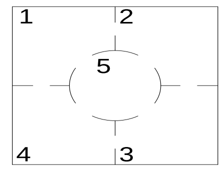
A Markov Chain Example



A Markov chain example. In an experiment you are placed in a five room "house" (see Figure 3). Every hour the doors are opened and you must move from your current room to one of the adjacent rooms. Assuming the rooms are all equally attractive, what percentage of the time will you spend in each room? (The extent to which the experimental percentage differs from this measures the desirability of each room).

To solve this problem one introduces the 5×5 *transition matrix* $M=(m_{ij})$ of this *Markov* [1856–1922] *chain*: if you are currently in room j, then m_{ij} is the probability you will next be in room i (CAUTION: some mathematicians interchange the roles of i and j). For this, we number the rooms, say clockwise beginning in the upper left corner with p_5 referring to the center room. Then, for instance, $m_{12}=m_{32}=m_{52}=\frac{1}{3}$ since if you are in room 2, it is equally likely that you will next be in rooms 1, 3, or 5, but you won't be in rooms 2 or 4. Proceeding similarly we obtain

$$M=\begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

The elements of M are non-negative and the sum of every column is 1: no matter where you are now, at the next step you will certainly be in one of the rooms.

It is useful to introduce column *probability column vectors* $P=(p_1,\ldots,p_5)$ with the property that p_j gives the probability of being in the j^{th} room at a given time. Then $0\le p_j\le 1$ and $\sum p_j=1$. If P_{now} describes the probabilities of your current location, then $P_{\text{next}}=MP_{\text{now}}$, gives the probabilities of your location at the next time interval. Thus, if one begins in Room 1, then $P_0=(1,0,0,0,0)$, and after the first hour $P_1=(0,\frac{1}{3},0,\frac{1}{3},\frac{1}{3})=MP_0$. In the same way, at the end of the second hour $P_2:=MP_1=M^2P_0$, and $P_k:=MP_{k-1}=M^kP_0$.

For a matrix M arising in a Markov process (non-negative elements and the sum of each column is one), if λ is any eigenvalue of M^* (and hence M), then $|\lambda|\le 1$. To see this, let $v:=(v_1,\ldots,v_n)$ be a corresponding eigenvector, $M^*v=\lambda v$, with largest component v_k, that is, $|v_i|\le|v_k|$. Then $|(\lambda-m_{kk})v_k|=|\sum_{i\ne k}m_{ik}v_i|\le(\sum_{i\ne k}m_{ik})|v_k|$. Since $\sum_i m_{ik}=1$ then $|\lambda-m_{kk}|\le 1-m_{kk}$. Consequently $|\lambda|\le|\lambda-m_{kk}|+m_{kk}\le 1$ (this reasoning is a special case of Gershgorin's theorem).

Moreover, if we assume all the elements of M are positive, then equality $|\lambda|=1$ occurs only if $\lambda=1$ and $v_1=v_2=\ldots=v_n$. Thus $|\lambda|<1$ except for the one dimensional eigenspace corresponding to $\lambda=1$.

In seeking the long-term probabilities, we are asking if the probability vectors $P_k = M^k P_0$, $k = 1, 2, \ldots$ converge to some "equilibrium" vector P *independent* of the initial probability vector P_0. If so, then in particular $P = \lim M^{k+1} P_0 = \lim M M^k P_0 = MP$, that is, $P = MP$ so P is an eigenvector of M with eigenvalue 1. Moreover, choosing P_0 to be *any* standard basis vector e_j and since the j^{th} column of M^n is $M^n e_j \to P$, it follows that $M^k \to M_\infty$ where all the columns of M_∞ are the *same* eigenvector P. In addition, still assuming convergence to equilibrium, every eigenvector of M with eigenvalue $\lambda = 1$ must be a multiple of P.

Although $\lambda = 1$ is always an eigenvalue of M (since it is an eigenvalue of M^* with eigenvector $(1, \ldots, 1)$), the limit $M^k P_0$ does not always exist. For example, it does not exist for the transition matrix $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ for a two room "house." If $M = I$, then the limit of $M^k P_0$ exists but is not independent of P_0. However the limit $M^k P_0$ does exist and is independent of the initial probability vector P_0 if all of the elements of M —or some power of M—are positive. In this case M is called the transition matrix of a *regular* Markov Chain.

If M is diagonalizable, the convergence of $M^k P_0$ to a limiting vector independent of P_0 follows from the above information on its eigenvalues. For the general case one must work harder. At the end of these notes I'll give the simplest proof I know that does not assume M is diagonalizable. In our case all the elements of M^2 are positive since after two steps there is a positive probability that one will be in each of the rooms. It remains to find this limiting probability distribution P by solving $P = MP$.

Here is where we can use symmetry. Since the four corner rooms are identical, M must commute with the matrices T_{ij} that interchange the probabilities of being in the corner rooms, p_i and p_j for $1 \le i, j \le 4$. Since $M(T_{ij}P) = T_{ij}MP = T_{ij}P$, we see that $T_{ij}P$ is also a probability eigenvector with eigenvalue $\lambda = 1$. Thus, by uniqueness of this probability eigenvector, $T_{ij}P = P$ so "by symmetry" P has the special form $P = (x, x, x, x, y)$ with $1 = \sum p_i = 4x + y$. The system of equations $P = MP$ now involves only two unknowns x, y. Its first equation is $x = \frac{1}{3}x + \frac{1}{3}x + \frac{1}{4}y$, that is $4x = 3y$. Combined with $4x + y = 1$ one finds $x = \frac{3}{16}$, $y = \frac{1}{4}$. Therefore 25% of the time is spent in the center room and 18.75% in each of the corner rooms. Symmetry turned a potentially messy computation into a simple one.

Proof that M^k converges

Recall that for the transition matrix M of a Markov chain all of its elements are non-negative and that the sum of each column is one. M is *regular* if for some power N all the elements of M^N are positive.

Theorem *If M is the transition matrix of a regular Markov chain, then as $k \to \infty$ the powers M^k converge to a matrix M_∞ whose columns are the same, so for any given row all the elements are the same.*

Before proving this, note that we already know that under these assumptions M has exactly one eigenvector P with eigenvalue 1, so $MP = P$. P is normalized so that the sum of its elements is one. Thus, assuming the above convergence is proved, then each of the columns of M_∞ must just this vector P.

The key to the proof we give here is the

Averaging Lemma: *If one takes a weighted average* $\bar{c} = c_1 w_1 + c_2 w_2 + \cdots + c_n w_n$ *of real numbers* $c_1, \ldots, c_n$*, where* $0 \le \gamma \le w_j$ *and* $w_1 + \cdots + w_n = 1$*, then the average lies between the max and min of the* c_j *with the quantitative estimate*

$$\gamma c_{\max} + (1-\gamma) c_{\min} \le \bar{c} \le (1-\gamma) c_{\max} + \gamma c_{\min}. \tag{1}$$

This is obvious if $\gamma = 0$, so the new information is when $\gamma > 0$.

To prove the Lemma, note that the largest weighted average would occur if all but one of the c_j's equals $c_{\max}$ and the remaining c_j equals $c_{\min}$, and this smallest entry had the smallest weight, γ. This proves the right-hand side of (1). Here is a more algebraic version of the same proof. Say $c_{\min} = c_1$. Since $w_2 + \cdots + w_n = 1 - w_1$,

$$\begin{aligned} \bar{c} &= c_1 w_1 + c_2 w_2 + \cdots + c_n w_n \le c_{\min} w_1 + c_{\max}(w_2 + \cdots + w_n) \\ &= c_{\min} w_1 + c_{\max}(1 - w_1) = c_{\max} - [c_{\max} - c_{\min}] w_1 \\ &\le c_{\max} - [c_{\max} - c_{\min}]\gamma = (1-\gamma) c_{\max} + \gamma c_{\min}. \end{aligned}$$

The left-hand side of (1)is proved similarly. This completes the proof of the lemma.

We now apply this to prove the theorem. We first treat the special case where all the elements of M are positive, so let $\gamma > 0$ be the smallest element of M. The plan is to show that the transpose matrices M^{*k} converge. Because the sum of the elements in any row of M^* is 1, if $v^{[0]} = v$ is any column vector let $v^{[j]} := M^{*j} v$. Observe that the elements of $v^{[1]} := M^* v$ are various averages of v. Thus the above estimate gives the upper bound for $v^{[1]}_{\max} \le (1-\gamma) v_{\max} + \gamma v_{\min}$ and similarly $\gamma v_{\max} + (1-\gamma) v_{\min} \le v^{[1]}_{\min}$. These imply

$$\begin{aligned} v^{[1]}_{\max} - v^{[1]}_{\min} &\le [(1-\gamma) v_{\max} + \gamma v_{\min}] - [\gamma v_{\max} + (1-\gamma) v_{\min}] \\ &= (1 - 2\gamma)(v_{\max} - v_{\min}). \end{aligned}$$

Because $0 < 1 - 2\gamma < 1$, this shows that applying the matrix M^* to a vector "squeezes" the output vector. Since $v^{[k]} = M v^{[k-1]} = M^{*k} v$ we can repeat this contraction to obtain

$$0 \le v^{[k]}_{\max} - v^{[k]}_{\min} \le (1 - 2\gamma)(v^{[k-1]}_{\max} - v^{[k-1]}_{\min}) \le (1 - 2\gamma)^k (v_{\max} - v_{\min}).$$

Thus $v_{\max}^{[k]} - v_{\min}^{[k]} \to 0$. This shows that each element of the vector $M^{*k}v$ converges to the *same* number, so

$$\lim_{k\to\infty} M^{*k}v = \begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix}$$

for some constant c that depends on v. In particular, to get the j^{th} column of M^*_∞ use the case where v is the j^{th} standard basis vector. This proves the convergence of M^{*k} and hence of M^k.

As an exercise, show that the above theorem also holds if one assumes only that for some power N all the elements of M^N are positive.

Problems

1a). If A and B are both $n \times n$ transition matrices for Markov chains, does their product AB also have this property? Proof or counterexample.

b). With A and B as above, if either all the elements of A are positive or all the elements of B are positive, does this imply that all the elements of AB are positive? Proof or counterexample.

2. Let M be the transition matrix of a Markov process.

a). Show that the property that "all the elements of M^N are positive" is equivalent to the statement "it is possible to go from any state to any other in exactly N steps."

b). Use the example $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ to show that the above property is *not* equivalent to the statement "it is possible to go from any state to any other in at most N steps."

3. Let M be the transition matrix of a Markov process with n states $s_1, \ldots, s_n$ so M is an $n \times n$ matrix. Say that within the first n steps you cannot go from state s_1 to state s_2. Does this mean it is *impossible* to go from state s_1 to state s_2 no matter how many more steps you try? Proof or counterexample.

4. Say A is the matrix of a Markov chain all of whose elements are positive. Construct a new Markov chain whose transition matrix M agrees with A, except that its first column has been replaced by the standard column unit vector $e_1 = (1, 0, \ldots, 0)$. Show that for any initial state P (a probability vector) one has $\lim_{k\to\infty} M^k P = e_1$.

This is essentially one model of the spreading of a fatal disease (no one recovers), where everyone has a positive probability of catching the disease. Eventually everyone dies of this disease. The ebola virus, for which the death rate is about 80% is alarmingly close to this.

Solution to Problem 4

Here is one approach. We show that M^k converges to a matrix whose first row is all 1's and all the other elements are 0. Say $M = (m_{ij})$ is our given $n \times n$ matrix. The key facts we use are

- Each element in the first row of M is positive: $m_{1j} \geq \alpha > 0$ where $1 \leq j \leq n$ and $0 < \alpha \leq 1$ is the smallest element in the first row of M. [We will not use the positivity of the other rows.]
- The sum of each column of M is 1: $m_{1j} + m_{2j} + \cdots + m_{nj} = 1$, so $m_{2j} + \cdots + m_{nj} \leq 1 - \alpha$ for every j.

Now for $k = 1, 2, \ldots$, the matrix M^k has the form

$$M^k = \begin{pmatrix} 1 & b_{12} & \cdots & b_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & b_{n2} & \cdots & b_{nn} \end{pmatrix},$$

where the values of the b_{ij}'s of course depend on k. We want to show that for $i \geq 2$ the elements $b_{ij} \to 0$ as $k \to \infty$. Assuming this is done, then since the sum of each column of M^k is 1, this will also show that each of the elements in the first row of M^k each converge to 1.

A key ingredient in the proof is to write $M^{k+1} = M^k M$ (rather than the equally reasonable choice $M^{k+1} = M M^k$). Using this

$$M^{k+1} = M^k M = \begin{pmatrix} 1 & b_{12} & \cdots & b_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & b_{n2} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} 1 & m_{12} & \cdots & m_{1n} \\ 0 & m_{22} & \cdots & m_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & m_{n2} & \cdots & m_{nn} \end{pmatrix}.$$

Write $M^{k+1} = (c_{ij})$ and let β_k be the largest of the numbers b_{ij} for $i \geq 2$. Since for $q \geq 2$ we know $b_{q1} = 0$, then for $i \geq 2$, we have

$$c_{ij} = b_{i2} m_{2j} + \cdots + b_{in} m_{nj} \leq \beta_k (m_{2j} + \cdots + m_{nj}) \leq \beta_k (1 - \alpha).$$

Thus for any $i \geq 2$ we know the largest of the c_{ij} is at most $\beta_k(1 - \alpha)$, that is, $\beta_{k+1} \leq (1 - \alpha)\beta_k$. Consequently, $\beta_{k+1} \leq (1 - \alpha)^k \beta_1 \to 0$ as $k \to \infty$.

This proves, as desired, that the elements in all but the first row of M^k converge to zero as k tends to infinity.

REMARK Given a probability vector P, let $Z_k = M^k P$ Then $\lim_{k\to\infty} Z_k = e_1$, so the first component of Z_k converges to 1 and the others to 0. It is easy to see that the first component of Z_k increases monotonically as k increases.

However, the other components of Z_k are not necessarily decreasing monotonically, even though we know they converge to 0 to zero. Here is an example.

$$Z_1 = MP = \begin{pmatrix} 1 & .1 & .1 \\ 0 & .8 & .8 \\ 0 & .1 & .1 \end{pmatrix} \cdot \begin{pmatrix} .4 \\ .3 \\ .3 \end{pmatrix} = \begin{pmatrix} .46 \\ .48 \\ .06 \end{pmatrix}$$

The second component of $Z_1 = MP$ is larger than the second component of P, even though for k large the second component of $M^k P$ will converge to 0.