PERRON FROBENIUS THEOREM

R. CLARK ROBINSON

Definition 1. A $n \times n$ matrix M with real entries m_{ij}, is called a *stochastic matrix* provided (i) all the entries m_{ij} satisfy $0 \leq m_{ij} \leq 1$, (ii) each of the columns sum to one, $\sum_i m_{ij} = 1$ for all j, (iii) each row has some nonzero entry (it is possible to make a transition to each of the i^{th}-states from some other state), and (iv) some column has more than one nonzero entry (from one of the j^{th}-states there are two possible following states). The entry m_{ij} is the probability that something taken from the j^{th}-state is returned to the i^{th}-state.

In probability, usually the rows are assumed to sum to one rather than the columns. However, in that situation, the columns are written as rows and the matrix is multiplied on the right of the vector. We retain column vectors which are multiplied on the left by the matrix.

A stochastic matrix M is called *regular* or *eventually positive* provided there is a $q_0 > 0$ such that M^{q_0} has all positive entries. This means that for this iterate, it is possible to make a transition from any state to any other state. It then follows that M^q has all positive entries for $q \geq q_0$. A regular stochastic matrix automatically satisfies conditions (iii) and (iv) in the definition of a stochastic matrix.

Let $x_j^{(0)} \geq 0$ is the amount of the material at the j^{th}-location at time 0. Then $m_{ij}x_j^{(0)}$ is the amount of material from the j^{th}-location that is returned to the i^{th}-location at time 1. The total amount at the i^{th}-location at time 1 is the sum of the material from all the sites,

$$x_i^{(1)} = \sum_j m_{ij} x_j^{(0)}.$$

Let

$$\mathbf{x}^{(q)} = \begin{pmatrix} x_1^{(q)} \\ \vdots \\ x_n^{(q)} \end{pmatrix}$$

be the vector of the amount of material at all the sites. By the above formula,

$$\mathbf{x}^{(1)} = M\mathbf{x}^{(0)}$$

and more generally

$$\mathbf{x}^{(q+1)} = M\mathbf{x}^{(q)}.$$

Date: April 12, 2002.

Notice that the total amount of material at time 1 is the same as at time 0:

$$\begin{aligned}\sum_i x_i^{(1)} &= \sum_i \left(\sum_j m_{ij} x_j^{(0)} \right) \\ &= \sum_j \left(\sum_i m_{ij} \right) x_j^{(0)} \\ &= \sum_j x_j^{(0)}.\end{aligned}$$

Call this amount X. Then,

$$p_j^{(q)} = \frac{x_j^{(q)}}{X}$$

is the proportion of the material at the j^{th}-site at time q. Letting

$$\mathbf{p}^{(q)} = \left(p_j^{(q)} \right) = \frac{1}{X} \mathbf{x}^{(q)}$$

be the vector of these proportions,

$$\begin{aligned}M\mathbf{p}^{(q)} &= M \frac{\mathbf{x}^{(q)}}{X} \\ &= \frac{1}{X} M\mathbf{x}^{(q)} \\ &= \frac{1}{X} \mathbf{x}^{(q+1)} \\ &= \mathbf{p}^{(q+1)}\end{aligned}$$

also transforms by multiplication by the matrix M.

For a stochastic matrix M, 1 is always an eigenvalue of the the transpose of M, M^T, with eigenvector $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$:

$$M^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} \sum_i m_{i1} 1 \\ \vdots \\ \sum_i m_{in} 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Since M^T and M have the same eigenvalues, M always has 1 as an eigenvalue.

Before stating the general result, we give some examples.

Example 1. Let

$$M = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{pmatrix}.$$

This has eigenvalues 1, 0.5, and 0.2. (We do not give the characteristic polynomial, but do derive an eigenvector for each of these values.)

For $\lambda = 1$,

$$M - I = \begin{pmatrix} -0.5 & 0.2 & 0.3 \\ 0.3 & -0.2 & 0.3 \\ 0.2 & 0 & -0.6 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & -0.4 & -0.6 \\ 0 & -0.08 & 0.48 \\ 0 & 0.08 & -0.48 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & -6 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus, $v_1 = 3v_3$ and $v_2 = 6v_3$. Since we want $1 = v_1 + v_2 + v_3 = (3 + 6 + 1)v_3 = 10v_3$, $v_3 = 0.1$, and

$$\mathbf{v}^1 = \begin{pmatrix} 0.3 \\ 0.6 \\ 0.1 \end{pmatrix}.$$

For $\lambda_2 = 0.5$,

$$M - 0.5\,I = \begin{pmatrix} 0 & 0.2 & 0.3 \\ 0.3 & 0.3 & 0.3 \\ 0.2 & 0 & -0.1 \end{pmatrix}$$
$$\sim \begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix}$$
$$\sim \begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & 3 \\ 0 & 1 & 1.5 \end{pmatrix}$$
$$\sim \begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

Thus, $2v_1 = v_3$ and $2v_2 = 3v_3$, and

$$\mathbf{v}^2 = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}.$$

Notice that $v_1^2 + v_2^2 + v_3^2 = 1 - 3 + 2 = 0$. This is always the case for the eigenvectors of the other eigenvalues.

For $\lambda_3 = 0.2$,

$$\begin{aligned} M - 0.2\,I &= \begin{pmatrix} 0.3 & 0.2 & 0.3 \\ 0.3 & 0.6 & 0.3 \\ 0.2 & 0 & 0.2 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 1 \\ 3 & 6 & 3 \\ 3 & 2 & 3 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 6 & 0 \\ 0 & 2 & 0 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Thus, $v_1 = -v_3$ and $v_2 = 0$, and

$$\mathbf{v}^3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Notice that $v_1^3 + v_2^3 + v_3^3 = 1 + 0 - 1 = 0$ is true in this case as well.

If the original distribution is given by

$$\mathbf{p}^{(0)} = \begin{pmatrix} 0.45 \\ 0.45 \\ 0.1 \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.6 \\ 0.1 \end{pmatrix} + \frac{1}{20}\begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} + \frac{1}{10}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},$$

then

$$M^q\mathbf{p}^{(0)} = \begin{pmatrix} 0.3 \\ 0.6 \\ 0.1 \end{pmatrix} + \frac{1}{20}\left(\frac{1}{2}\right)^q\begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} + \frac{1}{10}\left(\frac{1}{5}\right)^q\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

which converges to the distribution

$$\mathbf{v}^1 = \begin{pmatrix} 0.3 \\ 0.6 \\ 0.1 \end{pmatrix}$$

Take any initial distribution $\mathbf{p}^{(0)}$ with $\sum_i p_i^{(0)} = 1$. Writing $\mathbf{p}^{(0)} = y_1\mathbf{v}^1 + y_2\mathbf{v}^2 + y_3\mathbf{v}^3$,

$$\begin{aligned} 1 &= \sum_i p_i^{(0)} \\ &= y_1\left(\sum_i v_i^1\right) + y_2\left(\sum_i v_i^2\right) + y_3\left(\sum_i v_i^3\right) \\ &= y_1(1) + y_2(0) + y_3(0) \\ &= y_1. \end{aligned}$$

Thus,

$$\mathbf{p}^{(0)} = \mathbf{v}^1 + y_2\mathbf{v}^2 + y_3\mathbf{v}^3,$$

and

$$M^q \mathbf{p}^{(0)} = \mathbf{v}^1 + y_2(0.5)^q \mathbf{v}^2 + y_3(0.2)^q \mathbf{v}^3$$

converges to $\mathbf{v}^1$, the eigenvector for $\lambda = 1$.

Example 2 (Complex Eigenvalues)**.** The following stochastic matrix illustrates the fact that a regular stochastic matrix can have complex eigenvalues. Let

$$M = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.1 & 0.6 \end{pmatrix}.$$

The eigenvalues are $\lambda = 1$ and $0.4 \pm i\,0.1\sqrt{3}$. Notice that $|0.4 \pm i\,0.1\sqrt{3}| = \sqrt{0.16 + 0.03} = \sqrt{0.19} < 1$.

Example 3 (Not Regular)**.** An example of a stochastic matrix which is not regular (nor transitive) is given by

$$M = \begin{pmatrix} 0.8 & 0.3 & 0 & 0 \\ 0.2 & 0.7 & 0 & 0 \\ 0 & 0 & 0.6 & 0.3 \\ 0 & 0 & 0.4 & 0.7 \end{pmatrix},$$

which has eigenvalues $\lambda = 1$, 1, 0.5, and 0.3. Notice that states 1 and 2 interchange with each other and states 3 and 4 interchange, but there is no interchange between the pair of sites 1 and 2 with the pair of sites 3 and 4.

An example of a stochastic matrix which is transitive but not regular is given by

$$M = \begin{pmatrix} 0 & 0 & 0.8 & 0.3 \\ 0 & 0 & 0.2 & 0.7 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

which has eigenvalues $\lambda = 1$, -1, and $\pm\sqrt{0.5}$. Here is is possible to get from any site to any other site, but starting at site one, the even iterates are always at either sites 3 or 4 and the odd iterates are always at either sites 1 or 2. Thus there is no one power for which all the transition probabilities are positive. Thus, M is not regular.

The following theorem summarizes some of the results about regular stochastic matrices which the above examples illustrated.

Theorem 0.1. *Let M be a regular stochastic matrix.*

(a) The matrix M has 1 as a eigenvalue of multiplicity one. The eigenvector $\mathbf{v}^1$ can be chosen with all positive entries and $\sum_i v_i^1 = 1$. (It must have either all positive entries or all negative entries.)

(b) All the other eigenvalues λ_j have $|\lambda_j| < 1$. If $\mathbf{v}^j$ is the eigenvector for λ_j, then $\sum_i v_i^j = 0$.

(c) If $\mathbf{p}$ is any probability distribution with $\sum_i = 1$, then

$$\mathbf{p} = \mathbf{v}^1 + \sum_{j=2}^{n} y_j \mathbf{v}^j.$$

Also, $M^q \mathbf{p}$ goes to $\mathbf{v}^1$ as q goes to infinity.

Sketch of the proof. We assume below that all the $m_{ij} > 0$, which can be done by taking a power if necessary.

(a) The multiplicity of the eigenvalue 1 is one. We noted above M^T always has 1 as an eigenvalues so it is always an eigenvalue of M. Again, to discuss the multiplicity, we look at M^T. Assume $M^T\mathbf{v} = \mathbf{v}$ and not all the v_j are equal. Assume that k is the index for which $|v_k$ is largest. By scalar multiplication by -1 if necessary we can take v_k positive. Thus, $v_k = |v_k| \geq |v_j|$ for all j and $v_k > |v_\ell|$ for some ℓ. Then,

$$\begin{aligned} v_k &= \sum_j m_{jk} v_j \\ &< \sum_j m_{jk} v_k \\ &= v_k. \end{aligned}$$

The second strict inequality uses the fact that all the $m_{ij} > 0$, i.e., that M is regular. Since this shows $v_k > v_k$, the contradiction implies that there are no such other vectors, i.e., that there can only be one eigenvector for the eigenvalue 1.

To complete the proof, we would have to consider the case with only one eigenvector but an algebraic multiplicity of the characteristic equation. We leave this detail to the references.

(b) Case (i): Assume λ is a real eigenvalue. We show that $\lambda < 1$. Again, assume that $M^T\mathbf{v} = \lambda\mathbf{v}$. Let k be such that $v_k = |v_k| \geq |v_j|$ for all j and $v_k > |v_\ell|$ for some ℓ. Then,

$$\begin{aligned} \lambda v_k &= \sum_j m_{jk} v_j \\ &< \sum_j m_{jk} v_k \\ &= v_k. \end{aligned}$$

This shows that $\lambda v_k < v_k$, so $\lambda < 1$.

Case (ii): Assume λ is a real eigenvalue. We show that $\lambda > -1$. Again, assume that $M^T\mathbf{v} = \lambda\mathbf{v}$. Let k be such that $v_k = |v_k| \geq |v_j|$ for all j and $v_k > |v_\ell|$ for some ℓ, i.e., $v_j \geq -v_k$ for all j and $v_\ell > -v_k$ for some ℓ. Then,

$$\begin{aligned} \lambda v_k &= \sum_j m_{jk} v_j \\ &> \sum_j m_{jk} (-v_k) \\ &= -v_k. \end{aligned}$$

This shows that $\lambda v_k > -v_k$, so $\lambda > -1$.

Care (iii) Assume $\lambda = r\, e^{2\pi\omega i}$ is a complex eigenvalue with complex eigenvector $\mathbf{v}$. Assume the v_j are chosen with v_k real and $v_k \geq \mathrm{Re}(v_j)$ for all j. Then,

$$\begin{aligned}
\mathrm{Re}(\lambda^q v_k) &= \mathrm{Re}(r^q\, e^{2\pi q\omega i} v_k) \\
&= \mathrm{Re}((M^q \mathbf{v})_k) \\
&= (M^q \mathbf{v})\mathrm{Re}(\mathbf{v}) \\
&= \sum_j \left(m_{jk}^{(q)} \mathrm{Re}(v_j) \right) \\
&< \sum_j \left(m_{jk}^{(q)} v_k \right) \\
&= v_k.
\end{aligned}$$

Therefore, $r^q \mathrm{Re}(e^{2\pi q\omega i}) < 1$ for all q. Since we can find a q_1 for which $\mathrm{Re}(e^{2\pi q_1 \omega i})$ is very close to 1, we need $r^{q_1} < 1$ so $r = |\lambda| < 1$.

(c) Let $\mathbf{p}$ be a probability distribution with $\sum_i p_i = 1$. The eigenvectors are a basis, so there exist $y_1, \ldots y_n$ such that

$$\mathbf{p} = \sum_{j=1}^{n} y_j\, \mathbf{v}^j.$$

Then,

$$\begin{aligned}
1 &= \sum_i p_i \\
&= y_1 \sum_i v_i^1 + \sum_{j=2}^{n} y_j \sum_i v_i^j \\
&= y_1 + \sum_{j=2}^{n} y_j\, (0) \\
&= y_1.
\end{aligned}$$

Thus,

$$\mathbf{p} = \mathbf{v}^1 + \sum_{j=2}^{n} y_j\, \mathbf{v}^j$$

as claimed.

Writing the iteration as if all the eigenvalues are real,

$$M^q \mathbf{p} = M^q \mathbf{v}^1 + \sum_{j=2}^{n} y_j\, M^q \mathbf{v}^j \qquad = \mathbf{v}^1 + \sum_{j=2}^{n} y_j\, \lambda_j^q \mathbf{v}^j$$

which tends to $\mathbf{v}^1$ because all the $\lambda_j^q| < 1$ for $j \geq 2$. □

References

1. Gantmacher, F.R., *The Theory of Matrices, Volume I, II*, Chelsea Publ. Co., New York, 1959.
2. Strang, G., *Linear Algebra and Its Applications, Third Edition*, Harcourt Brace Jovanovich Publishers, San Diego, 1988.

Department of Mathematics, Northwestern University, Evanston IL 60208
E-mail address: clark@math.northwestern.edu