MARKOV CHAINS

by Peter Mitchell

website http://meikleriggs.co.uk
e-mail pm@meikleriggs.co.uk

Notes prepared for the MEI Conference 2003 in the University of Nottingham; July, 2003.
Revised for Conferences 2004, 2006 in Reading.

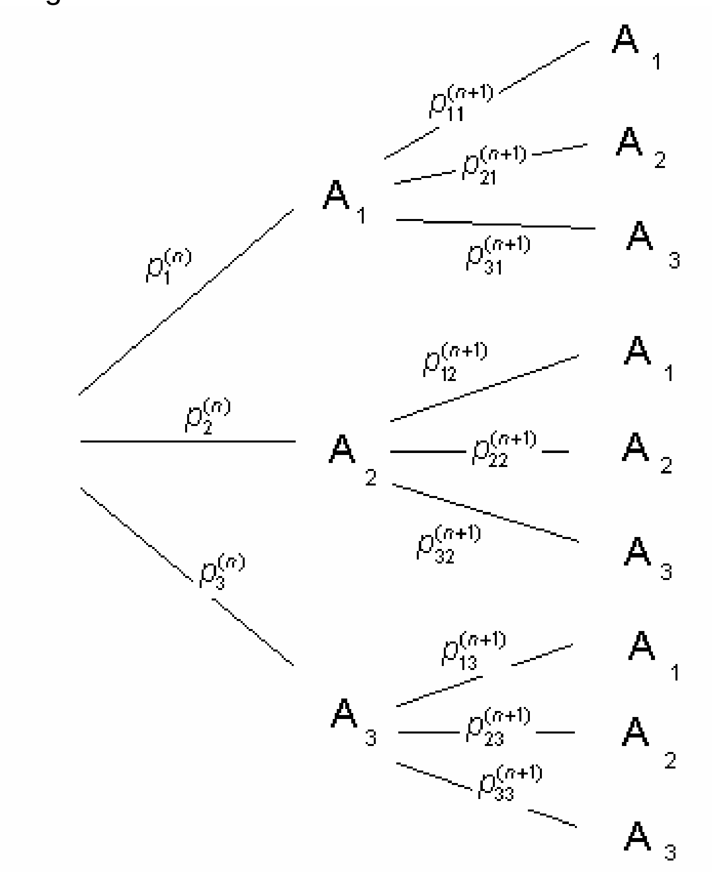
Version 4

What is a Markov process?

The weather forecast for tomorrow is certainly related to the weather today; on the other hand it is not nearly so dependent on the weather 3 months ago. This idea of dependence on previous events is fundamental to the idea of Markov process.

A Markov process is a series of experiments with an identical set of outcomes where the probability distribution of the latest set of outcomes is dependent only on the immediately previous outcome of the experiment.

Suppose the experiment has outcomes A_1, A_2, A_3 – only 3 to make the diagram less cumbersome!



This diagram shows the transition probabilities $p_{ij}^{(n+1)}$ from the nth experiment to the (n+1)th. In a Markov process, ***these probabilities are constant*****,** ie there is no need for the superfices (n+1).

Note that the suffices on the transition probabilities need to be read from right to left. The reason for this (perhaps) counter-intuitive sense will become apparent later!

The row vector $\begin{bmatrix} p_1^{(n)} \\ p_2^{(n)} \\ p_3^{(n)} \end{bmatrix} = \underline{p}^{(n)}$, say, is called the state vector. Using the usual rules for a probability tree,

$$p_m^{(n+1)} = p_{m1}p_1^{(n)} + p_{m2}p_2^{(n)} + p_{m3}p_3^{(n)}$$

$$\begin{bmatrix} p_1^{(n+1)} \\ p_2^{(n+1)} \\ p_3^{(n+1)} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} p_1^{(n)} \\ p_2^{(n)} \\ p_3^{(n)} \end{bmatrix}$$

$$\Rightarrow \quad \underline{p}^{(n+1)} = M\underline{p}^{(n)}$$

where M is the transition matrix. [Some books use the notation $\underline{p}^{(n+1)} = \underline{p}^{(n)}M$, where $\underline{p}^{(n)}$, $\underline{p}^{(n+1)}$ are *row* vectors, but MEI - from 2004 on! - uses the column vector approach.] Now the perversity in the suffix notation is explained by the need to conform to standard matrix notation!

Since we have

$$\begin{aligned} \underline{p}^{(n)} &= M\underline{p}^{(n-1)} \\ &= M^2\underline{p}^{(n-2)} \\ &\cdots\cdots \\ &= M^n\underline{p}^{(0)} \end{aligned}$$

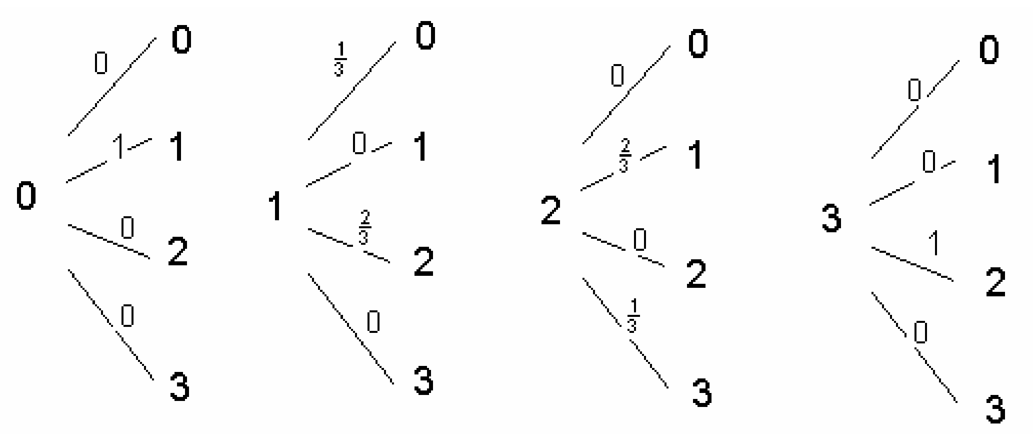
there is interest in matrix products studied elsewhere in module PM6. Hence too a Markov process can be subdivided into blocks of events within the series, each of which is itself a Markov process. If the block has length k, then the transition matrix is M^k.

Example A bag contains 3 balls each of which is black or white. A trial consists of drawing a ball and replacing it by a ball of the opposite colour. Is either of the following a Markov process? If so determine its transition matrix. (a) The colour of the ball drawn is recorded each time; (b) the number of black balls in the bag is recorded each time.

Solution

(a) There are two states B and W (with obvious notation). Suppose the current state is B, ie we drew a black (and replaced it by a white) last time. Knowing this is insufficient to determine the probability of B next time. We need to know also how many of each are currently in the bag, which in turn depends on the previous history of the experiment. Thus this is *not* a Markov process.

(b) There are four states characterised by the number of black balls, 0, 1, 2 or 3. The probabilities of transition are



$$M=\begin{bmatrix}0 & \tfrac{1}{3} & 0 & 0\\ 1 & 0 & \tfrac{2}{3} & 0\\ 0 & \tfrac{2}{3} & 0 & 1\\ 0 & 0 & \tfrac{1}{3} & 0\end{bmatrix}$$

Note that the column sums of the transition matrix are in each case 1; as is also the column sum of the state vector. This property is the defining characteristic of a ***stochastic matrix***.

For interest,

$$M^2=\begin{bmatrix}\tfrac{1}{3} & 0 & \tfrac{2}{9} & 0\\ 0 & \tfrac{7}{9} & 0 & \tfrac{2}{3}\\ \tfrac{2}{3} & 0 & \tfrac{7}{9} & 0\\ 0 & \tfrac{2}{9} & 0 & \tfrac{1}{3}\end{bmatrix}$$

The product of two stochastic matrices is again stochastic (as M^2 illustrates but does not prove).

Let the p x q stochastic matrix $A=\left[a_{ij}\right]$ where $\sum_{i=1}^{p} a_{ij}=1$, and the q x n stochastic matrix $B=\left[b_{jk}\right]$ where $\sum_{j=1}^{q} b_{jk}=1$. Then $AB=\left[\sum_{j=1}^{q} a_{ij}b_{jk}\right]$. Now consider

$$\sum_{i=1}^{p}\sum_{j=1}^{q} a_{ij}b_{jk}=\sum_{j=1}^{q} b_{jk}\sum_{i=1}^{p} a_{ij} \quad \text{(since the sum is finite)}$$

$$=\sum_{j=1}^{q} b_{jk}.1 \quad \text{(since a column sum is 1)}$$

$$=1$$

This establishes the result.

Consider now

$$M^3 = \begin{bmatrix} 0 & 7/27 & 0 & 2/9 \\ 7/9 & 0 & 20/27 & 0 \\ 0 & 20/27 & 0 & 7/9 \\ 2/9 & 0 & 7/27 & 0 \end{bmatrix}$$

$$M^4 = \begin{bmatrix} 7/27 & 0 & 20/81 & 0 \\ 0 & 61/81 & 0 & 20/27 \\ 20/27 & 0 & 61/81 & 0 \\ 0 & 20/81 & 0 & 7/27 \end{bmatrix}$$

$$M^5 = \begin{bmatrix} 0 & 61/243 & 0 & 20/81 \\ 61/81 & 0 & 182/243 & 0 \\ 0 & 182/243 & 0 & 61/81 \\ 20/81 & 0 & 61/243 & 0 \end{bmatrix}$$

$$M^6 = \begin{bmatrix} 61/243 & 0 & 487/729 & 0 \\ 0 & 487/729 & 0 & 182/243 \\ 182/243 & 0 & 242/729 & 0 \\ 0 & 242/729 & 0 & 61/243 \end{bmatrix}$$

$$M^{2n} \overset{?}{=} \begin{bmatrix} 1/4 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 3/4 \\ 3/4 & 0 & 1/3 & 0 \\ 0 & 1/3 & 0 & 1/4 \end{bmatrix}$$

Two questions now arise: is there a limit for $\underline{p}^{(n)}$?; is there a limit for M^n? They are clearly connected, since $\underline{p}^{(n)} = M^n \underline{p}^{(0)}$, so

$M^n \to M_\infty$ *as* $n \to \infty \quad \Rightarrow \quad \underline{p}^{(n)} \to \underline{p} = M_\infty \underline{p}^{(0)}$ *as* $n \to \infty$. Is it possible that $\underline{p}$ is independent of $\underline{p}^{(0)}$?

Consider the 2 x 2 stochastic matrix $S = \begin{bmatrix} a & b \\ 1-a & 1-b \end{bmatrix}$ (where $0 \le a \le 1, 0 \le b \le 1$. Then consider the characteristic equation

$$\begin{vmatrix} a-l & b \\ 1-a & 1-b-l \end{vmatrix} = 0$$

$$\Rightarrow \quad (a-l)(1-b-l) - (1-a)b = 0$$

$$\Rightarrow a - l - ab - al + lb + l^2 - b + ab = 0$$

$$\Rightarrow \quad l^2 - l(1+a-b) - b + a = 0$$

$$\Rightarrow \quad (l-1)(l-a+b) = 0$$

$$\Rightarrow \quad l = 1 \quad \text{or} \quad a-b$$

Hence in every case a 2 x 2 stochastic matrix has 1 as an eigenvalue. To find the eigenvector which corresponds see next page.

$$S\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\Rightarrow \begin{matrix} ax + by = x \\ (1-a)x + (1-b)y = y \end{matrix}$$

$$\Rightarrow \quad (a-1)x + by = 0$$

in each case.

If $a = 1$, then y=0 and the eigenvector is $\begin{bmatrix} l \\ 0 \end{bmatrix}$; case 1.

If $b = 0$, then x=0 and the eigenvector is $\begin{bmatrix} 0 \\ l \end{bmatrix}$; case 2.

If $a = 1$ and $b = 0$, then any vector is an eigenvector; case 3.

If $a \neq 1$ and $b \neq 0$, then we have $y = \frac{(1-a)}{b}x$ so an eigenvector is $l\begin{bmatrix} 1 \\ \frac{1-a}{b} \end{bmatrix}$; case 4.

In case1, the corresponding probability vector is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (so that the column sum is 1); in case 2, $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$; and in case 3 any $\begin{bmatrix} l \\ 1-l \end{bmatrix}$ for $0 \leq l \leq 1$. These are trivial cases, for $a = 1$ corresponds to one absorbing state, $b = 0$ to another. We focus on case 4, ie we exclude both $a = 1$ and $b = 0$.

Now the eigenvector which corresponds to the eigenvalue 1 is $l\begin{bmatrix} 1 \\ \frac{1-a}{b} \end{bmatrix}$. For a stochastic vector we must have

$$l(1 + \tfrac{1-a}{b}) = 1$$

$$\Rightarrow l(b + 1 - a) = b$$

$$\Rightarrow l = \tfrac{b}{b-a+1}$$

provided that $a - b \neq 1 \Leftrightarrow a \neq 1, b \neq 0$. Then we have the stochastic vector

$$l\begin{bmatrix} 1 \\ \frac{1-a}{b} \end{bmatrix} = \tfrac{b}{b-a+1}\begin{bmatrix} 1 \\ \frac{1-a}{b} \end{bmatrix} = \underline{u}_1, \text{ say.}$$

Now let any $\underline{p} = \underline{u}_1 + e\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. This ensures that $\frac{b}{b-a+1} + \frac{1-a}{b-a+1} + e - e = 1$ still.

Then

$$S = \underline{u}_1 + e\begin{bmatrix} a-b \\ a-b \end{bmatrix}$$

$$= \underline{u}_1 + e(a-b)\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

We have that $|(a-b)| < 1$ unless $a = 1, b = 0$ or $a = 0, b = 1$

$$\Rightarrow S^n \underline{p} = \underline{u}_1 + e(a-b)^n \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$\to \underline{u}_1$$

as $n \to \infty$. Thus in the 2 x 2 case, when the trivial cases $a = 1, b = 0$ and $a = 0, b = 1$ are excluded, it is always the case that $\underline{p}^{(n)} \to \underline{u}_1$ as $n \to \infty$ so *there is a limiting state vector and it is independent of the initial state.*

Now we investigate our matrix *M.*

$$|M - lI| = 0$$
$$\Rightarrow \begin{vmatrix} -l & \frac{1}{3} & 0 & 0 \\ 1 & -l & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & -l & 1 \\ 0 & 0 & \frac{1}{3} & -l \end{vmatrix} = 0$$
$$\Rightarrow \begin{vmatrix} -l & 1 & 0 & 0 \\ 1 & -3l & 2 & 0 \\ 0 & 2 & -3l & 1 \\ 0 & 0 & 1 & -l \end{vmatrix} = 0$$
$$\Rightarrow -l \begin{vmatrix} -3l & 2 & 0 \\ 2 & -3l & 1 \\ 0 & 1 & -l \end{vmatrix} - \begin{vmatrix} 1 & 2 & 0 \\ 0 & -3l & 1 \\ 0 & 1 & -l \end{vmatrix} = 0$$
$$\Rightarrow -l(-3l)(3l^2 - 1) + 4l(-l) - 1(3l^2 - 1) = 0$$
$$\Rightarrow 9l^4 - 3l^2 - 4l^2 - 3l^2 + 1 = 0$$
$$\Rightarrow 9l^4 - 10l^2 + 1 = 0$$
$$\Rightarrow (l^2 - 1)(9l^2 - 1) = 0$$
$$\Rightarrow l = \pm 1, \pm \tfrac{1}{3}$$

What about M^2?

$$|M^2 - lI| = 0$$
$$\Rightarrow 0 = \begin{vmatrix} \frac{1}{3} - l & 0 & \frac{2}{9} & 0 \\ 0 & \frac{7}{9} - l & 0 & \frac{2}{3} \\ \frac{2}{3} & 0 & \frac{7}{9} - l & 0 \\ 0 & \frac{2}{9} & 0 & \frac{1}{3} - l \end{vmatrix}$$
$$\Rightarrow 0 = \begin{vmatrix} 1 - 3l & 0 & 2 & 0 \\ 0 & 7 - 9l & 0 & 2 \\ 2 & 0 & 7 - 9l & 0 \\ 0 & 2 & 0 & 1 - 3l \end{vmatrix}$$

Now, to save effort, if we wish to satisfy ourselves that $l = 1$ is a root in this case too, it will suffice to examine the last determinant with $l = 1$. Thus we consider

$$\begin{vmatrix} -2 & 0 & 2 & 0 \\ 0 & -2 & 0 & 2 \\ 2 & 0 & -2 & 0 \\ 0 & 2 & 0 & -2 \end{vmatrix}$$

The third row is a multiple of the first so this determinant is zero as expected.

Now we consider the general situation (with stochastic matrix *T*) such that $T^n \underline{p} \to \underline{q}$ as $n \to \infty$. Then $T^{n+1}\underline{p} \to T\underline{q}$ as $n \to \infty$. Hence $T\underline{q} = \underline{q}$ and so $\underline{q}$ is an eigenvector corresponding to the eigenvalue 1 and any limit vector must be invariant. In the 2 x 2 case, apart from the exceptional cases, such an invariant vector certainly exists, and we have shown it is a limit vector.

Any stochastic matrix has 1 as an eigenvalue. Let *N* be a stochastic matrix. Then by definition $\underline{u}'N = \underline{u}'$, where here $\underline{u}'$ denotes a *row* of 1's; we use a prime to indicate that we are considering a transposed *column* vector $\underline{u}$. Then we have

$$N'\underline{u} = \underline{u}$$
$$\Rightarrow (N'-I)\underline{u} = \underline{0}$$
$$\Rightarrow \quad |N'-I| = 0$$
$$\Rightarrow \quad |N-I| = 0$$

so 1 is an eigenvalue of N.

A significant omission in this development (I now realise) is the demonstration that the corresponding eigenvector can be a probability vector. This boils down to demonstrating that in such an eigevector all the elements have the same sign, since a multiplier can be found then so that the elements are all positive and add up to 1. This insertion must await a fifth edition!

The questions posed so far relate to the limiting form of T^n and the limiting form of $\underline{p}^{(n)}$, and as to whether the form of lim $\underline{p}^{(n)}$ is independent of the initial state. In the 4 x 4 case considered, no limit of M^n exists, since the form of the matrix clearly oscillates, but we suspect nonetheless a limiting form of M^{2n}. We now establish conditions in general sufficient for lim T^n to exist, and consider its consequence.

An n x n matrix T is said to be regular if, for some k, T^k has at least one column with no zeros. A regular matrix is therefore such that there is at least one state which is accessible from any starting state. Note that the 4-state example considered is not regular; neither is M^{2n} in that case.

Theorem (which may be extended to larger square matrices) If T is an n x n regular matrix, there is an unique invariant probability vector $\underline{t}$, and $\underline{t}$ = lim $T^n\underline{q}$ for any probability vector $\underline{q}$.

Proof

We show that lim T^n is a matrix whose columns are all identical.

Assume for a moment that this is the case, and call that limiting column $\underline{t}$.
$T^n \to [\underline{t} \quad \underline{t} \quad \underline{t}] \Rightarrow [\underline{t} \quad \underline{t} \quad \underline{t}]\underline{q} = \underline{t}$, if $\underline{q}$ is a probability vector ($q_1 + q_2 + q_3 = 1 \Rightarrow q_1t_1 + q_2t_1 + q_3t_1 = t_1$, etc). The invariance of $\underline{t}$ follows from taking $\underline{t}$=$\underline{q}$.

Let A=T^k have a column of non-zero entries. Take it is as the first (by reordering the states if necessary). Let $e > 0$ be the smallest entry in that column. We illustrate the argument in the 3x3 case, but it is perfectly general. Let

$$A^n = \begin{bmatrix} u_1 & u_2 & u_3 \\ & & \\ & & \end{bmatrix}, A^{n+1} = \begin{bmatrix} v_1 & v_2 & v_3 \\ & & \\ & & \end{bmatrix} \text{ and } A = \begin{bmatrix} a_1 & & \\ a_2 & & \\ a_3 & & \end{bmatrix}.$$

Then we have $a_1 + a_2 + a_3 = 1$ and

$$\begin{aligned} v_1 &= a_1u_1 + a_2u_2 + a_3u_3, \quad \text{etc} \\ &\geq a_1u_1 + (a_2 + a_3)u_{\min}, \quad \textit{where } u_{\min} = \min\{u_1, u_2, u_3\} \\ &\geq a_1u_1 + (1 - a_1)u_{\min} \\ &\geq eu_1 + (1 - e)u_{\min}, \text{ since } a_1(u_1 - u_{\min}) \geq e(u_1 - u_{\min}) \end{aligned}$$

and similarly

$$v_1 \leq eu_1 + (1 - e)u_{\max}$$

By identical reasoning, there are corresponding inequations satisfied by v_2, v_3. So we have

$$\begin{aligned} v_{\min} &\geq eu_1 + (1 - e)u_{\min} \\ v_{\max} &\leq eu_1 + (1 - e)u_{\max} \\ \Rightarrow -v_{\min} &\leq -eu_1 - (1 - e)u_{\min} \\ \Rightarrow v_{\max} - v_{\min} &\leq (1 - e)(u_{\max} - u_{\max}) \end{aligned}$$

This means the differences between the elements of the first row tend to zero as n increases. The same argument works for each row. Hence

$\lim\limits_{n \to \infty} A^n = [\underline{t} \quad \underline{t} \quad \underline{t}]$ where $\underline{t}$ is a probability vector. Now we have for any n by the division algorithm

$$\begin{aligned} n &= r + kq \quad (0 \leq r < k) \\ \Rightarrow T^n &= T^r (T^k)^q \\ &= T^r A^q \end{aligned}$$

so

$$\begin{aligned} \lim_{q \to \infty} T^n &= \lim(T^r A^q) \\ &= \lim(T^r)\lim(A^q) \\ &= T^r [\underline{t} \quad \underline{t} \quad \underline{t}] \\ &= [\underline{t} \quad \underline{t} \quad \underline{t}] \end{aligned}$$

since T is a stochastic matrix. Done!