1 What is a *Markov* Chain ?

A *Markov* Chain is a stochastic system in which state transition probabilities are independent of past history, i.e. if $\mathbf{X}_k$ represents the state vector at time k and $\mathbf{s}_k$ its value, then

$$\text{Prob}\left\{\mathbf{X}_{k+1} = \mathbf{s}_{k+1} | \mathbf{X}_k = \mathbf{s}_k, \mathbf{X}_{k-1} = \mathbf{s}_{k-1}, \ldots, \mathbf{X}_1 = \mathbf{s}_1, \mathbf{X}_0 = \mathbf{s}_0\right\} = \text{Prob}\left\{\mathbf{X}_{k+1} = \mathbf{s}_{k+1} | \mathbf{X}_k = \mathbf{s}_k\right\}$$

A *Markov* Chain can be discrete-time (time is essentially measured by the integers) or continous-time, finite (the number of values taken by the state vector is finite) or infinite, time-invariant (the transition probabilities don't depend on the time of transition: also called autonomous or homogenous [1]) or time varying.
The rest of this note will deal with ....

2 Discrete-time Finite Time-invariant *Markov* Chains

We introduce some notation

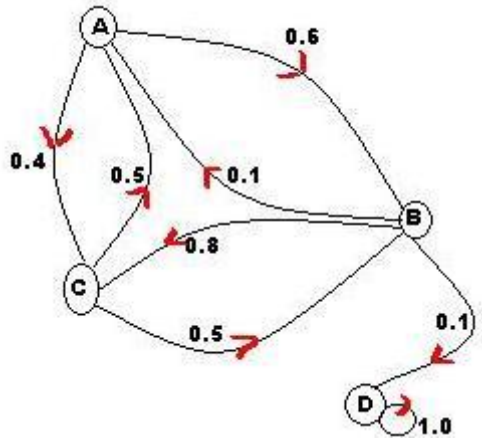
(i) the probability that the system is in state $\mathbf{s}$ at time k is given by

$$p_{\mathbf{s}}(k) = \text{Prob}\left\{\mathbf{X}_k = \mathbf{s}\right\}$$

(ii) the single time step state transition probability from state $\mathbf{s}$ to state $\mathbf{t}$ is given by

$$p_{\mathbf{st}} = \text{Prob}\left\{\mathbf{X}_{k+1} = \mathbf{t} | \mathbf{X}_k = \mathbf{s}\right\}.$$

A *Markov* Chain is typically described by graph or a table of transition probabilities. For example consider the following 4-state *Markov* Chain



next \ current	↓	A	B	C	D
	A	0.0	0.1	0.5	0.0
	B	0.6	0.0	0.5	0.0
	C	0.4	0.8	0.0	0.0
	D	0.0	0.1	0.0	1.0

Figure 1: A 4-state Markov Chain

Both representations contain the same information: the single step transition probabilities. Note that since the next state must be one of the system's states, the sum of exiting probabilities from any one of the current states must add to 1; this corresponds to the sum of the weights on the exiting edges from any node

[1] Homogenous is the term preferred by the "*Markov* Chain " community. Note that it has a meaning different to its use in solving linear recurrences

of the graph must be 1, and similarly the sum of entries in any column of the corresponding table must be 1.

From the representation, it is clear that the state probabilities evolve according to

$$\begin{pmatrix} p_A(k+1) \\ p_B(k+1) \\ p_C(k+1) \\ p_D(k+1) \end{pmatrix} = \begin{pmatrix} 0.0 & 0.1 & 0.5 & 0.0 \\ 0.6 & 0.0 & 0.5 & 0.0 \\ 0.4 & 0.8 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.0 & 1.0 \end{pmatrix} \begin{pmatrix} p_A(k) \\ p_B(k) \\ p_C(k) \\ p_D(k) \end{pmatrix}$$

In addition, we have the *consistency condition* that at any time the system must be in one or other of the system's states i.e.

$$p_A(k) + p_B(k) + p_C(k) + p_D(k) = 1$$

More generally, for a n-state *Markov* Chain with state set $\{A_1, A_2, \ldots A_n\}$, we can define

$$\mathbf{V}_k = \begin{pmatrix} p_{A_1}(k) \\ p_{A_2}(k) \\ \vdots \\ p_{A_n}(k) \end{pmatrix}$$

and describe the evolution of state probabilities by

$$\mathbf{V}_{k+1} = M\mathbf{V}_k$$

where the entries of the *transition* matrix M are given by $M_{ij} = p_{A_i,A_j}$. To determine the complete solution of this linear recurrence, the initial probability distribution $\mathbf{V}_0$ must be specified. Then

$$\mathbf{V}_k = M^k\mathbf{V}_0$$

In addition, the system satisfies the consistency condition

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \mathbf{V}_k = 1$$

3 Absorbing *Markov* Chains

A state $\mathbf{s}$ of a *Markov* Chain is called *absorbing* if it is impossible to leave it, i.e. $p_{\mathbf{s},\mathbf{s}} = 1$. A *Markov* Chain is called *absorbing* if it has at least one absorbing state and if from every state it is possible to go to an absorbing state in a finite number of steps. In an absorbing *Markov* Chain , a state which is not absorbing is called *transient*.

Absorbing *Markov* Chains are a natural context for stochastic algorithms where the algorithm eventually stops...or enters an absorbing state. The example in Section 2 is an absorbing *Markov* Chain , state D is absorbing while states A, B and C are transient.

Absorbing *Markov* Chains have a canonical form. Consider an absorbing *Markov* Chain with with q transient states and $p = n - q$ absorbing ones. Without loss of generality, we can partition the states into transient states $Tr = \{A_1, A_2, \ldots A_q\}$ and absorbing states $Ab = \{A_{q+1}, A_{q+2}, \ldots A_n\}$ with the respective probability vectors $\mathbf{V}_k^{Tr}$ and $\mathbf{V}_k^{Ab}$. Then the evolution of probabilities is governed by

$$\begin{pmatrix} \mathbf{V}_{k+1}^{Tr} \\ \mathbf{V}_{k+1}^{Ab} \end{pmatrix} = \begin{pmatrix} Q & 0 \\ \mathbf{r} & I \end{pmatrix} \begin{pmatrix} \mathbf{V}_k^{Tr} \\ \mathbf{V}_k^{Ab} \end{pmatrix}, \qquad \begin{pmatrix} \mathbf{V}_0^{Tr} \\ \mathbf{V}_0^{Ab} \end{pmatrix} = \mathbf{V}_0$$

In the context of stochastic algorithms, typically $\mathbf{V}_0^{Ab} = 0$, and we'll assume this below.

We can decompose and solve the recurrence as (i)

$$\begin{aligned} & \mathbf{V}_{k+1}^{Tr} &=& Q\mathbf{V}_k^{Tr} \\ \Rightarrow \quad & \mathbf{V}_k^{Tr} &=& Q^k\mathbf{V}_0^{Tr} \end{aligned}$$

and (ii)

$$\begin{aligned} & \mathbf{V}_{k+1}^{Ab} &=& \mathbf{V}_k^{Ab} + \mathbf{r}\mathbf{V}_k^{Tr} \\ \Rightarrow \quad & \mathbf{V}_k^{Ab} &=& \mathbf{V}_0^{Ab} + \sum_{j=0}^{k-1} \mathbf{r}\mathbf{V}_j^{Tr} \\ & &=& \mathbf{r}\left(\sum_{j=0}^{k-1} Q^j\right)\mathbf{V}_0^{Tr} \\ & &=& \mathbf{r}\left(I + Q + Q^2 + \cdots + Q^{k-1}\right)\mathbf{V}_0^{Tr} \\ & &=& \mathbf{r}(I-Q)^{-1}(I-Q^k)\mathbf{V}_0^{Tr} \end{aligned}$$

(This last expression above is identical with that for the sum of a (scalar) geometric series. To see that $I - Q$ is invertible, we first note that as $k \to \infty$, $Q^k \to 0$. Then consider

$$\begin{aligned} (I-Q)x &= 0 \\ \Rightarrow x &= Qx \\ \Rightarrow x &= Q(Qx) = Q^2x \\ \Rightarrow x &= Q^kx \\ \Rightarrow (\text{as } k \to \infty)\ x &\to 0 \end{aligned}$$

Thus $x = 0$ is the only solution of $(I - Q)x = 0$ which implies that $I - Q$ is invertible.)
Letting $k \to \infty$ in the expressions for the state probabilities gives (i) $\mathbf{V}_k^{Tr} \to 0$ and (ii) $\mathbf{V}_k^{Ab} \to \mathbf{r}(I-Q)^{-1}\mathbf{V}_0^{Tr}$. This last expression must be a vector whose entries add to one, since with probability 1 the system reaches an absorbing state. Indeed if the system has only 1 absorbing state, then $\mathbf{r}$ is a row vector and hence $\mathbf{r}(I-Q)^{-1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}$.

Consider the random variable T which measures the time required for first arrival at an absorbing state. This occurs at time k provided that the system was in a transient state at time $k-1$ and made the required transition. Thus

$$\begin{aligned} \text{Prob}\{T=k\} &= \mathbf{r}\mathbf{V}_{k-1}^{Tr} \\ &= \mathbf{r}Q^{k-1}\mathbf{V}_0^{Tr} \end{aligned}$$

The expected value of this random variable is given by

$$\begin{aligned} E(T) &= \sum_{k=0}^{\infty} k\ \text{Prob}\{T=k\} \\ &= \sum_{k=0}^{\infty} k\ \mathbf{r}Q^{k-1}\mathbf{V}_0^{Tr} \\ &= \mathbf{r}\left(\sum_{k=0}^{\infty} kQ^{k-1}\right)\mathbf{V}_0^{Tr} \\ &= \mathbf{r}(I-Q)^{-2}\mathbf{V}_0^{Tr} \end{aligned}$$

This last is best evaluated as

$$E(T) = \left[\mathbf{r}(I-Q)^{-1}\right]\left[(I-Q)^{-1}\mathbf{V}_0^{Tr}\right]$$

(In order to see where the last identity comes from, consider the (scalar) geometric series identity: when $|x| < 1$,

$$\frac{1}{1-x} = (1-x)^{-1} = \sum_{k=0}^{\infty} x^k$$

Differentiate both sides with respect to x

$$\frac{1}{(1-x)^2} = (1-x)^{-2} = \sum_{k=0}^{\infty} k x^{k-1}$$

This identity is the basis for the matrix identity used above which holds provided all eigenvalues of the matrix have magnitude less than 1. In the absorbing *Markov* Chain scenario above, this is indeed the case for Q.)

Example: Consider the 4-state *Markov* Chain with 1 absorbing state introduced in Section 2. Let's assume that we are told that the system starts in state C. For this system

$$\begin{aligned}
Q &= \begin{pmatrix} 0.0 & 0.1 & 0.5 \\ 0.6 & 0.0 & 0.5 \\ 0.4 & 0.8 & 0.0 \end{pmatrix} \qquad \mathbf{r} = \begin{pmatrix} 0.0 & 0.1 & 0.0 \end{pmatrix} \qquad \mathbf{V}_0^{Tr} = \begin{pmatrix} 0.0 \\ 0.0 \\ 1.0 \end{pmatrix} \\
\Rightarrow I - Q &= \begin{pmatrix} 1.0 & -0.1 & -0.5 \\ -0.6 & 1.0 & -0.5 \\ -0.4 & -0.8 & 1.0 \end{pmatrix} \\
\Rightarrow (I-Q)^{-1} &= \begin{pmatrix} 7.5 & 6.25 & 6.875 \\ 10.0 & 10.0 & 10.0 \\ 11.0 & 10.5 & 11.75 \end{pmatrix} \\
\text{check} \quad \mathbf{r}(I-Q)^{-1} &= \begin{pmatrix} 1.0 & 1.0 & 1.0 \end{pmatrix} \\
\text{and} \quad \mathbf{r}(I-Q)^{-1}\mathbf{V}_0^{Tr} &= 1 \\
\text{Now} \quad \mathbf{r}(I-Q)^{-2}\mathbf{V}_0^{Tr} &= \left[\mathbf{r}(I-Q)^{-1}\right]\left[(I-Q)^{-1}\mathbf{V}_0^{Tr}\right] \\
&= \begin{pmatrix} 1.0 & 1.0 & 1.0 \end{pmatrix} \begin{pmatrix} 6.875 \\ 10.0 \\ 11.75 \end{pmatrix} = 28.625
\end{aligned}$$

Thus starting in state C, the systems requires 28.625 state transitions on average to reach state D.