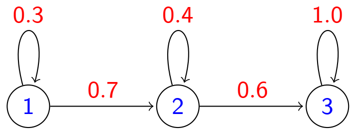
Lecture 12

Markov chains

Consider a system that can be in three different states.



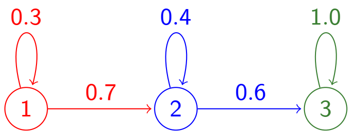
Once per second, it can change state in a random way. If it is in state 1, it jumps to state 2 with probability 0.7 and stays in state 1 with probability 0.3. If it is in state 2, it jumps to state 3 with probability 0.6 and stays in state 1 with probability 0.4. If it is in state 3, it stays there (with probability 1).

This is an example of a *Markov chain*. These are widely used to model (pseudo)-random processes in economics, population biology, information technology and other areas. Some questions about a Markov chain:

- How much time to we spend in state i, on average?
- If we start in state i, what is the average wait before reaching j?
- If we start in state i, what is the probability of reaching j before k?

We will take the first steps towards answering such questions.

Markov chains



Notation: $p_{j\leftarrow i}$ is the probability of jumping from state i to state j.
The *transition matrix* has $p_{j\leftarrow i}$ in the i'th column of the j'th row.

$$P = \begin{bmatrix} p_{1\leftarrow 1} & p_{1\leftarrow 2} & p_{1\leftarrow 3} \\ p_{2\leftarrow 1} & p_{2\leftarrow 2} & p_{2\leftarrow 3} \\ p_{3\leftarrow 1} & p_{3\leftarrow 2} & p_{3\leftarrow 3} \end{bmatrix} = \begin{bmatrix} 0.3 & 0.0 & 0.0 \\ 0.7 & 0.4 & 0.0 \\ 0.0 & 0.6 & 1.0 \end{bmatrix}.$$

All entries are probabilities so they lie between 0 and 1.
The entries in column 1 are the probabilities of all possible steps when we start in state 1, so they must add up to 1.
Similarly, each column has nonnegative entries adding up to 1, in other words it is a *probability vector*. By definition, this means that P is a *stochastic matrix*.

Distribution vectors

Suppose that the probability of being in state i (at a certain time) is q_i.
Let q'_j be the probability of being in state j one second later. Then $q'_j = \sum_i p_{j\leftarrow i} q_i$.

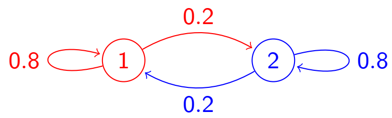
In terms of distribution vectors $q = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}^T$ and $q' = \begin{bmatrix} q'_1 & \cdots & q'_n \end{bmatrix}^T$ this says that $q' = Pq$. For example, when there are three states we have

$$q' = \begin{bmatrix} q'_1 \\ q'_2 \\ q'_3 \end{bmatrix} = \begin{bmatrix} p_{1\leftarrow 1}q_1 + p_{1\leftarrow 2}q_2 + p_{1\leftarrow 3}q_3 \\ p_{2\leftarrow 1}q_1 + p_{2\leftarrow 2}q_2 + p_{2\leftarrow 3}q_3 \\ p_{3\leftarrow 1}q_1 + p_{3\leftarrow 2}q_2 + p_{3\leftarrow 3}q_3 \end{bmatrix} = \begin{bmatrix} p_{1\leftarrow 1} & p_{1\leftarrow 2} & p_{1\leftarrow 3} \\ p_{2\leftarrow 1} & p_{2\leftarrow 2} & p_{2\leftarrow 3} \\ p_{3\leftarrow 1} & p_{3\leftarrow 2} & p_{3\leftarrow 3} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = Pq.$$

Thus, if r_t is the distribution vector at time t we have $r_t = P^t r_0$. This can be calculated using the eigenvalues and eigenvectors of P.

Markov chain example

Consider a two-state Markov chain which stays in the same state with probability 0.8, and flips to the other state with probability 0.2.



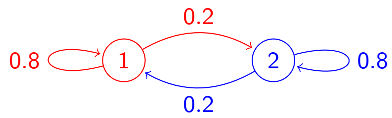
$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

The characteristic polynomial is $\chi_P(t) = t^2 - 1.6t + 0.6$ so the eigenvalues are $(1.6 \pm \sqrt{2.56 - 4 \times 0.6})/2$, which works out as $\lambda_1 = 0.6$ and $\lambda_2 = 1$.
Corresponding eigenvectors: $u_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Now $P = UDU^{-1}$ and so $P^n = UD^nU^{-1}$, where

$$U = [u_1|u_2] = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} 0.6 & 0 \\ 0 & 1 \end{bmatrix} \qquad U^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

$$P^n = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} (0.6)^n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.5(1+0.6^n) & 0.5(1-0.6^n) \\ 0.5(1-0.6^n) & 0.5(1+0.6^n) \end{bmatrix}$$

Markov chain example



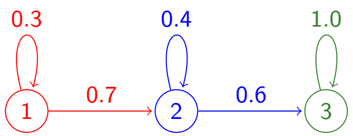
$$P^n = \begin{bmatrix} 0.5(1+0.6^n) & 0.5(1-0.6^n) \\ 0.5(1-0.6^n) & 0.5(1+0.6^n) \end{bmatrix}$$

Suppose we are given that the system starts at $t = 0$ in state 1, so $r_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. It follows that

$$r_n = P^n r_0 = \begin{bmatrix} 0.5(1+0.6^n) & 0.5(1-0.6^n) \\ 0.5(1-0.6^n) & 0.5(1+0.6^n) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5(1+0.6^n) \\ 0.5(1-0.6^n) \end{bmatrix}.$$

Thus, at time n the probability of being in state 1 is $0.5(1 + 0.6^n)$, and the probability of being in state 2 is $0.5(1 - 0.6^n)$.
When n is large, we observe that $(0.6)^n$ will be very small, so $r_n \simeq \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$, so it is almost equally probable that X will be in either of the two states. This should be intuitively plausible, given the symmetry of the situation.

Markov chain example



$$P = \begin{bmatrix} 0.3 & 0.0 & 0.0 \\ 0.7 & 0.4 & 0.0 \\ 0.0 & 0.6 & 1.0 \end{bmatrix}.$$

We start in state 1 at $t = 0$. What is the probability that we are in state 3 at $t = 5$? We are given $r_0 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$ and we need to find $r_5 = P^5 r_0$.

$$\chi_P(t) = \det \begin{bmatrix} 0.3-t & 0.0 & 0.0 \\ 0.7 & 0.4-t & 0.0 \\ 0.0 & 0.6 & 1.0-t \end{bmatrix} = (0.3-t)(0.4-t)(1-t),$$

so the eigenvalues are 0.3, 0.4 and 1.
To find an eigenvector of eigenvalue 0.3, we row-reduce the matrix $P - 0.3I$:

$$\begin{bmatrix} 0 & 0 & 0 \\ 7/10 & 1/10 & 0 \\ 0 & 6/10 & 7/10 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1/7 & 0 \\ 0 & 1 & 7/6 \end{bmatrix} \to \begin{bmatrix} 1 & 1/7 & 0 \\ 0 & 1 & 7/6 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1/6 \\ 0 & 1 & 7/6 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus take $u_1 = \begin{bmatrix} 1 & -7 & 6 \end{bmatrix}^T$ as an eigenvector of eigenvalue 0.3.
Eigenvectors u_2 and u_3 can be found similarly.

Markov chain example

$$P = \begin{bmatrix} 0.3 & 0.0 & 0.0 \\ 0.7 & 0.4 & 0.0 \\ 0.0 & 0.6 & 1.0 \end{bmatrix} \quad u_1 = \begin{bmatrix} 1 \\ -7 \\ 6 \end{bmatrix} \quad u_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \quad u_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\lambda_1 = 0.3 \qquad \lambda_2 = 0.4 \qquad \lambda_3 = 1$$

We have $P = UDU^{-1}$ where

$$D = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3) = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 1.0 \end{bmatrix} \qquad U = [u_1|u_2|u_3] = \begin{bmatrix} 1 & 0 & 0 \\ -7 & 1 & 0 \\ 6 & -1 & 1 \end{bmatrix}$$

Now find U^{-1} by row-reducing $[U|I_3]$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ -7 & 1 & 0 & 0 & 1 & 0 \\ 6 & -1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 7 & 1 & 0 \\ 0 & -1 & 1 & -6 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 7 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]$$

$$P^k = UD^kU^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -7 & 1 & 0 \\ 6 & -1 & 1 \end{bmatrix} \begin{bmatrix} (0.3)^k & 0 & 0 \\ 0 & (0.4)^k & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 7 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} (0.3)^k & 0 & 0 \\ 7(0.4)^k - 7(0.3)^k & (0.4)^k & 0 \\ 1 + 6(0.3)^k - 7(0.4)^k & 1 - (0.4)^k & 1 \end{bmatrix}.$$

Markov chain example

$$P^k = \begin{bmatrix} (0.3)^k & 0 & 0 \\ 7(0.4)^k - 7(0.3)^k & (0.4)^k & 0 \\ 1 + 6(0.3)^k - 7(0.4)^k & 1 - (0.4)^k & 1 \end{bmatrix}$$

We are definitely in state 1 at $t = 0$, so $r_0 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$. It follows that

$$r_k = P^k r_0 = \begin{bmatrix} (0.3)^k & 0 & 0 \\ 7(0.4)^k - 7(0.3)^k & (0.4)^k & 0 \\ 1 + 6(0.3)^k - 7(0.4)^k & 1 - (0.4)^k & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$
$$= \begin{bmatrix} (0.3)^k \\ 7(0.4)^k - 7(0.3)^k \\ 1 + 6(0.3)^k - 7(0.4)^k \end{bmatrix}.$$

For the probability p that X is in state 3 at time $t = 5$, we need to take $k = 5$ and look at the third component, giving

$$p = 6(0.3)^5 - 7(0.4)^5 + 1 \simeq 0.94290.$$

Stochastic matrices have eigenvalue 1

In both of the last two examples, one of the eigenvalues of the transition matrix P was equal to one. This was not a coincidence.

Proposition 17.10: If P is an $n \times n$ stochastic matrix, then 1 is an eigenvalue of P.

We will prove this after two lemmas.

A and A^T have the same eigenvalues

Lemma: Let B be an $n \times n$ matrix.
Then 0 is an eigenvalue of B iff 0 is an eigenvalue of B^T.

Proof: We can divide B and B^T into columns, say

$$B = \left[\begin{array}{c|c|c} v_1 & \cdots & v_n \end{array} \right] \qquad B^T = \left[\begin{array}{c|c|c} w_1 & \cdots & w_n \end{array} \right]$$

Now 0 is an eigenvalue of B
iff $\exists \alpha \neq 0$ with $B\alpha = 0$ or $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$
iff the columns v_i are linearly dependent
iff the v_i are not a basis (using the fact that there are n columns)
iff the w_j are not a basis (by duality)
iff the w_j are linearly dependent (using the fact that there are n columns)
iff $\exists \beta \neq 0$ with $\beta_1 w_1 + \cdots + \beta_n w_n = 0$ or $B^T \beta = 0$
iff 0 is an eigenvalue of B^T. □

A and A^T have the same eigenvalues

Corollary: For any $n \times n$ matrix A, the eigenvalues of A are the same as the eigenvalues of A^T.

Proof.
Let λ be an eigenvalue of A, so there is a nonzero vector u with $Au = \lambda u$.
This means that $(A - \lambda I_n)u = 0$, so 0 is an eigenvalue of $A - \lambda I_n$.
The lemma then tells us that 0 is also an eigenvalue of $(A - \lambda I_n)^T = A^T - \lambda I_n$.
This means that there is a nonzero vector v with $(A^T - \lambda I_n)v = 0$, or equivalently $A^T v = \lambda v$.
This proves that λ is also an eigenvalue of A^T.

The whole argument can be reversed to prove the converse as well: if λ is an eigenvalue of A^T, then it is also an eigenvalue of A. □

Stochastic matrices have eigenvalue 1

Corollary: For any $n \times n$ matrix A, the eigenvalues of A are the same as the eigenvalues of A^T.

Proposition 17.10: If P is an $n \times n$ stochastic matrix, then 1 is an eigenvalue of P.

Proof.
Let the columns of P be $v_1, \dots, v_n$.
Put $d = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \end{bmatrix}^T \in \mathbb{R}^n$.
Because P is stochastic we know that the sum of the entries in v_i is one, or in other words that $v_i.d = 1$. This means that

$$P^T d = \begin{bmatrix} v_1^T \\ \hline \vdots \\ \hline v_n^T \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} v_1.d \\ \vdots \\ v_n.d \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = d.$$

Thus, d is an eigenvector of P^T with eigenvalue 1.
It follows by the Corollary that 1 is also an eigenvalue of P, as required. □

Stationary distribution

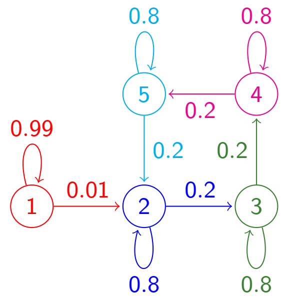
Definition 17.11: A *stationary distribution* for a Markov chain is a probability vector q that satisfies $Pq = q$ (so q is an eigenvector of eigenvalue 1).
Remark 17.12: It often happens that the distribution vectors r_n converge (as $n \to \infty$) to a distribution r_∞, whose i'th component is the long term average probability of the system being in state i. Because $Pr_n = r_{n+1}$ we then have

$$Pr_\infty = P \lim_{n\to\infty} r_n = \lim_{n\to\infty} Pr_n = \lim_{n\to\infty} r_{n+1} = r_\infty,$$

so r_∞ is a stationary distribution. Moreover, it often happens that there is only one stationary distribution. There are many theorems about this sort of thing, but we will not explore them in this course.

Stationary distribution example

We will use a heuristic argument to guess what the stationary distribution should be, and then give a rigorous proof that it is correct.
At each time there is a (small but) nonzero probability of leaving state 1 and entering the square, so if we wait long enough we can expect this to happen.
After we have entered the square there is no way we can ever return to state 1, so the long-run average probability of being in state 1 is zero.



Once we have entered the square things are symmetric so we spend $\frac{1}{4}$ of the time in each of states $2, \dots, 5$. Thus $q = \begin{bmatrix} 0 & 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}^T$ should be a stationary distribution. It is a probability vector and

$$Pq = \begin{bmatrix} 0.99 & 0 & 0 & 0 & 0 \\ 0.01 & 0.8 & 0 & 0 & 0.2 \\ 0 & 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} = q \text{ as required.}$$

Stationary distribution counterexamples

- The Markov chain on the left has transition matrix $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$.
  It follows that *every* probability vector q satisfies $Pq = q$, and so is a stationary distribution.
  In particular, there is more than one stationary distribution.
- The Markov chain on the right has transition matrix $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.
  One can check that the only stationary distribution is $q = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$.
  However, if $r_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ then
  $r_n = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ when n is even, and $r_n = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ when n is odd.
  Thus, r_n does not converge to q.