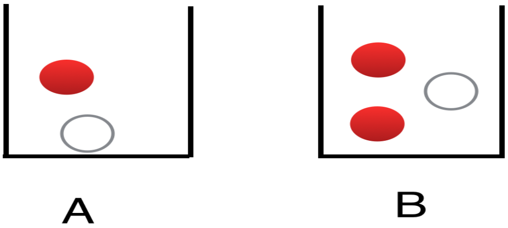
Stochastic Systems: solving Markov chains.

May 10, 2018

Urn model: balls of various colours in a pot

Suppose two pots, A and B, each contain 3 red balls and 2 white balls between them so that A always has 2 balls and B always has 3 balls. Note: this is Exercise 5.1. of Reichl

What are the states of the system?



What are the transition probabilities?

To find these have to define the rules that govern the model.

Dynamical rule: At each time step pick a ball out of pot A at random and one out of pot B at random, and interchange them

- Start from state 1. It can only go to state 2.

$$\Rightarrow \quad Q_{11} = 0\,,\ Q_{21} = 1\,,\ Q_{31} = 0$$

- Start from state 2.

$$\begin{aligned}
\text{To go to } n=1 \quad &: \quad \text{probability } = \frac{1}{2}.\frac{1}{3} = \frac{1}{6} \\
\text{To go to } n=3 \quad &: \quad \text{probability } = \frac{1}{2}.\frac{2}{3} = \frac{1}{3} \\
\text{To go to } n=2 \quad &: \quad \text{probability } = 1 - \frac{1}{6} - \frac{1}{3} = \frac{1}{2} \\
n=2 \text{ directly} \quad &: \quad \text{probability } = \frac{1}{2}.\frac{1}{3} + \frac{1}{2}.\frac{2}{3} = \frac{1}{2}
\end{aligned}$$

$$\Rightarrow \quad Q_{12} = \frac{1}{6}\,,\ Q_{22} = \frac{1}{2}\,,\ Q_{32} = \frac{1}{3}$$

- Start from state 3.

$$\begin{aligned} \text{To go to } n=1 \quad &: \quad \text{probability } = 0 \\ \text{To go to } n=2 \quad &: \quad \text{probability } = 1.\frac{2}{3} = \frac{2}{3} \\ \text{To go to } n=3 \quad &: \quad \text{probability } = 1.\frac{1}{3} = \frac{1}{3} \end{aligned}$$

$$\Rightarrow \quad Q_{13} = 0\,,\ Q_{23} = \frac{2}{3}\,,\ Q_{33} = \frac{1}{3}$$

So the transition probability matrix is:

$$Q = \begin{pmatrix} 0 & \frac{1}{6} & 0 \\ 1 & \frac{1}{2} & \frac{2}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Notice that each of the columns add up to 1. Matrices which have this property and in addition have entries which are all non-negative, are called stochastic matrices.

So suppose we are given $P(0)$ — say that the system starts in state 3 at $t = 0$. This means that

$$P(0) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Then

$$P(1) = QP(0)\,;\; P(2) = QP(1) = Q^2P(0)\,;\ldots;\; P(3) = Q^3P(0)\;\ldots$$

Continuing in this way we arrive at the general result

$$P(t) = Q^tP(0)$$

We will discuss later how to study this for general t, but for the moment let's just look at what happens after one and two time steps in this example

Since

$$Q^2 = \begin{pmatrix} \frac{1}{6} & \frac{1}{12} & \frac{1}{9} \\ \frac{1}{2} & \frac{23}{36} & \frac{20}{36} \\ \frac{1}{3} & \frac{10}{36} & \frac{1}{3} \end{pmatrix},$$

we find that if we start in state 2, then after one step

$$P(1) = \begin{pmatrix} 0 & \frac{1}{6} & 0 \\ 1 & \frac{1}{2} & \frac{2}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{6} \\ \frac{1}{2} \\ \frac{1}{3} \end{pmatrix}$$

and after two steps

$$P(2) = \begin{pmatrix} \frac{1}{6} & \frac{1}{12} & \frac{1}{9} \\ \frac{1}{2} & \frac{23}{36} & \frac{20}{36} \\ \frac{1}{3} & \frac{10}{36} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{12} \\ \frac{23}{36} \\ \frac{10}{36} \end{pmatrix}$$

So, for example, the probability of being in state 2 having started in state 2 two time steps earlier, is $23/36$

General solution of a Markov chain

Let us return to the general equation for Markov chains, when the transition probability is time-independent:

$$P(t) = Q^t P(0)$$

How do we solve it for general t?

Answer: We need to find the eigenvalues and eigenvectors of the matrix Q.

We first need to have an aside on the eigenvalues and eigenvectors of a (typically non-symmetric) matrix.

Let us look at the specific example:

$$M = \begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix},$$

which clearly is not a stochastic matrix (its columns don't sum to unity, and it has negative entries).

So, we compute its eigenvalues in the usual way; subtract λ from the diagonals, and set the determinant to zero;

$$\begin{vmatrix} -1-\lambda & 2 \\ -3 & 4-\lambda \end{vmatrix} = 0 \quad \Rightarrow \quad \lambda^2 - 3\lambda + 2 = 0.$$

Thus, the characteristic equation (the one on the right, above) has solutions

$$\lambda^{(1)} = 1, \quad \lambda^{(2)} = 2.$$

We find what we will now call the *right eigenvectors* the way we "normally" find eigenvectors. That is, by solving

$$\begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$$

for each of the cases $\lambda = 1, \lambda = 2$ separately

We shall denote the right eigenvector as $\psi^{(i)}$, corresponding to eigenvalue $\lambda^{(i)}$.

We find that

$$\lambda^{(1)} = 1 \quad \Rightarrow \quad \psi^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and

$$\lambda^{(2)} = 2 \quad \Rightarrow \quad \psi^{(2)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Now, to find the left eigenvectors $\chi^{(i)}$, which correspond to eigenvector $\lambda^{(i)}$, we solve

$$(x \quad y) \begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix} = \lambda^{(i)} (x \quad y),$$

for each $\lambda^{(i)}$. So, for $\lambda^{(1)} = 1$, we see that

$$-x - 3y = x \quad \Rightarrow \quad \text{left eigenvector is} \quad (3 \quad -2),$$

and that for $\lambda^{(2)} = 2$, we see that

$$x = -y \quad \Rightarrow \quad \text{left eigenvector is} \quad (1 \quad -1).$$

Note that these eigenvectors are rows, rather than columns. We therefore denote them as $(\chi^{(i)})^T$. That is,

$$(\chi^{(1)})^T = (3 \quad -2) \quad \Leftrightarrow \quad \chi^{(1)} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

$$(\chi^{(2)})^T = (1 \quad -1) \quad \Leftrightarrow \quad \chi^{(2)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Now, by way of convenient notation, we denote right eigenvectors as kets $|\psi^{(i)}\rangle$, and left eigenvectors as bras $\langle\chi^{(i)}|$. Thus, notice the orthogonality of the two sets of eigenvectors:

$$\langle\chi^{(1)}|\psi^{(2)}\rangle = 6 - 6 = 0, \quad \langle\chi^{(2)}|\psi^{(1)}\rangle = 1 - 1 = 0.$$

They can also be normalised:

$$\langle\chi^{(1)}|\psi^{(1)}\rangle = 3 - 2 = 1,$$

so this set is already normalised, and

$$\langle\chi^{(2)}|\psi^{(2)}\rangle = 2 - 3 = -1,$$

so we simply multiply $\chi^{(2)}$ by -1.

Summary

$$\psi^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad , \quad \psi^{(2)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

$$\chi^{(1)} = \begin{pmatrix} 3 \\ -2 \end{pmatrix} \quad , \quad \chi^{(2)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$(\chi^{(1)})^T = (3 \quad -2) \quad , \quad (\chi^{(2)})^T = (-1 \quad 1)$$

Warning: M is not a stochastic matrix, so do not read anything into the nature of the eigenvalues or eigenvectors. In particular, the fact that one of the eigenvalues happens to be 1 and one of the eigenvectors $(1 \quad 1)^T$ will not be true for another choice of matrix.

With this, let us return to the study of Markov chains, and stochastic matrices (whose entries are all non-negative and where the entries in any one column add up to unity). We saw this latter property in the random walk model example, but this is obviously true in general, since "something must happen".

Mathematically, this idea that some event must happen can be formulated by using the definition of Q. We see that it corresponds to

$$\sum_n Q_{nm} = \sum_n P(n, t+1|m, t) = 1,$$

so that the probability that the state transitions from state m at time t, to all other states n at time $t+1$, is unity.

Properties of stochastic matrices

1. If Q_1 and Q_2 are stochastic matrices, so is $Q_1 Q_2$.
2. If Q is a stochastic matrix, then so is any power of that matrix, Q^t.
3. In our brief example, we saw that an eigenvalue $\lambda = 1$ appeared. This is a general property of stochastic matrices. All stochastic matrices have one eigenvalue which is unity. Corresponding to that eigenvalue, the left eigenvector is "unit". That is,

$$\lambda^{(1)} = 1, \quad (\chi^{(1)})^T = (1 \quad 1 \quad \dots \quad 1).$$

This follows just by writing the condition $\sum_n Q_{nm} = 1$ as

$$\sum_n \chi_n^{(1)} Q_{nm} = 1.\chi_m^{(1)},$$

where $\chi_n^{(1)} = 1$ for all n. Note that our example during the brief aside, was not a stochastic matrix, and therefore this left eigenvector did not appear.

4. Suppose that Q is time-independent. Then at large times $P(t)$ may approach a time-independent value called P^{st} — the *stationary state*. In that case $P(t+1)$ and $P(t)$ both equal P^{st}, so that

$$P(t+1) = QP(t),$$

becomes

$$P^{st} = QP^{st},$$

which is merely an eigenvalue equation, corresponding to eigenvalue 1. That is,

$$QP^{st} = 1.P^{st}.$$

so that P_n^{st} is a right-eigenvector of Q with eigenvalue 1

Note that

$$(1 \quad 1 \quad \dots \quad 1)\begin{pmatrix} P_1^{st} \\ P_2^{st} \\ \vdots \end{pmatrix} = \sum_n P_n^{st} = 1,$$

so that the correct normalisation of P^{st} implies that $(\chi^{(1)})^T \psi^{(1)} = 1$ and visa-versa.

5. All eigenvalues of a stochastic matrix have modulus ≤ 1,

$$|\lambda^{(i)}| \leq 1, \quad \forall i.$$

(see Reichl p236 for a proof).

Finally, if the matrix is symmetric, $Q^T = Q$, then the left- and right-eigenvectors are the same.

Example 3 red balls and 2 white balls in urns A and B (continued)

Let us compute the right & left eigenvectors for our previous transition matrix for the urn model:

$$Q = \begin{pmatrix} 0 & 1/6 & 0 \\ 1 & 1/2 & 2/3 \\ 0 & 1/3 & 1/3 \end{pmatrix}.$$

We compute eigenvalues via

$$\begin{vmatrix} -\lambda & 1/6 & 0 \\ 1 & 1/2 - \lambda & 2/3 \\ 0 & 1/3 & 1/3 - \lambda \end{vmatrix} = 0,$$

which results in the characteristic equation

$$\lambda^3 - \frac{5}{6}\lambda^2 - \frac{2}{9}\lambda + \frac{1}{18} = 0.$$

To solve this cubic, we first note that we know that one factor is $\lambda = 1$ (it being a stochastic matrix implies that there is one unit eigenvalue). Thus, we factorise, correctly choosing the λ^3 and λ^0 coefficients:

$$(\lambda - 1)(\lambda^2 + a\lambda - \frac{1}{18}) = 0.$$

If we expand this out, and compare the powers of λ^2 with the original characteristic equation, we find that $a = \frac{1}{6}$.

Therefore,

$$(\lambda - 1)(\lambda^2 + \tfrac{1}{6}\lambda - \tfrac{1}{18}) = 0,$$

which further factorises to

$$(\lambda - 1)(\lambda + \tfrac{1}{3})(\lambda - \tfrac{1}{6}) = 0.$$

So the three eigenvalues are

$$\lambda^{(1)} = 1, \quad \lambda^{(2)} = -\tfrac{1}{3}, \quad \lambda^{(3)} = \tfrac{1}{6}.$$

Now, corresponding to $\lambda^{(1)} = 1$, we know that the left eigenvector is just

$$(\chi^{(1)})^T = (1 \quad 1 \quad 1), \quad \lambda^{(1)} = 1.$$

So, to find the corresponding right eigenvector, $\psi^{(1)}$, we solve in the usual way:

$$\begin{pmatrix} 0 & 1/6 & 0 \\ 1 & 1/2 & 2/3 \\ 0 & 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 1. \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

to give

$$\psi^{(1)} = \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix}.$$

This is the stationary state; but, to correctly normalise it so that $\langle \chi^{(1)} | \psi^{(1)} \rangle = 1$, we note that

$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} = 10; \quad \psi^{(1)} = P^{st} = \begin{pmatrix} 1/10 \\ 3/5 \\ 3/10 \end{pmatrix}.$$

And thus, we have found the stationary state. The other eigenvectors are fairly easily found to be

$$\lambda^{(2)} = -\tfrac{1}{3} \quad \Rightarrow \quad \psi^{(2)} = \begin{pmatrix} 1/6 \\ -1/3 \\ 1/6 \end{pmatrix}, \quad \chi^{(2)} = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$$

and

$$\lambda^{(3)} = \frac{1}{6} \quad \Rightarrow \quad \psi^{(3)} = \begin{pmatrix} -4/15 \\ -4/15 \\ 8/15 \end{pmatrix}, \quad \chi^{(3)} = \begin{pmatrix} -3/2 \\ -1/4 \\ 1 \end{pmatrix}.$$

Remember that we get the $\chi^{(i)}$ in the form of a row vector, $(\chi^{(i)})^T$. To continue, we make the identification

$$\psi^{(i)} \longmapsto |\psi^{(i)}\rangle, \quad (\chi^{(i)})^T \longmapsto \langle\chi^{(i)}|,$$

and so, using this notation, we see orthonormality:

$$\langle\chi^{(1)}|\psi^{(3)}\rangle = 0.$$

Also note, as $\psi^{(2),(3)}$ are orthogonal to $\chi^{(1)}$, their entries must sum to zero (which they do).

General theory of eigenvectors of Q

We now prove various relations, in a very similar fashion to quantum mechanics. This theory does not rely on the matrix being stochastic; it is true for any matrix.

So suppose Q is a (time-independent) $M \times M$ matrix. It will have M eigenvalues in general. If Q is symmetric the eigenvalues will be real, otherwise they may be complex

Corresponding to the ith eigenvalue $\lambda^{(i)}$, there will be a right-eigenvector (or eigenstate) $|\psi^{(i)}\rangle$ and a left-eigenvector (or eigenstate) $\langle\chi^{(i)}|$ such that

$$Q|\psi^{(i)}\rangle = \lambda^{(i)}|\psi^{(i)}\rangle \quad \text{and} \quad \langle\chi^{(i)}|Q = \lambda^{(i)}\langle\chi^{(i)}|$$

Orthogonality

Forming the product of the above eigenvalue equations, with a bra-state on the first, and a ket-state on the second, we have that

$$\langle\chi^{(j)}|Q|\psi^{(i)}\rangle = \lambda^{(i)}\langle\chi^{(j)}|\psi^{(i)}\rangle, \quad \langle\chi^{(i)}|Q|\psi^{(j)}\rangle = \lambda^{(i)}\langle\chi^{(i)}|\psi^{(j)}\rangle.$$

Interchanging indices in the second expression gives

$$\langle \chi^{(j)} | Q | \psi^{(i)} \rangle = \lambda^{(j)} \langle \chi^{(j)} | \psi^{(i)} \rangle.$$

Subtracting the two expressions now gives

$$0 = (\lambda^{(i)} - \lambda^{(j)}) \langle \chi^{(j)} | \psi^{(i)} \rangle.$$

So, if $\lambda^{(i)} \neq \lambda^{(j)}$, then this simply reads

$$\langle \chi^{(j)} | \psi^{(i)} \rangle = 0, \quad \lambda^{(i)} \neq \lambda^{(j)}.$$

If we have chosen to normalise the eigenvectors, then

$$\langle \chi^{(j)} | \psi^{(i)} \rangle = \delta_{ij}, \tag{1}$$

which is the statement of orthonormality: right and left eigenvectors corresponding to different eigenvalues are orthonormal to each other.

Completeness

Since $\{|\psi^{(i)}\rangle\}$ are a complete set of states, we can expand any probability vector in terms of them:

$$|P\rangle = \sum_{i=1}^{M} \alpha_i |\psi^{(i)}\rangle.$$

So, forming the product of this with a bra-state gives

$$\langle\chi^{(j)}|P\rangle = \sum_{i=1}^{M} \alpha_i \langle\chi^{(j)}|\psi^{(i)}\rangle,$$

which, by our orthonormality statement, is simply

$$\langle\chi^{(j)}|P\rangle = \sum_{i=1}^{M} \alpha_i \langle\chi^{(j)}|\psi^{(i)}\rangle = \sum_{i=1}^{M} \alpha_i \delta_{ij} = \alpha_j.$$

This identifies the α_i:

$$\alpha_i = \langle\chi^{(i)}|P\rangle.$$

Then, using this in our original expansion,

$$|P\rangle = \sum_{i=1}^{M} |\psi^{(i)}\rangle \alpha_i = \left\{ \sum_{i=1}^{M} |\psi^{i}\rangle\langle\chi^{(i)}| \right\} |P\rangle.$$

But this is true for any $|P\rangle$, so that

$$\sum_{i=1}^{M} |\psi^{(i)}\rangle\langle\chi^{(i)}| = I, \qquad (2)$$

where I is the $M \times M$ identity matrix.

We note the following important relation which follows by multiplying (2) by Q:

$$\begin{aligned} Q &= \sum_{i=1}^{M} Q|\psi^{(i)}\rangle\langle\chi^{(i)}| \\ &= \sum_{i=1}^{M} \lambda^{(i)}|\psi^{(i)}\rangle\langle\chi^{(i)}| \end{aligned}$$

How does all this help us to solve $P(t) = Q^t P(0)$?

Well, if Q has eigenvalues $\lambda^{(i)}$ and eigenvectors $\langle\chi^{(i)}|$ and $|\psi^{(i)}\rangle$, then Q^t has eigenvalues $(\lambda^{(i)})^t$ and also has eigenvectors $\langle\chi^{(i)}|$ and $|\psi^{(i)}\rangle$

This is easy to prove by induction: assume Q^N has eigenvalues $(\lambda^{(i)})^N$ and eigenvectors $\langle\chi^{(i)}|$ and $|\psi^{(i)}\rangle$. Then $Q^{N+1}|\psi^{(i)}\rangle = Q.Q^N|\psi^{(i)}\rangle$ which equals $Q(\lambda^{(i)})^N|\psi^{(i)}\rangle = (\lambda^{(i)})^{N+1}|\psi^{(i)}\rangle$. So if it is true for N, it is true for $N+1$. Since it is true for $N=1$, it is true for all N. A similar proof holds for the left eigenvector.

Now just as we obtained an important relation above by multiplying (2) by Q, we can obtain a generalisation by multiplying (2) by Q^t:

$$\begin{aligned} Q^t &= \sum_{i=1}^{M} Q^t|\psi^{(i)}\rangle\langle\chi^{(i)}| \\ &= \sum_{i=1}^{M} (\lambda^{(i)})^t|\psi^{(i)}\rangle\langle\chi^{(i)}| \qquad (3) \end{aligned}$$

Equation (3) is a key result, because it shows that if we can find the eigenvalues and eigenvectors of Q, we can calculate Q^t.

Notice that this implies that

$$P(t) = Q^t P(0) = \sum_{i=1}^{M} (\lambda^{(i)})^t |\psi^{(i)}\rangle\langle\chi^{(i)}|P(0)\rangle.$$

All the quantities on the right-hand side can be found from a knowledge of Q.

Example The urn model, Exercise 5.1 from Reichl, continued.

Let us return to the urn models matrix & eigenvectors, to compute an arbitrary power of the matrix.
For $\lambda^{(1)} = 1$, we see that

$$|\psi^{(1)}\rangle\langle\chi^{(1)}| = \begin{pmatrix} 1/10 \\ 3/5 \\ 3/10 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1/10 & 1/10 & 1/10 \\ 3/5 & 3/5 & 3/5 \\ 3/10 & 3/10 & 3/10 \end{pmatrix} \equiv Q_1.$$

For $\lambda^{(2)} = -1/3$, we see that

$$|\psi^{(2)}\rangle\langle\chi^{(2)}| = \begin{pmatrix} 1/6 \\ -1/3 \\ 1/6 \end{pmatrix} (3 \quad -1 \quad 1) = \begin{pmatrix} 1/2 & -1/6 & 1/6 \\ -1 & 1/3 & -1/3 \\ 1/2 & -1/6 & 1/6 \end{pmatrix} \equiv Q_2.$$

And finally, for $\lambda^{(3)} = 1/6$, we see that

$$|\psi^{(3)}\rangle\langle\chi^{(3)}| = \begin{pmatrix} 2/5 & 1/15 & -4/15 \\ 2/5 & 1/15 & -4/15 \\ -4/5 & -2/15 & 8/15 \end{pmatrix} \equiv Q_3.$$

So, an arbitrary power of Q may be found from

$$Q^t = Q_1 + (-\tfrac{1}{3})^t Q_2 + (\tfrac{1}{6})^t Q_3.$$

Thus, we have a way of computing

$$P(t) = Q^t P(0).$$

Now, we can notice a few things from this. First, note that the smallest eigenvalue will have little effect on the late-time behaviour of the system. That is, for high t, the last term will be negligible. The next largest eigenvalue gives the dominant large t behaviour. Second, notice that if any of the eigenvalues had been > 1, then the system would have diverged. Finally, the very-large t behaviour is completely determined by Q_1.

If the initial state of the system is, for instance, $P(0)^T = (1\ 0\ 0)$, then multiplying the above expression for Q^t into this gives

$$P(t) = \begin{pmatrix} 1/10 \\ 3/5 \\ 3/10 \end{pmatrix} + \left(-\frac{1}{3}\right)^t \begin{pmatrix} 1/2 \\ -1 \\ 1/2 \end{pmatrix} + \left(\frac{1}{6}\right)^t \begin{pmatrix} 2/5 \\ 2/5 \\ -4/5 \end{pmatrix}$$

From this we can calculate, for example, the mean

$$\langle n(t) \rangle = \sum_{n=1}^{3} nP(n,t) = P(1,t) + 2P(2,t) + 3P(3,t) = \frac{11}{5} - \frac{6}{5}\left(\frac{1}{6}\right)^t$$