More notes on Markov Chains

First recall that a Markov chain is a set of states $i = 1, \ldots, N$, with probabilistic transitions given by the $i \to j$ transition matrix T_{ij}. A state i is recurrent if after leaving i there is always some way back to i. A state i is periodic if it is recurrent but paths return to i only at fixed multiplies of some period d (i.e., the probability of returning to i after n steps is zero unless n is a multiple of d). A Markov chain is called ergodic if it has only one recurrent class and no periodic states.

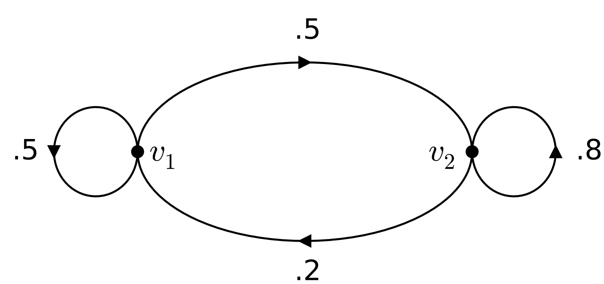
Consider the long term behavior of a Markov chain: is there some set of probabilities v_i for being in state i after some large number of steps, independent of the starting state? The answer is yes for an ergodic Markov chain, and we call v_i the stationary distribution of the chain. The v_i can also be thought of as the long term frequency of being in state i.

If the v_i are really stationary, in the sense that further steps leave them unchanged, then they satisfy the key relation

$$\sum_j v_j T_{ji} = v_i \ .$$

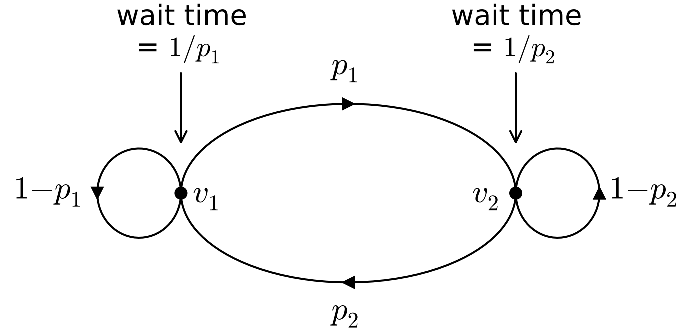
[Note that this takes the form of an eigenvalue equation $\vec{v}T = \lambda v$, with eigenvalue $\lambda = 1$. The relevant mathematical theorem (the Perron-Frobenius theorem) ensures that the matrix T has its largest eigenvalue is equal to 1, and the associated eigenvector is unique with all positive entries, so that suitably normalized it is interpretable as a set of probabilities.]

Consider for example the chain below, with two states:

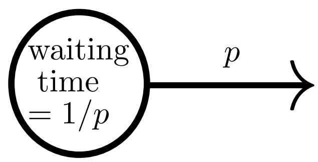


The stationary condition above becomes the two equations $.5v_1 + .8v_2 = v2$, $.2v_2 + .5v_1 = v1$, with solution $2v_2 = 5v_1$, and hence normalized as probabilities we have $v_1 = 2/7$, $v_2 = 5/7$.

More generally, for transition probabilities p_1 and p_2:



The two equations $p_1 v_1 + (1-p_2)v_2 = v_2$, $p_2 v_2 + (1-p_1)v_1 = v_1$ have solution $p_1 v_1 = p_2 v_2$, and hence normalized as probabilities we have $v_1 = p_2/(p_1+p_2)$ and $v_2 = p_1/(p_1+p_2)$. Intuitively it makes sense that in the steady state one spends more time at v_2 if the transition probability from v_1 to v_2 is greater than vice versa. The steady state probabilities depend only on the ratio p_1/p_2, but recall that the waiting times depend on their values: the expected number of steps to leave v_i is $1/p_i$, so that when the p_i are small it takes longer to equilibrate to the steady state distribution. [Equivalently the eigenvalues of the transition matrix $T = \begin{pmatrix} 1-p_1 & p_1 \\ p_2 & 1-p_2 \end{pmatrix}$ are 1 and $1-(p_1+p_2)$. The second eigenvalue determines the rate at which the steady state distribution is attained.]



If the probability of leaving a state (and never returning) is p, then the probability of leaving after exactly n steps is $(1-p)^{n-1}p$. With $q = 1-p$, the expectation value for the number of steps to leave is thus†

$$E[n] = \sum_{n=0}^{\infty} n\, q^{n-1} p = p \frac{\partial}{\partial q} \sum_{n=0}^{\infty} q^n = p \frac{\partial}{\partial q} \frac{1}{1-q} = p \frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p} \ .$$

So for example it takes on average two flips of a fair coin to get the first head, or on average six rolls of a die to get the first 3 (of course in any particular trial, sometimes more and sometimes less).

This result is important enough we'll prove it in a second way, without using the derivative as in the above. Let X be a random variable counting the number of steps it takes for a process with probability p to occur for the fist time, e.g., the number of flips of a coin to get the first H. Its expectation value is given by $E[X] = 1 \cdot \Pr[X=1] + 2 \cdot \Pr[X =$

† In the below, the infinite sum $S(q) = \sum_{n=0}^{\infty} q^n = 1 + q + q^2 + q^3 + \ldots$ satisfies the relation $qS(q) = S(q) - 1$, with solution $S(q) = 1/(1-q)$.

$2] + 3 \cdot \Pr[X = 3] + \ldots$. We can reorganize the j copies of $\Pr[X = j]$ in this expression by recalling that $\Pr[X \geq k] = \Pr[X = k] + \Pr[X = k+1] + \ldots$, and moreover that $\Pr[X \geq k]$ is just $(1-p)^{k-1}$ (the probability that it didn't happen in the first $k-1$ steps), and hence

$$\begin{aligned} E[X] &= 1 \cdot \Pr[X = 1] + 2 \cdot \Pr[X = 2] + 3 \cdot \Pr[X = 3] + \ldots \\ &= \Pr[X \geq 1] + \Pr[X \geq 2] + \Pr[X \geq 3] + \ldots \\ &= 1 + (1-p) + (1-p)^2 + \ldots = \frac{1}{1-(1-p)} = \frac{1}{p} \ . \end{aligned}$$

The expected number of steps for something to occur is sometimes called the "waiting time". We see that the "waiting time" to see the first H for a biased coin, with probability p of heads, is $1/p$. This formula also gives the waiting times indicated for the two nodes in the Markov chain in the figure at the top of the preceding page.

[As an illustration of the utility of the matrix representation, we can consider a systematic computation of the average number of steps from some start state i to any other state r. Consider constructing a new matrix Q by removing the r^{th} row and column from the transition matrix T. (In python this can be implemented as `Q=delete(delete(T, r, axis=1), r, axis=0)` .) The $(i,j)^{\text{th}}$ entry of the n^{th} power Q^n then calculates the probability of going from i to j in n steps without ever going through the state r, equivalently the expected number of times that an n step path from i will reach j without going through r. The $(i,j)^{\text{th}}$ entry of sum $S(Q) = \mathbf{1} + Q + Q^2 + Q^3 + \ldots$ thus calculates the total number of expected times that a path of any length will go from i to j without going through r ($\mathbf{1}$ is the $n-1$-dimensional identity matrix, in python represented as `eye(n-1)`). The sum of the entries along the i^{th} row of $S(Q)$ is the total expected number of times that paths of any length from i spend in states other than r, and hence gives the expected number of steps before reaching r. Finally, for Q constructed as above from a stochastic matrix T, the sum $S(Q)$ converges and can be calculated by the same simple algebra as if Q were a number: $QS(Q) = S(Q) - \mathbf{1}$, so $S(Q) = (\mathbf{1} - Q)^{-1}$. Here the inverse is the matrix inverse, i.e., M^{-1} is such that $MM^{-1} = M^{-1}M = \mathbf{1}$, and is implemented as `inv(eye(n-1)-Q)` ; so the sums along the rows, `sum(inv(eye(n-1)-Q) , axis=1)`, give a 1-dimensional array with the expected number of steps from each i before reaching the excised state r. This can be repeated successively for each r.]