Notes on Markov Chains

A Markov chain is a model of a stochastic process where the next step is determined purely by the current state. In other words, it only matters where you are, not how you got there. This is called the Markov Property and is formalised as:

$$P(X_{t+1} = s | X_t = s_t, X_{t-1} = s_{t-1}, \cdots, X_0 = s_0) = P(X_{t+1} | X_t = s_t)$$

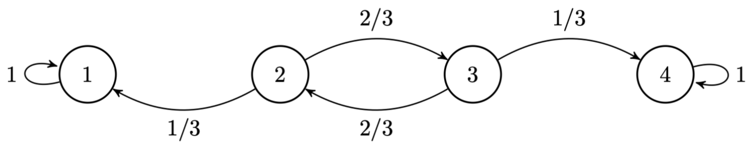
That is, the history of the process beyond the present state contains no information about the next state.

This assumption makes it easy to compute the conditional probabilities of reaching a certain state in the next step, since it only depends on the current state and not the full history of the process.

1 - Transition Matrices

Such a process is essentially a sequence of random variables $X_0, X_1, \cdots, X_n$. Each X_i is drawn based on a probability distribution over a state space $S = \{0, 1, \cdots, M\}$. Each X_t has its own probability distribution, and this distribution depends on the value of X_{t-1} took on.

These per-state probability distributions can be expressed through a transition diagram. A chain with state space $S = \{1, 2, 3, 4\}$ could look like this:



Often, we flatten them into matrix form as transition matrices Q. Just like the diagram, they show the probability of transitioning from one state into another. The row of a transition matrix is the current state i and the column is the target state j. The entry itself is then the probability that we reach state j from state i, namely $P(X_{t+1} = j | X_t = i)$. The Markov chain above has the following matrix representation:

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

A few useful things to keep in mind:

1. the vertical dimension of the matrix shows us 'now' (X_t) and the horizontal dimension shows us 'next' (X_{t+1}).
2. Because both X_t and X_{t+1} can take any value of S, we have M rows and M columns. This makes Q a square matrix.
3. As each row shows all the possible next states our Markov process could move into, the entries must sum to 1.

These matrices can be generalised to n steps, in which case Q^n would show the probability of moving from $i \to j$ over the course of n moves.

2 - Probability Distributions of X_i

As each step X_i in a Markov chain's trajectory is a random variable, it has a unique probability distribution. For a discrete state space S, we can represent it as a vector:

$$\mathbf{s}_{X_i} = (s_1, s_2, \cdots, s_M)$$

Where s_1 specifies the probability of moving to state 1, etc. The vector of course has an entry for

3 - Initial States

To simulate a Markov process, you need to specify an initial state X_0. The simple way to do this would be to assign X_0 to a certain state. However, it's often desirable to choose a random initial state based on some distribution. This distribution is analogous to the vector discussed above, where each entry is the probability of the initial state X_0 being a state i, more formally: $s_i = P(X_0 = i)$.

4 - Transient and Recurrent States

To think about how a Markov process develops, it helps to think of each state as either transient and recurrent. Early on in the process, we might spend some time in transient states, but as we increase the amount of steps, the chain will converge to spending time only in recurrent states.

A state i is recurrent if the probability of returning to that state after leaving it is 1. Otherwise, i is transient, meaning there is a positive probability of never returning to i.