Embeddable Markov Matrices

E B Davies

11 January 2010

Abstract

We give an account of some results, both old and new, about any $n \times n$ Markov matrix that is embeddable in a one-parameter Markov semigroup. These include the fact that its eigenvalues must lie in a certain region in the unit ball. We prove that a well-known procedure for approximating a non-embeddable Markov matrix by an embeddable one is optimal in a certain sense.

1   Introduction

A Markov matrix A is defined to be a real $n \times n$ matrix with non-negative entries satisfying $\sum_{j=1}^{n} A_{i,j} = 1$ for all i. The spectral properties of non-negative matrices and linear operators and in particular of Markov matrices have been studied in great detail, because of their great importance in finance, population dynamics, medical statistics, sociology and many other areas of probability and statistics. Theoretical accounts of parts of the subject may be found in [1, 5, 16, 17]. This paper develops ideas of [10], which investigated when the pth roots of Markov matrices were also Markov; this problem is related to the possibility of passing from statistics gathered at certain time intervals, for example every year, to the corresponding data for shorter time intervals.

Given an empirical Markov matrix, three major issues discussed in [19] are embeddability, uniqueness of the embedding and the effects of data/sampling error. All of these are also considered here. We call a Markov matrix A embeddable if there exists a matrix B such that $A = e^B$ and e^{Bt} is Markov for all $t \geq 0$. The matrix B involved need not be unique, but it must have non-negative off-diagonal entries and all its row sums must vanish; see [1] or

[5, section 12.3]. In probabilistic terms a Markov matrix A is embeddable if it is obtained by taking a snapshot at a particular time of an autonomous finite state Markov process that develops continuously in time. On the other hand a Markov matrix might not be embeddable if it describes the annual changes in a population that has a strongly seasonal breeding pattern; in such cases one might construct a more elaborate model that incorporates the seasonal variations. Embeddability may also fail because the matrix entries are not accurate; in such cases a regularization technique might yield a very similar Markov matrix that is embeddable; see [15] for examples arising in finance.

Theorem 9 describes some spectral consequences of embeddability. The earliest analysis of the structure of the set $\mathcal{E}$ of embeddable $n \times n$ Markov matrices and its topological boundary in the set of all Markov matrices was given by Kingman [14], who concluded that except in the case $n = 2$ it seemed unlikely that any very explicit characterisation of $\mathcal{E}$ could be given; see [12] for further work on this problem. Theorem 13 proves that a well-known method of approximating a Markov matrix by an embeddable Markov matrix is optimal in a certain sense. Many of the results in the present paper appear in one form or other in papers devoted to the wide variety of applications, and it is hoped that collecting them in one place may be of value.

2 The main theorem

For the sake of definiteness we define the principal logarithm of a number $z \in \mathbf{C} \backslash (-\infty, 0]$ to be the branch of the logarithm with values in $\{w : |\mathrm{Im}(w)| < \pi\}$. We define the principal logarithm of an $n \times n$ matrix A such that $\mathrm{Spec}(A) \cap (-\infty, 0] = \emptyset$ to be that defined by the functional integral

$$\log(A) = \frac{1}{2\pi i} \int_\gamma \frac{\log(z)}{zI - A} \, \mathrm{d}z \tag{1}$$

using the principal logarithm of z and a simple closed contour γ in $\mathbf{C} \backslash (-\infty, 0]$ that encloses the spectrum of A. This formula goes back to Giorgi in 1926; see [6, Theorem VII.1.10 and notes, p.607]. If $A = TDT^{-1}$ where D is diagonal, this is equivalent to $\log(A) = T \log(D) T^{-1}$ where $\log(D)$ is obtained from D by applying log to each diagonal entry of D. The non-diagonalisable case is discussed in some detail in [19] and yields the same matrix as (1).

Lemma 1 *If A is a Markov matrix and $\mathrm{Spec}(A) \cap (-\infty, 0] = \emptyset$ then the principal logarithm $L = \log(A)$ lies in the set $\mathcal{L}$ of all real $n \times n$ matrices L such that $\sum_{1 \leq j \leq n} L_{i,j} = 0$ for every i.*

2

Proof We use the formula (1) and take the contour γ to be symmetrical with respect to reflection about the x-axis. The statements of the lemma follow directly from two properties of the resolvent matrices.

The first is the identity

$$((\overline{z}I - A)^{-1})_{i,j} = \overline{(zI - A)^{-1})_{i,j}} \tag{2}$$

This holds for large $|z|$ by virtue of the identity

$$(zI - A)^{-1} = z^{-1} \sum_{n=0}^{\infty} (A/z)^n \tag{3}$$

and (2) then extends to all $z \notin \mathrm{Spec}(A)$ by analytic continuation.

The second identity needed is

$$(zI - A)^{-1}1 = (z - 1)^{-1}1,$$

whose proof follows the same route, using (3) and analytic continuation. ∎

The results in our next lemma are all well known and are included for completeness.

Lemma 2 *If A is embeddable then 0 is not an eigenvalue of A and every negative eigenvalue has even algebraic multiplicity. Moreover $\det(A) > 0$. If A is embeddable and $A_{i,j} > 0$, $A_{j,k} > 0$ then $A_{i,k} > 0$.*

Proof The first statement follows from the fact that

$$\mathrm{Spec}(A) = \exp(\mathrm{Spec}(B)).$$

Given an eigenvalue $\lambda < 0$ of A let

$$\begin{aligned}
S_+ &= \{z \in \mathrm{Spec}(B) : \mathrm{e}^z = \lambda \text{ and } \mathrm{Im}(z) > 0\}, \\
S_- &= \{z \in \mathrm{Spec}(B) : \mathrm{e}^z = \lambda \text{ and } \mathrm{Im}(z) < 0\},
\end{aligned}$$

and let $\mathcal{L}_\pm$ be the spectral projections of B associated with $S_\pm$. Since $\mathrm{e}^z = \lambda$ implies that $\mathrm{Im}(z) \neq 0$, we can deduce that $\mathcal{L}_- \cap \mathcal{L}_+ = 0$ and that $\mathcal{M} = \mathcal{L}_- + \mathcal{L}_+$ is the spectral projection of A associated with the eigenvalue λ. Since B is real $\mathcal{L}_-$ may be obtained from $\mathcal{L}_+$ by complex conjugation, so

$$\dim \mathcal{M} = \dim(\mathcal{L}_+) + \dim(\mathcal{L}_-) = 2 \dim(\mathcal{L}_+).$$

See [7].

By combining the reality of B with the formula $\det(A) = \mathrm{e}^{\mathrm{tr}(B)}$ we obtain $\det(A) > 0$. See [14].

3

The last statement follows from the general theory of Markov chains and is due to Ornstein and Levy, independently; see [1, Section 2.5, Theorem 2] and [5, Theorem 13.2.4]. We first note that one may write $B = C - \delta I$ where all the entries of C are non-negative and $\delta \geq 0$. Hence

$$\mathrm{e}^{Bt} = \mathrm{e}^{-\delta t} \sum_{n=0}^{\infty} C^n t^n / n!$$

where each entry of each C^n is non-negative. This implies that if $(\mathrm{e}^{Bt})_{i,j} > 0$ for some $t > 0$ the same holds for all $t > 0$. This quickly yields the final statement. ∎

Kingman [14] has shown that the set $\mathcal{E}$ of embeddable Markov matrices is a closed subset of the set of all $n \times n$ Markov matrices. The matrix norm used throughout this paper is

$$\|M\| = \max\{\|Mv\|_\infty : \|v\|_\infty \leq 1\} = \max_{1 \leq i \leq n} \left\{ \sum_{j=1}^{n} |M_{i,j}| \right\}. \qquad (4)$$

Lemma 3 *The set $\mathcal{S}$ of all $A \in \mathcal{E}$ with no negative eigenvalues is a dense relatively open subset of $\mathcal{E}$.*

Proof If $A \in \mathcal{S}$ then a simple perturbation theoretic argument implies that there exists $\varepsilon > 0$ such that C has no negative eigenvalues for any $n \times n$ matrix satisfying $\|A - C\| < \varepsilon$. This implies that $\mathcal{S}$ is relatively open in $\mathcal{E}$.

If $A \in \mathcal{E}$ then $A = \mathrm{e}^B$ for some Markov generator B. If $\{x_r + iy_r\}_{r=1}^n$ is the set of eigenvalues of B and $t > 0$ then e^{Bt} has a negative eigenvalue if and only if $ty_r = \pi(2m + 1)$ for some r and some integer m. The set of such t is clearly discrete. It follows that $\mathrm{e}^{Bt} \in \mathcal{S}$ for all t close enough to 1 except possibly for 1 itself. Since $\lim_{t \to 1} \|\mathrm{e}^{Bt} - A\| = 0$, we conclude that $\mathcal{S}$ is dense in $\mathcal{E}$. ∎

The following example shows that the density property in Lemma 3 depends on the embeddability hypothesis.

Example 4

The Markov matrix

$$A = \begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$$

has $\mathrm{Spec}(A) = \{1, -1/3\}$. If $0 < \varepsilon < 1/3$, any matrix close enough to A also has a single eigenvalue λ satisfying $|\lambda + 1/3| < \varepsilon$ by a standard perturbation theoretic argument. Since A has real entries the complex conjugate of λ is also

an eigenvalue, so λ must be real and negative. Therefore the set of Markov matrices with no negative eigenvalues is relatively open but not dense in the set of all Markov matrices, at least for $n = 2$. The example may be used to construct a similar example for every $n > 2$. ∎

We will need Lemma 5 and its corollary in the proof of Theorem 7.

Lemma 5 *There exists a polynomial p in the coefficients of an $n \times n$ matrix A such that A has a multiple eigenvalue in the algebraic sense if and only if $p = 0$. We call p the discriminant of A.*

Proof A has a multiple eigenvalue if and only if its characteristic polynomial $q(z) = z^n + a_1 z^{n-1} + \ldots + a_n$ has a multiple root; the coefficients of q are themselves polynomials in the entries of A. Moreover q has a multiple root if and only if its discriminant (the square of its Vandermonde determinant) vanishes, and the discriminant of q is a polynomial in $a_1, \ldots, a_n$. ∎

Corollary 6 *If A_0, A_1 are two $n \times n$ matrices then either $A_z = (1-z)A_0 + zA_1$ has a multiple eigenvalue for all $z \in \mathbf{C}$ or this happens only for a finite number of z.*

Proof The discriminant of A_z is a polynomial in z, which has a finite number of roots unless it vanishes identically. ∎

Theorem 7 *The set $\mathcal{T}$ of all $n \times n$ embeddable Markov matrices that have n distinct eigenvalues is relatively open and dense in the set $\mathcal{E}$ of all embeddable Markov matrices.*

Proof A standard argument from perturbation theory establishes that $\mathcal{T}$ is relatively open in $\mathcal{E}$, so we only need to prove its density.

Let $A = \mathrm{e}^{B_0}$ where B_0 is a Markov generator, and let $\varepsilon > 0$. Then put $B_t = (1-t)B_0 + tB_1$ where

$$(B_1)_{r,s} = \begin{cases} -1 & \text{if } r = s, \\ 1 & \text{if } r+1 = s, \\ 1 & \text{if } r = n,\ s = 1, \\ 0 & \text{otherwise.} \end{cases}$$

One sees immediately that B_t is a Markov generator for all $t \in [0,1]$ and that it has n distinct eigenvalues if $t = 1$. Corollary 6 now implies that the eigenvalues of B_t are distinct for all sufficiently small $t > 0$. By further restricting the size of $t > 0$ we may also ensure that $\|\mathrm{e}^{B_t} - A\| < \varepsilon/2$.

Having chosen t, we put $B = sB_t$ where $s \in \mathbf{R}$ is close enough to 1 so that $\|\mathrm{e}^{B_t} - \mathrm{e}^B\| < \varepsilon/2$; we also choose s so that if λ_1, λ_2 are any two eigenvalues of

B_t then $s(\lambda_1 - \lambda_2) \notin 2\pi i \mathbf{Z}$. These conditions ensure that $\|e^B - A\| < \varepsilon$ and that e^B has n distinct eigenvalues. ∎

The following lemma may be contrasted with the fact that a complex number λ such that $|\lambda| = 1$ is the eigenvalue of some $n \times n$ Markov matrix if and only if $\lambda^r = 1$ for some $r \in \{1, 2, \ldots, n\}$; see [16, Chap 7, Theorem 1.4]. Permutation matrices provide examples of such spectral behaviour. The lemma has been extended to an infinite-dimensional context in [4].

Lemma 8 (Elfving, [7]) *If A is an embeddable Markov matrix and $\lambda \neq 1$ is an eigenvalue of A then $|\lambda| < 1$.*

Proof Our hypotheses imply by [5, Lemma 12.3.5] that $\mathrm{Spec}(A) = \exp(\mathrm{Spec}(B))$ where $B = c(C - I)$, $c > 0$ and C is a Markov matrix. Since C is a contraction when considered as acting in $\mathbf{C}^n$ with the l^∞ norm, $\mathrm{Spec}(C) \subset \{z : |z| \leq 1\}$. Therefore every eigenvalue λ of B except 0 satisfies $\mathrm{Re}(\lambda) < 0$. The lemma follows. ∎

The main application of the following theorem may be to establish that certain Markov matrices arising in applications are not embeddable, and hence either that the entries are not numerically accurate or that the underlying process is not autonomous. The theorem is a quantitative strengthening of Lemma 8. It is of limited value except when n is fairly small, but this is often the case in applications.

Theorem 9 (Runnenberg, [18, 19]) *If $n \geq 3$ and the $n \times n$ Markov matrix A is embeddable then its spectrum is contained in the set*

$$\{re^{i\theta} : -\pi \leq \theta \leq \pi, 0 < r \leq r(\theta)\}$$

where

$$r(\theta) = \exp(-\theta \tan(\pi/n)).$$

Proof This depends on two facts, firstly that $\mathrm{Spec}(A) = \exp(\mathrm{Spec}(B))$ where $B = c(C - I)$, $c > 0$ and C is a Markov matrix. Secondly

$$\begin{aligned} \mathrm{Spec}(C - I) \quad &\subseteq \quad \{z : |\arg(z)| \geq \pi/2 + \pi/n\} \qquad &(5)\\ &= \quad \{-u + iv : u \geq 0, |v| \leq u \cot(\pi/n)\}. \qquad &(6) \end{aligned}$$

by applying a theorem of Karpelevič to C and then deducing

(6) from that; see [13] or [16, Chap. 7, Theorem 1.8]. The relevant boundary curve (actually a straight line segment from 1 to $e^{2\pi i/n}$) is the case $q = 0$, $p = 1$ and $r = n$ of $\lambda^q(\lambda^p - t)^r = (1 - t)^r$, where $0 \leq t \leq 1$. The small part of the theorem of Karpelevič that we need was proved by Dmitriev and Dynkin; see [16, Chap. 8, Theorem 1.7]. ∎

We turn now to the question of uniqueness. The first example of a Markov matrix A that can be written in the form $A = e^B$ for two different Markov generators was given by Speakman in [20]; Example 17 provides another. The initial hypothesis of our next result holds for most embeddable Markov matrices by Theorem 7.

Corollary 10 *Let A be an invertible $n \times n$ Markov matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$.*

1. *The solutions of $e^B = A$ form a discrete set and they all commute with each other and with A.*

2. *Only a finite number of the solutions of $e^B = A$ can be Markov generators.*

3. *If*
$$|\lambda_r| > \exp(-\pi \tan(\pi/n)) \tag{7}$$
   for all r then only one of the solutions of $e^B = A$ can be a Markov generator, namely the principal logarithm.

Proof

1. Since $\mathrm{Spec}(A) = \exp(\mathrm{Spec}(B))$, each B must have n distinct eigenvalues $\mu_1, \ldots, \mu_n$ and the corresponding eigenvectors form a basis in $\mathbf{C}^n$. These eigenvectors are also eigenvectors for A and
$$\lambda_r = e^{\mu_r} \tag{8}$$
   for all r. It follows that B can be written as a polynomial function of A. For each λ_r, the equation (8) has a discrete set of solutions μ_r.

2. If A is an invertible Markov matrix with distinct eigenvalues and the solution B of $e^B = A$ is a Markov generator then every eigenvalue μ_r of B lies in the sector $\{-u+iv : u \geq 0, |v| \leq u\cot(\pi/n)\}$ by (6). Combining this restriction on the imaginary parts of the eigenvalues with (8) reduces the set of such B to a finite number. See [11, Theorem 6.1] for items 1 and 2, and for an algorithm implementing item 2.

3. We continue with the assumptions and notation of item 2. The assumption (7) implies that if $\mu_r = -u_r + iv_r$ then $u_r < \pi \tan(\pi/n)$. Item 2 now yields $|v_r| < \pi$. Hence μ_r is the principal logarithm of λ_r and B is the principal logarithm of A.

The conclusions of the above corollary do not hold if A has repeated eigenvalues or a non-trivial Jordan form; see [3, 11]. For example the $n \times n$ identity matrix has a continuum of distinct logarithms B which do not all commute; if the eigenvalues of B are chosen to be $\{2\pi r i : 1 \leq r \leq n\}$, then the possible B are parametrized by the choice of an arbitrary basis as its set of eigenvectors. The general classification of logarithms is given in [8] and [9, Theorem 1.28]. These comments reveal a numerical instability in the logarithm of a matrix if it has two or more eigenvalues that are very close to each other.

The following provides a few other conditions that imply the uniqueness of a Markov generator B such that $A = \mathrm{e}^B$.

Theorem 11 (Cuthbert, [2, 3]) *Let $A = \mathrm{e}^B$ where B a Markov generator. Then (9)$\Rightarrow$(10)$\Rightarrow$(11)$\Rightarrow$(12), where*

$$
\begin{aligned}
\mathrm{e}^{-\pi} \;&<\; \det(A) \;\leq\; 1, & (9)\\
-\pi \;&<\; \mathrm{tr}(B) \;\leq\; 0, & (10)\\
\|B + \beta I\| \;&<\; \pi, \qquad \text{where } \beta = \max_{1 \leq i \leq n}\{|B_{i,i}|\}, & (11)\\
\mathrm{Spec}(B) \;&\subseteq\; \{z : \mathrm{Im}(z)| < \pi\}. & (12)
\end{aligned}
$$

If A is a Markov matrix that has distinct eigenvalues and $\det(A) > \mathrm{e}^{-\pi}$ then its only possible Markov generator is its principal logarithm $\log(A)$.

Proof
(9)$\Rightarrow$(10) This uses $\det(A) = \mathrm{e}^{\mathrm{tr}(B)}$.
(10)$\Rightarrow$(11) This uses the fact that $B + \beta I$ has non-negative entries and its row sums all equal β, which satisfies $\beta < \pi$.
(11)$\Rightarrow$(12) follows directly from $\mathrm{Spec}(B + \beta I) \subseteq \{z : |z| < \pi\}$.
The final statement of the theorem follows the proof of Corollary 10. ∎

If A^t is a one-parameter Markov semigroup then for every $t > 0$ one may define $L(t)$ to be the number of Markov generators B such that $\mathrm{e}^{Bt} = A^t$. Some general theorems concerning the dependence of $L(t)$ on t may be found in [3, 19].

3 Regularization

Let $\mathcal{G}$ denote the set of $n \times n$ Markov generators; following the notation of Lemma 1, $\mathcal{G}$ is the set of $G \in \mathcal{L}$ such that $G_{i,j} \geq 0$ whenever $i \neq j$.

Let A be a Markov matrix satisfying the assumptions of Lemma 1, for which $L = \log(A)$ does not lie in $\mathcal{G}$. There are several *regularizations* of L, that is

algorithms that replace L by some $G \in \mathcal{G}$ that are (nearly) as close to L as possible. Kreinin and Sidelnikova [15] have compared different regularization algorithms for several empirical examples arising in finance and it appears that they all have similar accuracy. The best approximation must surely depend on the matrix norm used, but if one considers the physically relevant matrix norm (4) then we prove that the simplest method, called diagonal adjustment in [15], also produces the best possible approximation. We emphasize that although $\mathcal{G}$ is a closed convex cone, this does not imply that the best approximation is unique, because the matrix norm (4) is not strictly convex.

Theorem 12 *Let $L \in \mathcal{L}$ and define $B \in \mathcal{G}$ by*

$$B_{i,j} = \begin{cases} L_{i,j} & \text{if } i \neq j \text{ and } L_{i,j} \geq 0, \\ 0 & \text{if } i \neq j \text{ and } L_{i,j} < 0, \end{cases}$$

together with the constraint $\sum_{j=1}^{n} B_{i,j} = 0$ for all i. Then

$$\|L - B\| = \min\{\|L - G\| : G \in \mathcal{G}\}.$$

Proof It follows from the definition of the matrix norm that we can deal with the matrix rows one at a time. We therefore fix i and put

$$\begin{aligned}
\ell_j &= L_{i,j}, \\
P &= \{j : j \neq i \text{ and } \ell_j \geq 0\}, \\
N &= \{j : j \neq i \text{ and } \ell_j < 0\}, \\
\ell_P &= \sum_{j \in P} \ell_j \geq 0, \\
\ell_N &= -\sum_{j \in N} \ell_j \geq 0,
\end{aligned}$$

so that $\ell_i = \ell_N - \ell_P$. We next put $b_j = B_{i,j}$, where B is defined as in the statement of the theorem. Thus

$$b_j = \begin{cases} \ell_j & \text{if } j \in P, \\ 0 & \text{if } j \in N, \\ -\ell_P & \text{if } j = i. \end{cases}$$

A direct calculation shows that

$$\ell_j - b_j = \begin{cases} 0 & \text{if } j \in P, \\ \ell_j & \text{if } j \in N, \\ \ell_N & \text{if } j = i. \end{cases}$$

Therefore

$$\|\ell - b\|_1 = 2\ell_N.$$

9

Finally given $G \in \mathcal{G}$ we define $g_j = G_{i,j}$ for all j. We have

$$
\begin{aligned}
\|\ell - g\|_1 &= \sum_{j \neq i} |\ell_j - g_j| + \left| \sum_{j \neq i} (\ell_j - g_j) \right| \\
&\geq \sum_{j \in P} |\ell_j - g_j| + \sum_{j \in N} |\ell_j - g_j| \\
&\quad + \left| \sum_{j \in N} (\ell_j - g_j) \right| - \left| \sum_{j \in P} (\ell_j - g_j) \right| \\
&\geq \sum_{j \in P} |\ell_j - g_j| + \ell_N \\
&\quad + \left| \sum_{j \in N} \ell_j \right| - \sum_{j \in P} |\ell_j - g_j| \\
&= 2\ell_N.
\end{aligned}
$$

$\blacksquare$

Theorem 13 *Let A be a Markov matrix such that $\mathrm{Spec}(A) \cap (-\infty, 0] = \emptyset$ and put $L = \log(A)$. If $B \in \mathcal{G}$ and $\|L - B\| = \varepsilon$ then*

$$
\|A - \mathrm{e}^B\| \leq \min\{2, \mathrm{e}^\varepsilon - 1\} \leq \min\{2, 2\varepsilon\}.
$$

Proof If we put $E = L - B$ then the series expansion

$$
\mathrm{e}^L = \mathrm{e}^B + \int_{t=0}^1 \mathrm{e}^{B(1-t)} E \mathrm{e}^{Bt}\, \mathrm{d}t + \int_{t=0}^1 \int_{s=0}^t \mathrm{e}^{B(1-t)} E \mathrm{e}^{B(t-s)} E \mathrm{e}^{Bs}\, \mathrm{d}s\mathrm{d}t + \ldots
$$

given in [5, Theorem 11.4.1] yields

$$
\begin{aligned}
\|A - \mathrm{e}^B\| &= \|\mathrm{e}^L - \mathrm{e}^B\| \\
&\leq \|E\| + \|E\|^2/2! + \ldots \\
&= \mathrm{e}^{\|E\|} - 1.
\end{aligned}
$$

The other part of the estimate uses $\|A\| = 1$ and $\|\mathrm{e}^B\| = 1$. $\blacksquare$

4   Some Numerical Examples

Example 14

The Markov matrix

$$
A = \begin{bmatrix} 0.30 & 0.45 & 0.25 \\ 0.14 & 0.84 & 0.02 \\ 0.14 & 0.52 & 0.34 \end{bmatrix}.
$$

10

has eigenvalues $1, 0.32, 0.16$, exactly. The matrix $L = \log(A)$ is given to four decimal places by

$$
L = \begin{bmatrix} -1.5272 & 0.5991 & 0.9281 \\ 0.3054 & -0.2371 & -0.0683 \\ 0.3054 & 0.9023 & -1.2078 \end{bmatrix},
$$

and has a negative off-diagonal entry. The closest Markov generator B to L as described above is

$$
B = \begin{bmatrix} -1.5272 & 0.5991 & 0.9281 \\ 0.3054 & -0.3054 & 0 \\ 0.3054 & 0.9023 & -1.2078 \end{bmatrix}.
$$

and the embeddable Markov matrix $\widetilde{A} = \mathrm{e}^{B}$ (where B is entered to full precision) is given by

$$
\widetilde{A} = \begin{bmatrix} 0.3000 & 0.4383 & 0.2617 \\ 0.1400 & 0.8046 & 0.0554 \\ 0.1400 & 0.5057 & 0.3543 \end{bmatrix}.
$$

One observes that all the entries of $A - \widetilde{A}$ are less than 0.036 in absolute value. ∎

The following exactly soluble example illustrates the use of some of our theorems.

Theorem 15 *Let*

$$
L_s = \begin{bmatrix} -1-s & 1 & s \\ s & -1-s & 1 \\ 1 & s & -1-s \end{bmatrix}
$$

where $s \in \mathbf{R}$, and let $A_s = \mathrm{e}^{L_s}$. Then

1. *If $s \geq 0$ then A_s is an embeddable Markov matrix.*

2. *If $s < \sigma \sim -0.5712$ then A_s has at least one negative entry.*

3. *If $\sigma \leq s < 0$ then A_s is Markov but not embeddable.*

Proof We first note that $L_s 1 = 0$ for every $s \in \mathbf{R}$, so $A_s 1 = 1$.

Item 1 follows from the fact that L_s satisfies all the conditions for the generator of a Markov semigroup.

To prove item 2 we note that $L_s = -(1+s)I + F + sB$ where F, B are permutation matrices that commute. Let $\mathcal{S}$ be the set of all s such that e^{L_s} is non-negative. If $t \in \mathcal{S}$ and $s \geq t$ then

$$
L_s = L_t + (s-t)B - (s-t)I
$$

11

where all the matrices commute, and

$$\mathrm{e}^{L_s} = \mathrm{e}^{-(s-t)}\mathrm{e}^{L_t}\mathrm{e}^{B(s-t)} \geq 0,$$

so $s \in \mathcal{S}$. Therefore $\mathcal{S}$ is an interval, which is obviously closed. Numerical calculations show that the smallest number $\sigma \in \mathcal{S}$ is approximately -0.5712. More rigorously if $s < -1$ then

$$\det(A_s) = \mathrm{e}^{\mathrm{tr}(L_s)} = \mathrm{e}^{-3(1+s)} > 1$$

so A_s cannot be a Markov matrix and must have a negative entry. This establishes that $\sigma \geq -1$.

Clearly L_s is not a Markov generator if $\sigma \leq s < 0$. We prove item 3 by obtaining a contradiction from the existence of a Markov generator B_s with $\mathrm{e}^{B_s} = A_s$. Since

$$\exp(\mathrm{Spec}(L_s)) = \mathrm{Spec}(A_s) = \exp(\mathrm{Spec}(B_s))$$

we conclude that every eigenvalue of L_s differs from an eigenvalue of B_s by an integral multiple of $2\pi i$. A direct computation shows that

$$\mathrm{Spec}(L_s) = \left\{ 0, -\frac{3(1+s)}{2} \pm \frac{\sqrt{3}(1-s)i}{2} \right\}.$$

For s in the stated range, each non-zero eigenvalue λ of L_s satisfies $|\arg(\lambda)| < 5\pi/6$ and the same applies if one adds an integral multiple of $2\pi i$ to the eigenvalue. Hence each non-zero eigenvalue λ of B_s satisfies $|\arg(\lambda)| < 5\pi/6$ and (5) implies that B_s cannot be a generator. ∎

Example 16

The following illustrates the difficulties in dealing with Markov matrices that have negative eigenvalues. If $c = 2\pi/\sqrt{3}$ and

$$B = c \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix},$$

then the eigenvalues of B are $0, -\sqrt{3}\pi \pm \pi i$. The matrix $A = \mathrm{e}^B$ is self-adjoint with eigenvalues $1, -\mathrm{e}^{-\sqrt{3}\pi}, -\mathrm{e}^{-\sqrt{3}\pi}$. If one uses Matlab's 'logm' command to compute $\log(A)$, one obtains a matrix with complex entries that is not close to B; it might be considered that 'logm' should produce a real logarithm of a real matrix if one exists, but it is not easy to see how to achieve this.

Example 17

We continue with the above example, but with the more typical choice $c = 4$. The eigenvalues of B are now 0, $-6\pm3.4641i$. Clearly $A = \mathrm{e}^B$ is an embeddable Markov matrix. If one rounds to four digits one obtains

$$A \sim \begin{bmatrix} 0.3318 & 0.3337 & 0.3346 \\ 0.3346 & 0.3318 & 0.3337 \\ 0.3337 & 0.3346 & 0.3318 \end{bmatrix},$$

The use of 'logm' yields

$$\log(A) \sim \begin{bmatrix} -4 & 0.3724 & 3.6276 \\ 3.6276 & -4 & 0.3724 \\ 0.3724 & 3.6276 & -4 \end{bmatrix},$$

which is also a Markov generator. We conclude that A is an embeddable Markov matrix in (at least) two distinct ways.

This is not an instance of a general phenomenon. If one defines the 5×5 cyclic matrix B by

$$B_{r,s} = \begin{cases} -4 & \text{if } r = s, \\ 4 & \text{if } r + 1 = s, \\ 4 & \text{if } r = 5,\ s = 1, \\ 0 & \text{otherwise,} \end{cases}$$

then B is a Markov generator with eigenvalues 0, $-7.2361\pm2.3511i$, $-2.7639\pm 3.8042i$. However $L = \log(\exp(B))$, with the principal choice of the matrix logarithm, is a cyclic matrix with some negative off-diagonal entries, so it cannot be a Markov generator. ∎

References

[1] K L Chung, Markov chains with stationary transition probabilities, Springer, Berlin, 1960.

[2] J R Cuthbert, On uniqueness of the logarithm for Markov semigroups, J. London Math. Soc. 4 (1972) 623-630.

[3] J R Cuthbert, The logarithm function for finite-state Markov semi-groups, J. London Math. Soc. (2), 6 (1973) 524-532.

[4] E B Davies, Triviality of the peripheral point spectrum, J. Evol. Eqns. 5 (2005) 407-415.

[5] E B Davies, Linear Operators and Their Spectra, Cambridge Univ. Press, 2007.

[6] N Dunford and J T Schwartz, Linear Operators, Part 1, Interscience Publ., New York, 1966.

[7] G Elfving, Zur Theorie der Markoffschen Ketten, Acta Soc. Sci. Fennicae, n. Ser. A2 no. 8, (1937) 1-17.

[8] F R Gantmacher, The Theory of Matrices, vol. 1, Chelsea, New York, 1959.

[9] N J Higham, Functions of Matrices: Theory and Computation, SIAM, Philadelphia, PA, USA, 2008.

[10] N J Higham and L Lin, On pth roots of stochastic matrices. MIMS preprint, 2009.

[11] R B Israel, J S Rosenthal, J Z Wei, Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings, Math. Finance 11, no. 2 (2001) 245-265.

[12] S Johansen, A central limit theorem for finite semi-groups and its application to the imbedding problem for finite-state Markov chains, Zeit. Wahrsch. 26 (1973) 171-90.

[13] F I Karpelevič, On the characteristic roots of matrices with nonnegative elements, Izv. Akad. Nauk SSSR Ser. Mat. 15 (1951) 361-383 (in Russian), Amer. Math. Soc. Transl. (2) 140 (1988) 79-100.

[14] J F C Kingman, The imbedding problem for finite Markov chains, Zeit. Wahrsch. 1 (1962) 14-24.

[15] A Kreinin and M Sidelnikova, Regularization algorithms for transition matrices, Algo Research Quarterly, 4 (2001) 23-40.

[16] H Minc, Nonnegative Matrices, Wiley, New York, 1988.

[17] R Nagel, One-Parameter Semigroups for Positive Operators, Lect. Notes Math. 1184, Springer, New York, 1986.

[18] J Th Runnenberg 1962. On Elfving's problem of imbedding a time-discrete Markov chain in a continuous time one for finitely many states, Proc. Koninklijke Nederlandse Akademie van Wetenschappen, Ser. A, Math. Sci. 65 (5) (1962) 536-41.

[19] B Singer, S Spilerman, The representation of social processes by Markov models, Amer. J. Sociology, 82 (1976) 1-54.

[20] J M O Speakman, Two Markov chains with a common skeleton, Zeit. Wahrsch. 7 (1967) 224.

[21] S Zahl, A Markov process model for follow-up studies, Human Biology, 27 (1955) 90-120.

Department of Mathematics
King's College
Strand
London
WC2R 2LS
UK

E.Brian.Davies@kcl.ac.uk