6.262: Discrete Stochastic Processes 2/28/11

Lecture 8: Markov eigenvalues and eigenvectors

Outline:

- Review of ergodic unichains
- Review of basic linear algebra facts
- Markov chains with 2 states
- Distinct eigenvalues for M > 2 states
- M states and M independent eigenvectors
- The Jordan form

Recall that for an ergodic finite-state Markov chain, the transition probabilities reach a limit in the sense that $\lim_{n\to\infty} P_{ij}^n = \pi_j$ where $\vec{\pi} = (\pi_1, \dots, \pi_{\mathsf{M}})$ is a strictly positive probability vector.

Multiplying both sides by P_{jk} and summing over j,

$$\pi_k = \lim_{n\to\infty} \sum_j P_{ij}^n P_{jk} = \sum_j \pi_j P_{jk}$$

Thus $\vec{\pi}$ is a steady-state vector for the Markov chain, i.e., $\vec{\pi} = \vec{\pi}[P]$ and $\vec{\pi} \geq 0$.

In matrix terms, $\lim_{n\to\infty}[P^n] = \vec{e}\vec{\pi}$ where $\vec{e} = (1, 1, \dots, 1)^{\mathsf{T}}$ is a column vector and $\vec{\pi}$ is a row vector.

The same result almost holds for ergodic unichains, i.e., one ergodic class plus an arbitrary set of transient states.

The sole difference is that the steady-state vector is positive for all ergodic states and 0 for all transient states.

$$[P] = \left[\begin{array}{c|c} [P_{\mathcal{T}}] & [P_{\mathcal{TR}}] \\ \hline [0] & [P_{\mathcal{R}}] \end{array}\right] \quad \textbf{where} \quad [P_{\mathcal{T}}] = \begin{bmatrix} P_{11} & \cdots & P_{1t} \\ \cdots & \cdots & \cdots \\ P_{t1} & \dots & P_{tt} \end{bmatrix}$$

The idea is that each transient state eventually has a transition (via $[P_{TR}]$) to a recurrent state, and the class of recurrent states lead to steady state as before.

Review of basic linear algebra facts

Def: A complex number λ is an eigenvalue of a real square matrix $[A]$, and a complex vector $\vec{v} \neq 0$ is a right eigenvector of $[A]$, if $\lambda\vec{v} = [A]\vec{v}$.

For every stochastic matrix (the transition matrix of a finite-state Markov chain $[P]$), we have $\sum_j P_{ij} = 1$ and thus $[P]\vec{e} = \vec{e}$.

Thus $\lambda = 1$ is an eigenvalue of an arbitrary stochastic matrix $[P]$ with right eigenvector $\vec{e}$.

An equivalent way to express the eigenvalue/eigenvector equation is that $[P - \lambda I]\vec{v} = 0$ where I is the identity matrix.

Def: A square matrix $[A]$ is singular if there is a vector $\vec{v} \neq 0$ such that $[A]\vec{v} = 0$.

Thus λ is an eigenvalue of $[P]$ if and only if (iff) $[P - \lambda I]$ is singular for some $\vec{v} \neq 0$.

Let $\vec{a}_1, \dots, \vec{a}_{\mathsf{M}}$ be the the columns of $[A]$. Then $[A]$ is singular iff $\vec{a}_1, \dots, \vec{a}_{\mathsf{M}}$ are linearly dependent.

The square matrix $[A]$ is singular iff the rows of $[A]$ are linearly dependent and iff the determinant $\det[A]$ of $[A]$ is 0.

Summary: λ is an eigenvalue of $[P]$ iff $[P - \lambda I]$ is singular, iff $\det[P - \lambda I] = 0$, iff $[P]\vec{v} = \lambda\vec{v}$ for some $\vec{v} \neq 0$, and iff $\vec{u}[P] = \lambda\vec{u}$ for some $\vec{u} \neq 0$.

For every stochastic matrix $[P]$, $[P]\vec{e} = \vec{e}$ and thus $[P - I]$ is singular and there is a row vector $\pi \neq 0$ such that $\vec{\pi}[P] = \vec{\pi}$.

This does <u>not</u> show that there is a probability vector $\vec{\pi}$ such that $\vec{\pi}[P] = \vec{\pi}$, but we already know there is such a probability vector (i.e., a steady-state vector) if $[P]$ is the matrix of an ergodic unichain.

We show later that there is a steady-state vector π for all Markov chains.

The determinant of an M by M matrix can be determined as

$$\det A = \sum_{\mu} \pm \prod_{i=1}^{\mathsf{M}} A_{i,\mu(i)}$$

where the sum is over all permutations μ of the integers $1, \ldots, \mathsf{M}$. Plus is used for each even permutation and minus for each odd.

The important facet of this formula for us is that $\det[P - \lambda I]$ must be a polynomial in λ of degree M.

Thus there are M roots of the equation $\det[P - \lambda I] = 0$, and consequently M eigenvalues of $[P]$.

Some of these M eigenvalues might be the same, and if k of these roots are equal to λ, the eigenvalue λ is said to have algebraic multiplicity k.

Markov chains with 2 states

$$\begin{array}{ll} \pi_1 P_{11} + \pi_2 P_{21} = \lambda \pi_1 & \quad P_{11}\nu_1 + P_{12}\nu_2 = \lambda \nu_1 \\ \pi_1 P_{12} + \pi_2 P_{22} = \lambda \pi_2 & \quad P_{21}\nu_1 + P_{22}\nu_2 = \lambda \nu_2 \end{array}.$$

left eigenvector   right eigenvector

$$\det[P - \lambda I] = (P_{11} - \lambda)(P_{22} - \lambda) - P_{12}P_{21}$$

$$\lambda_1 = 1; \qquad \lambda_2 = 1 - P_{12} - P_{21}$$

If $P_{12} = P_{21} = 0$ (the chain has 2 recurrent classes), then $\lambda = 1$ has multiplicity 2. Otherwise $\lambda = 1$ has multiplicity 1.

If $P_{12} = P_{21} = 1$ (the chain is periodic), then $\lambda_2 = -1$. Otherwise $|\lambda_2| < 1$.

$$\begin{array}{rlrl} \pi_1 P_{11} + \pi_2 P_{21} & = \lambda \pi_1 & \qquad P_{11}\nu_1 + P_{12}\nu_2 & = \lambda \nu_1 \\ \pi_1 P_{12} + \pi_2 P_{22} & = \lambda \pi_2 & \qquad P_{21}\nu_1 + P_{22}\nu_2 & = \lambda \nu_2 \end{array}.$$

$$\lambda_1 = 1; \qquad \lambda_2 = 1 - P_{12} - P_{21}$$

Assume throughout that either $P_{12} > 0$ or $P_{21} > 0$. Then

$$\begin{array}{rlrl} \vec{\pi}^{(1)} & = \left(\frac{P_{21}}{P_{12}+P_{21}}, \frac{P_{12}}{P_{12}+P_{21}}\right) & \qquad \vec{\nu}^{(1)} & = (1, 1)^{\mathsf{c}} \\ \vec{\pi}^{(2)} & = (1, -1) & \qquad \vec{\nu}^{(2)} & = \left(\frac{P_{12}}{P_{12}+P_{21}}, \frac{-P_{21}}{P_{12}+P_{21}}\right)^{\mathsf{c}} \end{array}$$

Note that $\vec{\pi}^{(i)}\vec{\nu}^{(j)} = \delta_{ij}$. In general, if $\vec{\pi}^{(i)}[P] = \lambda_i \vec{\pi}^{(i)}$ and $[P]\vec{\nu}^{(i)} = \lambda_i \vec{\nu}^{(i)}$ for $i = 1, \dots,$ M, then $\vec{\pi}^{(i)}\vec{\nu}^{(j)} = 0$ if $\lambda_i \neq \lambda_j$. To see this,

$$\lambda_i \vec{\pi}^{(i)}\vec{\nu}^{(j)} = \vec{\pi}^{(i)}[P]\vec{\nu}^{(j)} = \vec{\pi}^{(i)}(\lambda_j \vec{\nu}^{(j)}) = \lambda_j \vec{\pi}^{(i)}\vec{\nu}^{(j)}$$

so if $\lambda_i \neq \lambda_j$, then $\vec{\pi}_i \vec{\nu}_j = 0$. Normalization (of either $\vec{\pi}_i$ or $\vec{\nu}_i$) can make $\vec{\pi}_i \vec{\nu}_i = 1$ for each i.

Note that the equations

$$P_{11}\nu_1^{(i)} + P_{12}\nu_2^{(i)} = \lambda_i \nu_1^{(i)}; \qquad P_{21}\nu_1^{(i)} + P_{22}\nu_2^{(i)} = \lambda_i \nu_2^{(i)}$$

can be rewritten in matrix form as

$$[P][U] = [U][\Lambda] \qquad \textbf{where}$$

$$[\Lambda] = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad \textbf{and} \quad [U] = \begin{bmatrix} \nu_1^{(1)} & \nu_1^{(2)} \\ \nu_2^{(1)} & \nu_2^{(2)} \end{bmatrix},$$

Since $\vec{\pi}^{(i)}\vec{\nu}^{(j)} = \delta_{ij}$, we see that

$$\begin{bmatrix} \pi_1^{(1)} & \pi_2^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} \end{bmatrix} \begin{bmatrix} \nu_1^{(1)} & \nu_1^{(2)} \\ \nu_2^{(1)} & \nu_2^{(2)} \end{bmatrix} = [I],$$

so $[U]$ is invertible and $[U^{-1}]$ has $\vec{\pi}^{(1)}$ and $\vec{\pi}^{(2)}$ as rows. Thus $[P] = [U][\Lambda][U^{-1}]$ and

$$[P^2] = [U][\Lambda][U^{-1}][U][\Lambda][U^{-1}] = [U][\Lambda^2][U^{-1}]$$

Similarly, for any $n \geq 2$,

$$[P^n] = [U][\Lambda^n][U^{-1}] \quad (1)$$

Eq. 3.29 in text has a typo and should be (1) above.

We can solve (1) in general (if all M eigenvalues are distinct) as easily as for $\mathsf{M} = 2$.

Break $[\Lambda]^n$ into M terms,

$$[\Lambda]^n = [\Lambda_1^n] + \cdots + [\Lambda_{\mathsf{M}}^n] \qquad \text{where}$$

$[\Lambda_i^n]$ has λ_i^n in position (i, i) and has zeros elsewhere. Then

$$[P^n] = \sum_{i=1}^{\mathsf{M}} \lambda_i^n \vec{\nu}^{(i)} \vec{\pi}^{(i)}$$

$$[P^n] = \sum_{i=1}^{\mathsf{M}} \lambda_i^n \vec{\nu}^{(i)} \vec{\pi}^{(i)}$$

$$\begin{aligned} \vec{\pi}^{(1)} &= \left(\tfrac{P_{21}}{P_{12}+P_{21}}, \tfrac{P_{12}}{P_{12}+P_{21}}\right) & \vec{\nu}^{(1)} &= (1, 1)^{\mathsf{c}} \\ \vec{\pi}^{(2)} &= (1, -1) & \vec{\nu}^{(2)} &= \left(\tfrac{P_{12}}{P_{12}+P_{21}}, \tfrac{-P_{21}}{P_{12}+P_{21}}\right)^{\mathsf{c}} \end{aligned}$$

The steady-state vector is $\vec{\pi} = \vec{\pi}^{(1)}$ and

$$\vec{\nu}^{(1)}\vec{\pi} = \begin{bmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{bmatrix} \qquad \vec{\nu}^{(2)}\vec{\pi}^{(2)} = \begin{bmatrix} \pi_2 & -\pi_2 \\ -\pi_1 & \pi_1 \end{bmatrix}$$

$$[P^n] = \begin{bmatrix} \pi_1 + \pi_2\lambda_2^n & \pi_2 - \pi_2\lambda_2^n \\ \pi_1 - \pi_1\lambda_2^n & \pi_2 + \pi_1\lambda_2^n \end{bmatrix}$$

We see that $[P^n]$ converges to $\vec{e}\vec{\pi}$, and the rate of convergence is λ_2. This solution is exact. It essentially extends to arbitrary finite M.

Distinct eigenvalues for M > 2 states

Recall that, for an M state Markov chain, $\det[P-\lambda I]$ is a polynomial of degree M in λ. It thus has M roots (eigenvalues), which we assume here to be distinct.

Each eigenvalue λ_i has a right eigenvector $\vec{\nu}^{(i)}$ and a left eigenvector $\vec{\pi}^{(i)}$. Also $\vec{\pi}^{(i)}\vec{\nu}^{(j)} = 0$ for each $i, j \neq i$.

By scaling $\vec{\nu}^{(i)}$ or $\vec{\pi}^{(i)}$, we can satisfy $\vec{\pi}^{(i)}\vec{\nu}^{(i)} = 1$.

Let $[U]$ be the matrix with columns $\vec{\nu}^{(1)}$ to $\vec{\nu}^{(\mathsf{M})}$ and let $[V]$ have rows $\vec{\pi}^{(1)}$ to $\vec{\pi}^{(\mathsf{M})}$.

Then $[V][U] = I$, so $[V] = [U^{-1}]$. Thus the eigenvectors $\vec{\nu}^{(1)}$ to $\vec{\nu}^{(\mathsf{M})}$ are linearly independent and span M space. Same with $\vec{\pi}^{(1)}$ to $\vec{\pi}^{(\mathsf{M})}$.

Putting the right eigenvector equations together, $[P][U] = [U][\Lambda]$. Postmultiplying by $[U^{-1}]$, this becomes

$$[P] = [U][\Lambda][U^{-1}]$$

$$[P^n] = [U][\Lambda^n][U^{-1}]$$

Breaking $[\Lambda^n]$ into a sum of M terms as before,

$$[P^n] = \sum_{i=1}^{\mathsf{M}} \lambda_i^n \vec{\nu}^{(i)} \vec{\pi}^{(i)}$$

Since each row of $[P]$ sums to 1, $\vec{e}$ is a right eigenvector of eigenvalue 1.

Thm: The left eigenvector $\vec{\pi}$ of eigenvalue 1 is a steady-state vector if it is normalized to $\vec{\pi}\vec{e} = 1$.

Thm: The left eigenvector $\vec{\pi}$ of eigenvalue 1 is a steady-state vector if it is normalized to $\vec{\pi}\vec{e} = 1$.

Pf: There must be a left eigenvector $\vec{\pi}$ for eigenvalue 1. For every $j, 1 \le j \le \mathsf{M}$, $\pi_j = \sum_k \pi_k P_{kj}$. Taking magnitudes,

$$|\pi_j| \le \sum_k |\pi_k| P_{kj} \qquad (2)$$

with equality iff $\pi_j = |\pi_j| e^{i\phi}$ for all j and some ϕ. Summing over j, $\sum_j |\pi_j| \le \sum_k |\pi_k|$. This is satisfied with equality, so (2) is satisfied with equality for each j.

Thus $(|\pi_1|, |\pi_2|, \dots, |\pi_{\mathsf{M}}|$ is a nonnegative vector satisfying the steady-state vector equation. Normalizing to $\sum_j |\pi_j| = 1$, we have a steady-state vector.

Thm: Every eigenvalue λ_ℓ satisfies $|\lambda_\ell| \le 1$.

Pf: We have seen that if $\vec{\pi}^{(\ell)}$ is a left eigenvector of $[P]$ with eigenvalue λ_ℓ, then it is also a left eigenvector of $[P^n]$ with eigenvalue λ_ℓ^n. Thus

$$\lambda_\ell^n \, \pi_j^{(\ell)} \;=\; \sum_i \pi_i^{(\ell)} P_{ij}^n \qquad \text{for all } j.$$

$$|\lambda_\ell^n| \, |\pi_j^{(\ell)}| \;\le\; \sum_i |\pi_i^{(\ell)}| \, P_{ij}^n \qquad \text{for all } j.$$

Let β be the largest of $|\pi_j^{(\ell)}|$ over j. For that maximizing j,

$$|\lambda_\ell^n| \, \beta \;\le\; \sum_i \beta \, P_{ij}^n \;\le \beta \mathsf{M}$$

Thus $|\lambda_\ell^n| \le \mathsf{M}$ for all n, so $|\lambda_\ell| \le 1$.

These two theorems are valid for all finite-state Markov chains. For the case with M distinct eigenvalues, we have

$$[P^n] = \sum_{i=1}^{\mathsf{M}} \lambda_i^n \vec{\nu}^{(i)} \vec{\pi}^{(i)}$$

If the chain is an ergodic unichain, then one eigenvalue is 1 and the rest are strictly less than 1 in magnitude.

Thus the rate at which $[P^n]$ approaches $\vec{e}\vec{\pi}$ is determined by the second largest eigenvalue.

If $[P]$ is a periodic unichain with period d, then there are d eigenvalues equally spaced around the unit circle and $[P^n]$ does not converge.

M states and M independent eigenvectors

Next assume that one or more eigenvalues have multiplicity greater than 1, but that if an eigenvalue has multiplicity k, then it has k linearly independent eigenvectors.

We can choose the left eigenvectors of a given eigenvalue to be orthonormal to the right eigenvectors of that eigenvalue.

After doing this and defining $[U]$ as the matrix with columns $\vec{\nu}^{(1)}, \dots, \vec{\nu}^{(\mathsf{M})}$, we see $[U]$ is invertible and that $[U^{-1})$ is the matrix with rows $\vec{\pi}^{(1)}, \dots, \vec{\pi}^{(\mathsf{M})}$. We then again have

$$[P^n] = \sum_{i=1}^{\mathsf{M}} \lambda_i^n \vec{\nu}^{(i)} \vec{\pi}^{(i)}$$

Example: Consider a Markov chain consisting of ℓ ergodic sets of states.

Then each ergodic set will have an eigenvalue equal to 1 with a right eigenvector equal to 1 on the states of that set and 0 elsewhere.

There will also be a 'steady-state' vector, nonzero only on that set of states.

Then $[P^n]$ will converge to a block diagonal matrix where for each ergodic set, the rows within that set are the same.

The Jordan form

Unfortunately, it is possible that an eigenvalue of algebraic multiplicity $k \geq 2$ has fewer than k linearly independent eigenvectors.

The decomposition $[P] = [U][\Lambda][U^{-1}]$ can be replaced in this case by a Jordan form, $[P] = [U][J][U^{-1}]$ where $[J]$ has the form

$$[J] = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 1 & 0 \\ 0 & 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{bmatrix}.$$

The eigenvalues are on the main diagonal and ones are on the next diagonal up where needed for deficient eigenvectors.

Example:

$$[P] = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{bmatrix}.$$

The eigenvalues are 1 and 1/2, with algebraic multiplicity 2 for $\lambda = 1/2$**.**

There is only one eigenvector (subject to a scaling constant) for the eigenvalue 1/2. $[P^n]$ **approaches steady-state as** $n(1/2)^n$**.**

Fortunately, if $[P]$ **is stochastic, the eigenvalue 1 always has as many linearly independent eigenvectors as its algebraic multiplicity.**