whose coordinates are positive. (This condition has a simple interpretation in terms of the graph $G=(V,E)$ where $V=\{1,\dots,n\}$ and $E=\{(i,j) : a_{ij}>0\}$.)

Lemma 1 Let $T: V\to V$ be a linear transformation of a finite-dimensional real vector space. Let $E\subseteq V$ be a polytope (= bounded set which is an intersection of a finite number of closed half-spaces and has a nonempty interior). Assume that for some $k\geq 1$, $T^k(E)$ is contained in the interior of E. Then all eigenvalues of T have absolute value (strictly) less than 1.

<u>Proof:</u> Clearly, it's enough to prove this in the case $k=1$, i.e. we assume $T(E)\subseteq E^\circ$. Because $T(E)\subseteq E$, any eigenvalue λ satisfies $|\lambda|\leq 1$, so we need to explain why it can't happen that $|\lambda|=1$. First, if $\lambda=e^{2\pi i j/k}$ is a root of unity, then 1 is an eigenvalue of T^k, so there is a vector v on the boundary of E s.t.

$$T^k(v)=v\in\partial E,$$ in contradiction to the assumption that $T(E)\subseteq E^\circ$.

Second, if $\lambda = e^{i\theta}$ but λ is not a root of unity (i.e., θ is irrational), we know from linear algebra that there are vectors u, v such that $\begin{matrix} T(u) = \cos\theta\, u + \sin\theta\, v \\ T(v) = -\sin\theta\, u + \cos\theta\, u \end{matrix}$, i.e. T acts as a rotation by θ on $\operatorname{span}\{u,v\}$. In this case, if we take $w \in \partial E \cap \operatorname{span}\{u,v\}$ then the powers of T acting on w, $(T^k(w))_{k \geq 0}$, will have w as an accumulation point. This is impossible since these points are in the compact set $T(E) \subseteq E^\circ$. $\square$

<u>Proof of the Perron theorem</u>. First, we show there is an eigenvector with positive coordinates. Let Δ denote the standard simplex $\{v = (x_1 \ldots x_n) : x_i \geq 0 \ \sum_j x_j = 1\}$. Define a function $F : \Delta \to \Delta$ by $F(v) = \frac{Av}{\sum_{i=1}^{n} (Av)_i}$. (Note that Av is nonnegative and has at least one positive coordinate - otherwise A has a column of zeroes and A^k cannot be positive for any k.) Since F is continuous, there exists $v \in \Delta$ s.t. $F(v) = v$ (by the Brouwer fixed point theorem - is there a more constructive argument?), so v is an eigenvector: $Av = \lambda v$, and $\lambda > 0$. By the assumption that A is primitive, $v = \frac{1}{\lambda^k} A^k v$ so all of the coordinates of v are positive.

Next, we reduce the verification of the remaining claims to the case when $\lambda = 1$ and $v = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$. Note that the eigenvalue equation $\sum_{j=1}^{n} a_{ij} v_j = \lambda v_i$ can be rewritten as $\sum_{j=1}^{n} \frac{a_{ij} v_j}{\lambda v_i} = 1$. So, if we define a new matrix $B = (b_{ij})_{i,j=1}^{n}$ by $b_{ij} = \frac{a_{ij} v_j}{\lambda v_i}$, then $u = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is an eigenvector of B with eigenvalue 1. Furthermore, it is easy to see that B is still a primitive matrix (think of the graph paths interpretation). Finally note that any number μ is an eigenvalue of A if and only if $\frac{\mu}{\lambda}$ is an eigenvalue of B: If $w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$ satisfies $\sum_j a_{ij} w_j = \mu w_i$ then we can write $\frac{\mu}{\lambda} \frac{w_i}{v_i} = \sum_j \frac{a_{ij} v_j}{\lambda v_i} \frac{w_j}{v_j} = \sum_j b_{ij} \cdot \frac{w_j}{v_j}$.

It remains to show that in the case $\lambda = 1$, $v = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, all other eigenvalues satisfy $|\lambda|<1$. The trick is to now consider the action of A on row vectors, not column vectors: since $A\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}=\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$, it follows that for any $u=(u_1,\dots,u_n)\in\Delta$, $uA\in\Delta$, since $\sum_i (uA)_i=\langle uA,\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\rangle=uA\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}=\langle u,1\rangle=1$. So A acting from the left maps Δ to itself, and by the primitivity assumption, some power A^k maps Δ to its interior.

By the earlier part of the proof, there is a left-eigenvector $u=(u_1,\dots,u_n)\in\Delta$ with positive coordinates such that $uA=u$. Thus if we define $V=\{w=(w_1,\dots,w_n)\in\mathbb{R}^n:\ \sum w_j=0\}$, $E=\Delta-u$, then V is an A-invariant subspace of dimension $n-1$, and $E\subseteq V$ is a polytope such that $A^k(E)\subseteq E^o$ for some k. Applying Lemma 1, we see that all eigenvalues of $A_{|V}$ satisfy $|\lambda|<1$, which is just what we needed.

$\square$

<u>Corollary</u> 1: If A is a primitive matrix and u is a nonnegative eigenvector of A with eigenvalue β, then β is the eigenvalue of largest absolute value (so u is a scalar multiple of the positive eigenvector from the theorem).

<u>Proof</u>: By primitivity, $A^k u$ is positive for some k, so u is positive and $\beta>0$. If v is a positive eigenvector corresponding to the eigenvalue λ of largest absolute value, and we choose it such that $v\le u$, then for any $n\ge 1$, $\lambda^n v = A^n v\le A^n u=\beta^n u$. This cannot happen if $\beta<\lambda$, so $\beta=\lambda$.

$\square$

A square matrix A is called <u>irreducible</u> if for any $1 \le i,j \le n$ there is a $k \ge 1$ s.t. $(A^k)_{ij} > 0$. This means the graph associated with the matrix is (strongly) connected. It is a weaker condition than primitivity.

The <u>period</u> of an irreducible matrix is the gcd of all k such that $(A^k)_{ii} > 0$.

Suppose that A is irreducible with period p. Fix some vertex v_0. For $0 \le i \le p-1$ define $C_i = \{\text{vertex } u : \text{there is a path of length} \equiv i \pmod p \text{ from } v_0 \text{ to } u\}$.

The sets $C_0, C_1, \dots, C_{p-1}$ partition the vertex set V. All outgoing edges from a vertex in C_i lead to a vertex in $C_{i+1 \bmod p}$. By relabelling the rows and column indices of A, we can bring it to the form of a block matrix

$$A = \begin{pmatrix} 0 & A_1 & 0 & & & \\ 0 & 0 & A_2 & & 0 & \\ 0 & 0 & 0 & A_3 & & \\ & & & & \ddots & \\ 0 & & & & 0 & A_{p-1} \\ A_p & 0 & \cdots & & & 0 \end{pmatrix}$$

<u>Proposition</u>: Let A be a square nonnegative matrix. Then A is primitive if and only if it is irreducible with period one.

<u>Proof</u>: If A is primitive then it is irreducible, and the period must be 1 (otherwise A^k will have blocks of 0's). Conversely, if A is irreducible with period 1, using the exercise below it is not difficult to see that for large enough k there will be a path of length k in the graph associated with A between any two vertices. $\square$

Exercise: Prove that if $n_1, n_2, \dots, n_m$ are positive integers with $\gcd(n_1, \dots, n_m) = 1$, then the set of linear combinations of the form $\sum_{j=1}^{m} a_j n_j$ where $a_1, \dots, a_m$ are nonnegative integers contains all sufficiently large integers.

Theorem 2 (The Perron-Frobenius Theorem): Let A be an irreducible matrix with period p. Then A has a nonnegative right eigenvector v, unique up to multiplication by a scalar. The corresponding eigenvalue λ is the spectral radius. v has only positive coordinates, and has the following structure: if we write $v = \begin{pmatrix} v^0 \\ v^1 \\ \vdots \\ v^{p-1} \end{pmatrix}$ where v^i is the vector corresponding to the coordinates in the component $C(i)$ defined above, then v^0 is the positive eigenvector of the matrix $A_p \cdots A_3 A_2 A_1$, and $v^j = A_j v^{j-1}$

(Note: it is easy to see that $A_p \cdots A_2 A_1$ is a primitive matrix.)

Proof: The vector described in the theorem is clearly a positive eigenvector. Conversely, if $u = \begin{pmatrix} u^0 \\ \vdots \\ u^{p-1} \end{pmatrix}$ is a nonnegative eigenvector, then $A_1 u^0 = \lambda u^1$, $A_2 u^1 = \lambda u^2, \dots$, $A_{p-1} u^{p-2} = \lambda u^{p-1}$, $A_p u^{p-1} = \lambda u^0$, so

$A_p \cdots A_2 A_1 u^0 = \lambda^p u^0$. This shows that u is determined up to scalar multiplication. The other claims are left as an exercise. $\square$

Corollary 2: Powers of a primitive matrix: Suppose $A = (a_{ij})_{i,j=1}^{n}$ is a primitive matrix, and $u = (u_1, \dots, u_n)$, $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ are (respectively) a positive left eigenvector and a positive right eigenvector for the spectral radius λ, normalized so that $\langle u, v \rangle = uv = 1$. Then

$\frac{1}{\lambda^k} A^k$ converges exponentially fast to the matrix vu as $k\to\infty$. That is, for any $1\le i,j\le n$ we have

$$\frac{1}{\lambda^k}(A^k)_{i,j} \xrightarrow[k\to\infty]{} v_i u_j$$

Alternatively, we can write

$$(A^k)_{i,j} = v_i u_j \lambda^k + O(\mu^k) \qquad \text{for some } 0\le\mu<\lambda$$

Example: $A = \begin{pmatrix}1&3\\2&2\end{pmatrix}$. Check that $A^k = 4^k\begin{pmatrix}2/5&3/5\\2/5&3/5\end{pmatrix} + (-1)^k\begin{pmatrix}3/5&-3/5\\-2/5&2/5\end{pmatrix}$

Here $u=(2,3)$ $v=\binom{1}{1}$, $\lambda=4$.

Interpretation of the corollary for Markov chains: If A is the transition matrix for an irreducible, aperiodic Markov chain, then $\lambda=1$, $v=\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$, and $u=\pi=$ the unique stationary probability vector. In this case we get that for any $1\le i,j\le n$,

$$p_{i,j}^{(k)} = (A^k)_{i,j} \longrightarrow \pi_j \quad \text{exponentially fast as } k\to\infty.$$

I.e., after a long time the chain has forgotten its initial state and the distribution is very close to stationary.

Proof of the corollary: Let $B=\frac{1}{\lambda}A$. We have $uB=u$, $Bv=v$. If w is in one of the generalized eigenspaces $\operatorname{Ker}(A-\alpha I)^n$ for some eigenvalue $\alpha\ne\lambda$, then $|\alpha|<\lambda$, so

$$B^k w = \frac{1}{\lambda^k}A^k w \longrightarrow 0 \quad \text{exponentially fast as } k\to\infty.$$

($\|A^k w\|$ is bounded by α^k times a polynomial factor - exercise).

The same holds for row generalized eigenvectors. Therefore B^k converges to the unique matrix M that satisfies $uM=u$, $Mv=v$, and $Mw=0$, $w'M=0$ for any other generalized eigenvectors w, w'.

The matrix $N = vu$ is such a matrix: $uN = u(vu) = (uv)u = u$, $Nv = (vu)v = v(uv) = v$. If w is as before a generalized eigenvalue for an eigenvalue α, $|\alpha| < \lambda$, then

$$Mw = vuw = v(uB^k)w = vu(B^k w) \xrightarrow[k\to\infty]{} 0, \text{ and similarly}$$

$$w'M = w'vu = w'(B^k v)u = (w'B^k)vu \xrightarrow[k\to\infty]{} 0 \quad \text{for a generalized row eigenvector } w'.$$

$\square$

2. Finite state Markov chains

Let Σ be a finite set. A Markov chain with state space Σ is a sequence of random variables $X_0, X_1, X_2, \dots$ taking values in Σ such that

$$P(X_n = j \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) = p_{i_{n-1},j}$$

where $A = (p_{i,j})_{i,j\in\Sigma}$ is a fixed matrix of numbers called the transition matrix of the chain. Since $p_{i,j} \geq 0$ represent transition probabilities, we have $\sum_j p_{i,j} = 1$ for all i – the sums of all rows are 1. A matrix of nonnegative numbers satisfying this property is called stochastic. Note that this means that $A\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} = \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$, i.e., $\mathbb{1} = \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$ is a right-eigenvector of A with eigenvalue 1.

The vector $\pi = (P(X_0 = i))_{i\in\Sigma}$ is called the initial state distribution. The following lemma shows how to get the distribution of the n-th state X_n from π.

Lemma 2: $\left(P(X_n=j)\right)_{j\in\Sigma} = \pi A^n$ (matrix multiplication).

Proof: Induction:

$$P(X_n=j) = E\,P(X_n=j \mid X_{n-1}) = \sum_{i\in\Sigma} P(X_n=j \mid X_{n-1}=i)\,P(X_{n-1}=i)$$

$$= \sum_{i\in\Sigma} (\pi A^{n-1})_i \cdot p_{i,j} = (\pi A^n)_j . \qquad \square$$

~~The Markov chain is called irreducible if for any i,j∈Σ,~~

Denote $p^{(n)}_{i,j} = (A^n)_{i,j}$. $p^{(n)}_{i,j}$ represents the probability for a chain with transition matrix A started from (non-random) state $i\in\Sigma$ to be at state $j\in\Sigma$ after n steps. For a general initial state distribution π, we have $P(X_n=j) = \sum_i \pi_i\, p^{(n)}_{i,j}$.

The chain is called <u>irreducible</u> if for any $i,j\in\Sigma$ there exists an $n\geq 1$ s.t. $p^{(n)}_{i,j}>0$, i.e., if the transition matrix is irreducible. The chain is called <u>aperiodic</u> if the transition matrix is irreducible with period 1. (equivalently, A is primitive).

A state distribution vector $\pi = (\pi(i))_{i\in\Sigma}$ is called a <u>stationary vector</u> for the chain if $\pi A = \pi$ (i.e. π is a left eigenvector with eigenvalue 1). A chain started with this state distribution will satisfy

$$\left(P(X_n=j)\right)_{j\in\Sigma} = (\pi A^n)_j = \pi_j$$

for all n. That is, the variables $X_0, X_1, X_2, \ldots$ will be identically distributed.

Furthermore, because of the Markov property, it is easy to see that $X_0, X_1, X_2, \ldots$ is actually a stationary sequence.

Applying the Perron-Frobenius theory to the case of a Markov chain, we get the following result.

<u>Theorem 3:</u> Let A be the transition matrix for an irreducible Markov chain. There exists a unique ~~state~~ stationary distribution vector $\pi = (\pi(i))_{i\in\Sigma}$. If the chain is aperiodic, then we have for all $i,j \in \Sigma$, $p^{(n)}_{i,j} \xrightarrow[n\to\infty]{} \pi_j$ exponentially fast as $n \to \infty$.

3. Recurrence and ergodicity of Markov chains

<u>Theorem 4:</u> Let $X_0, X_1, \ldots$ be an irreducible Markov chain. Then all states $i \in \Sigma$ are <u>recurrent</u>, i.e.,

$$P(X_n = i \ \text{ i.o.}) = 1.$$

<u>Theorem 5:</u> Let $X_0, X_1, \ldots$ be a stationary irreducible Markov chain. Then the sequence $(X_n)_{n=0}^{\infty}$ is ergodic.

<u>Proof:</u> Let A be a shift-invariant event. Let π be the stationary vector. The probability $P(E) = P_\pi(E)$ of any event can be written as $P_\pi(E) = \sum_{j\in\Sigma} \pi_j P_j(E)$, where $P_j(\cdot)$ represents the distribution measure of the chain with the same transition matrix that is started from state $j \in \Sigma$. Thus we have $P(A) = \sum_j \pi_j P_j(A) = \sum_j \pi_j h(j)$, where we denote $h(j) = P_j(A) = E_j 1_A$.

Let S denote the shift transformation on $\mathbb{R}^{\mathbb{N}\cup\{0\}}$, and denote $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$. We have $1_A = 1_A \circ S^n$ a.s., so

$$E(1_A \mid \mathcal{F}_n) = E(1_A \circ S^n \mid \mathcal{F}_n) = \sum_{j\in\Sigma} 1_{\{X_n = j\}} E(1_A \circ S^n \mid X_n = j)$$

$$= \sum_{j\in\Sigma} 1_{\{X_n=j\}} E(1_A \mid X_0 = j) = \sum_j h(j) 1_{\{X_n = j\}} = h(X_n).$$

By Lévy's 0-1 law we have $E(1_A \mid \mathcal{F}_n) \xrightarrow[n\to\infty]{} 1_A$ a.s. So we have shown that $h(X_n) \xrightarrow[n\to\infty]{} 1_A$ a.s. Since by Theorem 4 all states are recurrent, it follows that $h \equiv \text{const}$ and $h(X_n) \xrightarrow[n\to\infty]{} \text{const.} = 1_A$ a.s., so $P(A) = 0$ or 1, which proves ergodicity. $\square$

<u>Proof of Theorem 4</u>. We assume the period is 1 - the general case is left as an exercise. In that case the transition matrix A is primitive. Let N_0 be such that all entries of A^{N_0} are positive. Let $c =$ the minimal entry of A^{N_0}. Fix $i \in \Sigma$ and consider the sequence of events $E_m = \{X_{N_0 m} = i\}$. We have $P(E_m \mid E_1, \dots, E_{m-1}) = E\left[P(E_m \mid X_1, \dots, X_{N_0(m-1)}) \mid E_1, \dots, E_{m-1}\right] \ge c$, so by a coupling argument, the number $S_m = \sum_{k=1}^m 1_{E_k}$ of occurrences of $E_1, \dots, E_m$ stochastically dominates a $\mathrm{Binom}(m,c)$ r.v.; in fact $(S_1, S_2, S_3, \dots)$ simultaneously dominate the process $\left(\sum_{k=1}^m Y_k\right)_{m\ge1}$ where $(Y_m)_m$ are i.i.d Bernoulli(c). It follows that $P(E_m \text{ i.o.}) \ge P(Y_k = 1 \text{ i.o.}) = 1$. $\square$