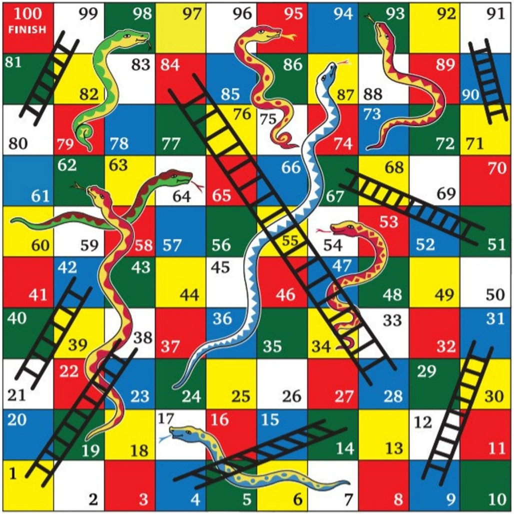
Markov Chains and Applications to Population Models

Steve Kirkland

Department of Mathematics
University of Manitoba

Warmup:



Source: StudioAllenShop

Both players start on square 1. Players take turns rolling the die, advancing on the board, moving up ladders and down snakes, depending on which square they land on. Game ends the first time that either of the players lands on square 100.

Key features:

• The state of the game can be represented by an ordered pair $(n_1, n_2), n_1, n_2 \in \{1, \ldots, 100\}$, where n_j is the number of the square that player j sits on, $j = 1, 2$.

• The state of the game after k rolls of the die depends on the state of the game after $k - 1$ rolls.

• The state of the game after k rolls of the die does *not* depend the earlier history of the game for rolls $k - 2, k - 3$, etc.

Generalize:

• Consider a particle that can occupy one of n states, labelled $1, \ldots, n$.

• At each discrete time step, the particle can either stay in its current state or move to a new state.

• For each $i, j = 1, \ldots, n$, let a_{ij} denote the probability that the particle moves from state i to state j in one time step. Note that these probabilities are assumed to be the same for all time steps.

• Starting from some initial probability distribution for the particle occupying the various states, we iterate the system in discrete time, according to these rules.

This is a discrete–time, time homogeneous Markov chain on a finite state space.

Formalize:
An $n \times n$ matrix A is stochastic if it is entrywise nonnegative, and $A\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ denotes the all ones vector in $\mathbb{R}^n$. Note that 1 is an eigenvalue of A, with $\mathbf{1}$ as a corresponding right eigenvector.

Let $x(0)^\top$ denote an initial probability vector in $\mathbb{R}^n$ – i.e. a nonnegative vector such that $x(0)^\top\mathbf{1} = 1$. The Markov chain associated with A is sequence of nonnegative vectors $x(k), k = 0, 1, 2, 3, \ldots$ satisfying $x(k+1)^\top = x(k)^\top A$ and $x(k)^\top\mathbf{1} = 1, k \in \mathbb{N} \cup \{0\}$. Evidently $x(k)^\top = x(0)^\top A^k$, for each $k \in \mathbb{N}$.

A is the transition matrix for the Markov chain.

Species succession:

.7725	.1022	.0170	.0040	.0150	.0010	.0180	.0120	.0020	.0140	.0030	.0020	.0050	.0030	.0291
.1450	.6090	.0310	.0110	.0280	.0050	.0220	.0250	.0110	.0150	.0120	.0080	.0050	.0040	.0690
.0519	.0609	.7093	.0040	.0200	.0040	.0080	.0080	.0250	.0030	.0050	.0070	.0020	.0080	.0839
.0170	.0541	.0060	.8398	.0050	0	.0040	.0060	.0080	.0040	.0060	.0110	0	.0030	.0360
.1169	.2178	.0350	.0040	.4036	.0080	.0330	.0320	.0130	.0070	.0060	.0050	.0060	.0050	.1079
.0090	.0240	.0120	0	.0160	.8647	.0010	.0070	.0160	.0030	.0040	.0070	0	0	.0361
.2412	.2232	.0511	.0160	.0801	.0240	.1051	.0410	.0140	.0330	.0250	.0050	.0140	.0120	.1151
.1986	.2345	.0379	.0180	.0888	.0070	.0439	.1537	.0150	.0269	.0160	.0200	.0090	.0090	.1218
.0559	.1469	.0260	.0110	.0200	.0060	.0110	.0260	.5854	.0210	.0060	.0050	.0010	.0050	.0739
.3084	.2275	.0309	.0100	.0269	.0060	.0419	.0309	.0100	.1647	.0130	.0080	.0120	.0060	.1038
.0559	.2216	.0279	.0080	.0359	0	.0250	.0200	.0070	.0070	.5060	.0020	.0050	.0030	.0758
.0250	.0680	.0180	.0300	.0160	0	.0100	.0160	.0040	.0030	.0010	.5370	.0030	.0030	.2660
.3210	.1790	.0230	0	.0630	0	.0300	.0200	.0030	.0200	.0170	0	.2480	0	.0760
.1583	.4489	.0180	.0180	.0852	.0060	.0301	.0180	.0180	.0301	.0060	.0060	0	.0301	.1273
.1010	.3200	.0250	.0090	.0620	.0050	.0480	.0340	.0130	.0310	.0170	.0170	.0110	.0130	.2940

Random walk on a graph:
Let G be a connected undirected graph on vertices labelled $1, \ldots, n$. Think of a random walker who vbegins at an initial vertex j_0 of G, and proceeds as follows: if the random walker is at vertex j_k at time k, it selects one of the neighbours of j_k at random, and moves to that vertex at time $k+1$. One may think of an intruder entering a network and randomly wandering from one vertex to another.

The corresponding transition matrix, A, which can be written as $D^{-1}C$, where:
i) C is the adjacency matrix for G, i.e. $c_{jk} = 1$ if vertices j and k are adjacent and $c_{jk} = 0$ otherwise;
ii) D is the diagonal matrix of vertex degrees, i.e. $D = \mathrm{diag}(C\mathbf{1})$.

Google uses a variation of this notion in its random surfer model that underpins its ranking of web pages.

A simple SIS model:
Fixed population of size N. Susceptible individuals may be infected via contact with infected individuals, and infected individuals may recover and become susceptible again. Individuals mix homogeneously.

Markov chain with states $0, \dots, N$, where the state of the chain is the number of infected individuals. Let β, γ represent the transmission and recovery rates, respectively. Under some simplifying assumptions, we have a transition matrix A such that:
$a_{jj+1} = \frac{\beta j(N-j)}{N}, j = 0, \dots, N-1,$
$a_{jj-1} = \gamma j, j = 1, \dots, N,$
$a_{jj} = 1 - \frac{\beta j(N-j)}{N} - \gamma j, j = 1, \dots, N,$
$a_{00} = 1.$

For $N = 5$, we have

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \gamma & 1-\gamma-\frac{4\beta}{5} & \frac{4\beta}{5} & 0 & 0 & 0 \\ 0 & 2\gamma & 1-2\gamma-\frac{6\beta}{5} & \frac{6\beta}{5} & 0 & 0 \\ 0 & 0 & 3\gamma & 1-3\gamma-\frac{6\beta}{5} & \frac{6\beta}{5} & 0 \\ 0 & 0 & 0 & 4\gamma & 1-4\gamma-\frac{4\beta}{5} & \frac{4\beta}{5} \\ 0 & 0 & 0 & 0 & 5\gamma & 1-5\gamma \end{bmatrix}.$$

Note that A is not irreducible, though it has 1 as a simple eigenvalue.

Theorem

Suppose that A is an irreducible stochastic matrix. There is a unique vector $w > 0$ such that $w^\top A = w^\top, w^\top \mathbf{1} = 1$.
If A is primitive, then $A^k \to \mathbf{1} w^\top$ as $k \to \infty$.
If A is irreducible, then $\frac{1}{m}\sum_{k=0}^{m-1} A^k \to \mathbf{1} w^\top$ as $m \to \infty$.

The vector w is called the stationary distribution for the Markov chain.

Corollary

Given a primitive stochastic matrix A, the corresponding Markov chain $x(k)$ converges to the stationary distribution, independently of the initial vector $x(0)$.
Given an irreducible stochastic matrix A, the corresponding Markov chain $x(k)$ has the property that the sequence of averages $\frac{1}{m}(x(0) + \ldots + x(m-1))$ converges to the stationary distribution, independently of the initial vector $x(0)$.

So, the stationary distribution vector w carries the long–term information about the behaviour of the Markov chain associated with A.

E.g. For the species succession example the stationary distribution is:

$$\begin{bmatrix} .3387 & .2631 & .0760 & .0454 & .0435 & .0272 & .0232 & .0220 & .0217 & .0170 & .0162 & .0139 & .0069 & .0050 & .0801 \end{bmatrix}.$$

E.g. For the example of a random walk on a connected graph, recall that $A = D^{-1}C$, where $D = \mathrm{diag}(D\mathbf{1})$. Observe that $(\mathbf{1}^\top D)A = (\mathbf{1}^\top D)D^{-1}C = \mathbf{1}^\top C = (\mathbf{1}^\top D)$. Hence the stationary distribution is $\frac{1}{\mathbf{1}^\top D\mathbf{1}}\mathbf{1}^\top D$.

Stationary vectors: another view

Suppose that A is irreducible $n \times n$ and stochastic, with stationary vector $w^\top$. Then
$(I - A)\mathrm{adj}(I - A) = \mathrm{adj}(I - A)(I - A) = (\det(I - A))I = 0.$
Deduce that $\mathrm{adj}(I - A) = c\mathbf{1}w^\top$, for some constant c.

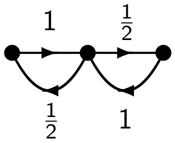
Hence $w^\top$ is a scalar multiple of the vector whose j–th entry is $\det(I - A_{(j)})$, where $A_{(j)}$ is formed from A by deleting the j–th row and column.

That determinant can be evaluated via the Matrix Tree Theorem.

Consider the weighted directed graph associated with the off–diagonal entries in A. Look for the directed trees that a) use all of the vertices, b) have all of the arcs directed towards j. Compute the weight of those trees as the product of the corresponding arc weights, then sum.

That sum is $\det(I - A_{(j)})$.

E.g. $A = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$.



$w^\top = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$

Suppose that we have an irreducible $n \times n$ stochastic matrix A. The short–term properties of the corresponding Markov chain are captured by its mean first passage times. The mean first passage time from state j to state k, m_{jk}, is the expected number of steps required for the Markov chain to reach state k for the first time, given that it started at state j. Informally we may think of these mean first passage times as 'travel times' between states.

Condition on the state after one step:
$m_{jk} = a_{jk} + \sum_{\ell \neq k} a_{j\ell}(m_{\ell k} + 1) = 1 + \sum_{\ell \neq k} a_{j\ell} m_{\ell k}$.

In matrix terms: $M = A(M - M_{\mathrm{dg}}) + J$, where $J =$all ones, and M_{dg} is the diagonal matrix whose diagonal entries coincide with those of M.

Denote the stationary vector by $w^\top$ and observe that $w^\top M = w^\top A(M - M_{\mathrm{dg}}) + w^\top J = w^\top M - w^\top M_{\mathrm{dg}} + \mathbf{1}^\top$. Deduce that $w^\top M_{\mathrm{dg}} = \mathbf{1}^\top$, i.e. $m_{jj} = \frac{1}{w_j}$ for $j = 1, \ldots, n$.

Offdiagonal entries of M? Have $(I - A)M = J - AM_{\mathrm{dg}}$, so look at the last column on both sides. Set $I - A = \left[\begin{array}{c|c} I - A_{(n)} & -(I - A_{(n)})\mathbf{1} \\ \hline * & * \end{array}\right]$, $M = \left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline * & m_{nn} \end{array}\right]$. Then $(I - A_{(n)})M_{12} - m_{nn}(I - A_{(n)})\mathbf{1} = \mathbf{1} - m_{nn}(I - A_{(n)})\mathbf{1}$, and deduce that $M_{12} = (I - A_{(n)})^{-1}\mathbf{1}$.

In general, for $j \neq k$ m_{jk} corresponds to the appropriate entry of $(I - A_{(k)})^{-1}\mathbf{1}$.

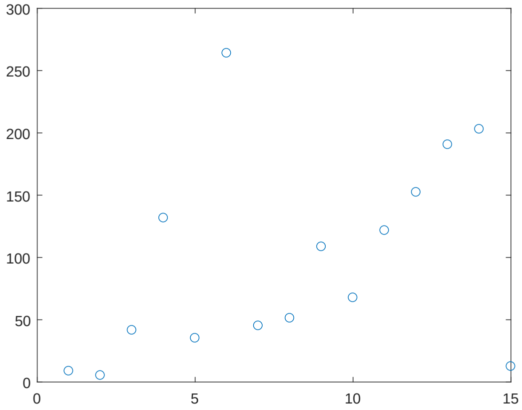
E.g. Random walk on a path on 3 vertices:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \text{ and } M = \begin{bmatrix} 4 & 1 & 4 \\ 3 & 2 & 3 \\ 4 & 1 & 4 \end{bmatrix}.$$

Sample computation: $(I - A_{(3)}) = \begin{bmatrix} 1 & -1 \\ -\frac{1}{2} & 1 \end{bmatrix}$ so

$(I - A_{(3)})^{-1} = \begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix}$, which has row sum vector $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$.

E.g. Species succession: mean first passage times from state 15 (bare rock)



For the SIS model, we can still consider the mean first passage times into state 0 (i.e. no infected individuals), computed as $(I - A_{(0)})^{-1}\mathbf{1}$.

For example if $N = 5$, we have

$$\begin{bmatrix} m_{10} \\ m_{20} \\ m_{30} \\ m_{40} \\ m_{50} \end{bmatrix} = \begin{bmatrix} \frac{24\beta^4+150\beta^3\gamma+500\beta^2\gamma^2+1250\beta\gamma^3+3125\gamma^4}{3125\gamma^5} \\ \frac{48\beta^4+360\beta^3\gamma+1375\beta^2\gamma^2+3750\beta\gamma^3+9375\gamma^4}{6250\gamma^5} \\ \frac{144\beta^4+1080\beta^3\gamma+4425\beta^2\gamma^2+13125\beta\gamma^3+34375\gamma^4}{18750\gamma^5} \\ \frac{288\beta^4+2160\beta^3\gamma+8850\beta^2\gamma^2+27750\beta\gamma^3+78125\gamma^4}{37500\gamma^5} \\ \frac{288\beta^4+2160\beta^3\gamma+8850\beta^2\gamma^2+27750\beta\gamma^3+85625\gamma^4}{37500\gamma^5} \end{bmatrix}.$$

Evidently $m_{10} < m_{20} < m_{30} < m_{40} < m_{50}$; each m_{j0} is increasing in β and decreasing in γ.

Consider our irreducible $n \times n$ stochastic matrix A with stationary vector $w^\top$, and denote the eigenvalues of A by $1, \lambda_2, \dots, \lambda_n$.

The matrix $I - A + \mathbf{1}w^\top$ has eigenvalues $1, 1 - \lambda_2, \dots, 1 - \lambda_n$, and so in particular it is nonsingular.

It turns out that the mean first passage matrix M is given by $M = (I - Z + JZ_{\mathrm{dg}})W^{-1}$, where: $Z = (I - A + \mathbf{1}w^\top)^{-1}$ and W is the diagonal matrix constructed from the entries of w.

Use the relationships between $A, Z, w^\top$ to show that this candidate satisfies $M = A(M - M_{\mathrm{dg}}) + J$.

Lecture 9 will present an alternate expression for the mean first passage matrix.

Recall that we used the equation $M = A(M - M_{\rm dg}) + J$, multiplied on the left by $w^\top$, to generate the connection between the m_{jj}'s and the w_j's.

How about multiplying on the right?
$Mw = A(M - M_{\rm dg})w + Jw = AMw - A\mathbf{1} + \mathbf{1} = A(Mw)$. So, Mw is an eigenvector of A corresponding to the eigenvalue 1.

Deduce that there is a constant c such that $Mw = c\mathbf{1}$.

For each index j the expression $\sum_{k=1}^n m_{jk} w_k (= 1 + \sum_{k \neq j} m_{jk} w_k)$ is independent of the index j.

In other words, the expected number of steps to go from state j to a randomly chosen state (i.e. randomly chosen according to the stationary distribution) does not depend on j.

What's the constant?

We have $Mw = (I - Z + JZ_{\mathrm{dg}})W^{-1}w = (I - Z + JZ_{\mathrm{dg}})\mathbf{1} = \mathbf{1} - Z\mathbf{1} + trace(Z)\mathbf{1} = trace(Z)\mathbf{1}$.

If the eigenvalues of A are $1, \lambda_2, \ldots, \lambda_n$, then the eigenvalues of Z are $1, \frac{1}{1-\lambda_2}, \ldots, \frac{1}{1-\lambda_n}$. Hence

$$Mw = \left(1 + \sum_{k=2}^{n} \frac{1}{1-\lambda_k}\right)\mathbf{1}.$$

The number $\kappa = \sum_{k=2}^{n} \frac{1}{1-\lambda_k}$ is called Kemeny's constant for the Markov chain. Evidently $\kappa + 1 = \sum_{j,k=1}^{n} w_j m_{jk} w_k$ so Kemeny's constant can be expressed in terms of the expected number of steps to go from a randomly chosen initial state to a randomly chosen destination state.

E.g. Random walk on a path on 3 vertices:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}, w^\top = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \text{ and } M = \begin{bmatrix} 4 & 1 & 4 \\ 3 & 2 & 3 \\ 4 & 1 & 4 \end{bmatrix}.$$

$Mw = \frac{5}{2}\mathbf{1}$, so $\kappa = \frac{3}{2}$.

E.g. Species succession: for this example, Kemeny's constant is 34.3260. We saw that some of the mean first passage times were large (on the order of 260), but those were into states corresponding to small entries in the stationary vector, hence the modest value of κ.

Kemeny's constant arises in a number of applied settings, including: vehicle traffic networks (overall efficiency), wireless networks (robustness), economics (flow of money between nations), and consensus algorithms (resistance to noise).

Sample results:
For an irreducible $n \times n$ stochastic matrix, the corresponding value of Kemeny's constant satisfies $\kappa \geq \frac{n-1}{2}$. Equality holds in the lower bound iff the transition matrix is the adjacency matrix of a directed n–cycle: $1 \rightarrow 2 \rightarrow \ldots \rightarrow n \rightarrow 1$.

For a random walk on a connected graph on n vertices, $\kappa \geq \frac{(n-1)^2}{n}$, with equality if and only if it is the complete graph.

References:

Kemeny and Snell, *Finite Markov Chains*, Springer–Verlag, 1976.

Seneta, *Non-negative Matrices and Markov Chains,* 2^{nd} *edition*, Springer, 1981.

Levene and Loizou, Kemeny's Constant and the Random Surfer, *American Mathematical Monthly* 109, 2002.

Hill, Witman and Caswell, Markov chain analysis of succession in a rocky subtidal community, *A*merican Naturalist 164, 2004.