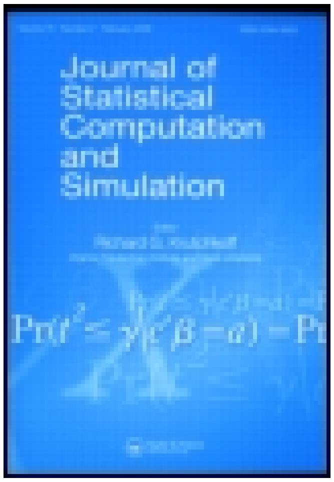

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

Computation of the stationary distribution of a markov chain

C.C. Paige [a] , George P.H. Styan [a] & Peter G. Wachter [a]
[a] McGill University , P.O.Box6070, Montreal , Quebec, Canada
Published online: 20 Mar 2007.

J. Statist. Comput. Simul., 1975, Vol. 4, pp. 173–186

Computation of the Stationary Distribution of a Markov Chain

C. C. PAIGE, GEORGE P. H. STYAN and PETER G. WACHTER†
McGill University, P.O. Box 6070, Montreal, Quebec, Canada.

(*Received November 27, 1973*)

Eight algorithms are considered for the computation of the stationary distribution $\mathbf{l}'$ of a finite Markov chain with associated probability transition matrix $\mathbf{P}$. The recommended algorithm is based on solving $\mathbf{l}'(\mathbf{I}-\mathbf{P}+\mathbf{eu}') = \mathbf{u}'$, where $\mathbf{e}$ is the column vector of ones and $\mathbf{u}'$ is a row vector satisfying $\mathbf{u}'\mathbf{e} \neq 0$. An error analysis is presented for any such $\mathbf{u}$ including the choices $\mathbf{u}' = \mathbf{e}_j\mathbf{P}$ and $\mathbf{u}' = \mathbf{e}'_j$, where $\mathbf{e}'_j$ is the jth row of the identity matrix. Computational comparisons between five of the algorithms are made based on twenty 8×8, twenty 20×20, and twenty 40×40 transition matrices. The matrix $(\mathbf{I}-\mathbf{P}+\mathbf{eu}')^{-1}$ is shown to be a non-singular generalized inverse of $\mathbf{I}-\mathbf{P}$ when the unit root of $\mathbf{P}$ is simple and $\mathbf{u}'\mathbf{e} \neq 0$. A simple closed form expression is obtained for the Moore-Penrose inverse of $\mathbf{I}-\mathbf{P}$ when $\mathbf{I}-\mathbf{P}$ has nullity one.

KEY WORDS and PHRASES: Stationary distribution, Markov chain, Generalized inverses, Numerical algebra, Rounding error analysis, Matrix theory, Stochastic processes, Statistical computation.

INTRODUCTION AND SUMMARY

The stationary distribution of a finite Markov chain with stationary transition probabilities is of widespread interest (Chung, 1960, Golub, Seneta, 1973, Kemeny, Snell, 1960). If $\mathbf{P}$ is the $n \times n$ stochastic matrix of transition probabilities, then the $1 \times n$ row vector $\mathbf{l}'$ denoting the associated stationary distribution satisfies

$$\mathbf{l}'\mathbf{P} = \mathbf{l}' \text{ and } \mathbf{l}'\mathbf{e} = 1, \tag{1.1}$$

where $\mathbf{e}$ is the $n \times 1$ column vector of ones. Moreover

$$\mathbf{Pe} = \mathbf{e}, \tag{1.2}$$

†Presently with Bell Canada, Montreal.

and so $\mathbf{l}'$ is a normalized left-hand characteristic vector of a matrix corresponding to a given (unit) characteristic root with right-hand characteristic vector specified (e). The Perron-Frobenius theorem (Marcus, Minc, 1964) assures that at least one $\mathbf{l}'$ with nonnegative elements satisfies (1.1).

In this paper we discuss (Section 2.1) the computation of $\mathbf{l}'$ based on

$$\mathbf{l}'(\mathbf{I}-\mathbf{P}+\mathbf{eu}') = \mathbf{u}'; \tag{1.3}$$

we show that an excellent choice of $\mathbf{u}'$ is a row of $\mathbf{P}$. We compare methods based on (1.3) with others (Section 2.2), including two which have been published (Decell, Odell, 1967, Albert, 1972). Error analyses and some computational results comparing the methods are also presented (Section 2.3 and 2.4). Further details are given by Wachter (1973).

An important ingredient of the recommended algorithms is the result (Lemma 2.1) that the characteristic roots of $\mathbf{I}-\mathbf{P}+\mathbf{eu}'$ are $\mathbf{u}'\mathbf{e}$ and $1-\lambda_2, \ldots, 1-\lambda_n$, where $1 = \lambda_1, \lambda_2, \ldots, \lambda_n$ are the characteristic roots of $\mathbf{P}$. Hence $\mathbf{I}-\mathbf{P}+\mathbf{eu}'$ is nonsingular if and only if $\mathbf{u}'\mathbf{e} \neq 0$ and the unit root of $\mathbf{P}$ is simple (this is assured when the associated Markov chain is irreducible). Then $(\mathbf{I}-\mathbf{P}+\mathbf{eu}')^{-1}$ is a nonsingular generalized inverse of $\mathbf{I}-\mathbf{P}$; when $\mathbf{u} = \mathbf{l}$ the matrix $\mathbf{Z} = (\mathbf{I}-\mathbf{P}+\mathbf{el}')^{-1}$ is the *fundamental matrix* introduced by Kemeny and Snell (1960). See also Hunter (1969). The Moore-Penrose inverse of $\mathbf{I}-\mathbf{P}$ is shown (Section 3) to be

$$(\mathbf{I}-\mathbf{P})^{+} = (\mathbf{I}-\mathbf{P}+\alpha\mathbf{le}')^{-1}-\alpha\mathbf{el}', \tag{1.4}$$

where

$$\alpha = (n\mathbf{l}'\mathbf{l})^{-1/2}. \tag{1.5}$$

Note the reversed position of e in (1.4) and (1.3).

THE ALGORITHMS

2.1 Recommended Algorithm

Let $\mathbf{P}$ be an $n \times n$ stochastic matrix satisfying

$$\mathbf{Pe} = \mathbf{e}, \tag{2.1.1}$$

where $\mathbf{e}$ is the $n \times 1$ column vector of ones. We seek the $1 \times n$ row vector $\mathbf{l}'$ satisfying

$$\mathbf{l}'\mathbf{P} = \mathbf{l}' \text{ and } \mathbf{l}'\mathbf{e} = 1. \tag{2.1.2}$$

It follows that

$$\mathbf{l}'(\mathbf{I}-\mathbf{P}+\mathbf{eu}') = \mathbf{u}' \tag{2.1.3}$$

for every $1\times n$ row vector $\mathbf{u}'$. We may compute $\mathbf{l}'$ using (2.1.3) when $\mathbf{I}-\mathbf{P}+\mathbf{eu}'$ is nonsingular. We need the following result (Brauer, 1952).

LEMMA 2.1 *Let the $n\times n$ matrix $\mathbf{A}$ have characteristic roots $0=\alpha_1, \alpha_2, \ldots, \alpha_n$, and suppose $\mathbf{Ae}=\mathbf{0}$. Then $\mathbf{A}+\mathbf{eu}'$ has characteristic roots $\mathbf{u}'\mathbf{e}, \alpha_2, \ldots, \alpha_n$, where $\mathbf{u}'$ is any $1\times n$ row vector.*

Proof The characteristic polynomial of $\mathbf{A}+\mathbf{eu}'$ is given by the determinant

$$|\mathbf{A}+\mathbf{eu}'-\lambda\mathbf{I}| = |(\mathbf{A}-\lambda\mathbf{I})(\mathbf{I}+(\mathbf{A}-\lambda\mathbf{I})^{-1}\mathbf{eu}')| = |\mathbf{A}-\lambda\mathbf{I}|\cdot(1+\mathbf{u}'(\mathbf{A}-\lambda\mathbf{I})^{-1}\mathbf{e}), \tag{2.1.4}$$

as $|\mathbf{I}+\mathbf{ab}'| = 1+\mathbf{b}'\mathbf{a}$ for any $n\times 1$ vectors $\mathbf{a}$ and $\mathbf{b}$. Since $\mathbf{Ae}=\mathbf{0}$, it follows that $(\mathbf{A}-\lambda\mathbf{I})\mathbf{e} = -\lambda\mathbf{e}$ and so $(\mathbf{A}-\lambda\mathbf{I})^{-1}\mathbf{e} = -\mathbf{e}/\lambda$ for any λ not a characteristic root of $\mathbf{A}$. Hence (2.1.4) yields

$$|\mathbf{A}+\mathbf{eu}'-\lambda\mathbf{I}| = (1-\mathbf{u}'\mathbf{e}/\lambda)\prod_{j=1}^{n}(\alpha_j-\lambda) = (\mathbf{u}'\mathbf{e}-\lambda)\prod_{j=2}^{n}(\alpha_j-\lambda), \tag{2.1.5}$$

which completes the proof.

It follows directly from Lemma 2.1 that if the unit characteristic root of $\mathbf{P}$ is simple then $\mathbf{I}-\mathbf{P}+\mathbf{eu}'$ is nonsingular if and only if the sum $\mathbf{u}'\mathbf{e}\neq 0$.

To solve (2.1.3) for $\mathbf{l}'$, we may, therefore, choose any $\mathbf{u}'$ such that $\mathbf{u}'\mathbf{e}\neq 0$, provided the unit root of $\mathbf{P}$ is simple (this is assured when the associated Markov chain is irreducible). Since the roots of $\mathbf{I}-\mathbf{P}+\mathbf{eu}'$ other than $\mathbf{u}'\mathbf{e}$ have maximum absolute value 2 we feel that a choice of $\mathbf{u}'$ should have the absolute value of the sum $\mathbf{u}'\mathbf{e}$ less than 2.

Our *recommended choice* (Golub, 1970, Tubilla, 1970) for $\mathbf{u}'$ is

$$\mathbf{u}' = \mathbf{e}_j'\mathbf{P}, \tag{2.1.6}$$

where $\mathbf{e}_j'$ is the $1\times n$ jth row of $\mathbf{I}_n$ and $j=1, 2, \ldots$, or n. This choice has the advantage that the jth row of $\mathbf{I}-\mathbf{P}+\mathbf{eu}' = \mathbf{I}-\mathbf{P}+\mathbf{ee}_j'\mathbf{P}$ is $\mathbf{e}_j'$, for

$$\mathbf{e}_j'(\mathbf{I}-\mathbf{P}+\mathbf{eu}') = \mathbf{e}_j'-\mathbf{e}_j'\mathbf{P}+\mathbf{u}' = \mathbf{e}_j'. \tag{2.1.7}$$

Another choice for $\mathbf{u}'$ is $\mathbf{e}_j'$, as suggested by Styan (1964). Again $j=1, 2, \ldots$, or n. Other choices for $\mathbf{u}'$ are considered by Wachter (1973).

We suggest using Gaussian elimination with pivoting to solve (2.1.3).

2.2 Other Algorithms

Decell and Odell (1967) suggested computing $\mathbf{l}'$ as

$$\mathbf{l}' = \mathbf{e}'[\mathbf{I}-(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})^{-}]/\mathbf{e}'[\mathbf{I}-(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})^{-}]\mathbf{e}, \tag{2.2.1}$$

where $(\mathbf{I}-\mathbf{P})^-$ is any generalized inverse [cf. (3.1)] of $\mathbf{I}-\mathbf{P}$ such that $(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})^-$ is symmetric. Decell and Odell (1967) suggested computing $(\mathbf{I}-\mathbf{P})^-$ explicitly; we do not recommend this, as the symmetric projector

$$(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})^- = \sum_{j=1}^{n} \mathbf{b}_j\mathbf{b}_j' \tag{2.2.2}$$

may be computed faster than any procedure we could find to compute any generalized inverse. The column vectors $\mathbf{b}_j$ form an orthonormal basis for the column space of $\mathbf{I}-\mathbf{P}$. An excellent way of computing (2.2.2) is the modified Gram-Schmidt orthonormalization procedure, cf. Golub (1969), for example.

Assuming that $\mathbf{I}-\mathbf{P}$ has rank $n-1$, the symmetric projector (see also Brauer, 1952, Keilson, Styan, 1973)

$$\mathbf{I}-(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})^- = \mathbf{l}\mathbf{l}'/\mathbf{l}'\mathbf{l}, \tag{2.2.3}$$

from which (2.2.1) follows immediately. To verify (2.2.3) note that

$$(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})^- + \mathbf{l}\mathbf{l}'/\mathbf{l}'\mathbf{l} = \mathbf{I}_n; \tag{2.2.4}$$

the two projectors on the left-hand side have ranks $n-1$ and 1, respectively, and their product is $\mathbf{0}$. Their sum is, therefore, idempotent of rank $n = (n-1)+1$, and so equals $\mathbf{I}_n$.

Albert (1972, p. 37) proposed computing $\mathbf{l}'$ by a rank reduction procedure. Let the $n \times n$ matrix

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = \mathbf{I}-\mathbf{P}. \tag{2.2.5}$$

Form, recursively, for $k = 1, 2, \ldots, n-1$,

$$\mathbf{A}_k = \mathbf{A}_{k-1} - \mathbf{A}_{k-1}\mathbf{a}_k\mathbf{a}_k'\mathbf{A}_{k-1}/\mathbf{a}_k'\mathbf{A}_{k-1}\mathbf{a}_k, \tag{2.2.6}$$

where $\mathbf{A}_0 = \mathbf{I}_n$. Then

$$\mathbf{A}_{n-1} = \mathbf{l}\mathbf{l}'/\mathbf{l}'\mathbf{l}. \tag{2.2.7}$$

To see this, note first that with $k = 1$,

$$\mathbf{A}_1 = \mathbf{I}_n - \mathbf{a}_1\mathbf{a}_1'/\mathbf{a}_1'\mathbf{a}_1 \tag{2.2.8}$$

is symmetric idempotent of rank $n-1$. Since $\mathbf{A}_1\mathbf{a}_1 = \mathbf{0}$ and $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n-1}$ are linearly independent, it follows that $\mathbf{a}_2'\mathbf{A}_1\mathbf{a}_2 \neq 0$. Hence

$$\mathbf{A}_2 = \mathbf{A}_1 - \mathbf{A}_1\mathbf{a}_2\mathbf{a}_2'\mathbf{A}_1/\mathbf{a}_2'\mathbf{A}_1\mathbf{a}_2 \tag{2.2.9}$$

is symmetric idempotent rank $n-2$. Proceeding recursively leads to $\mathbf{A}_{n-1}$ symmetric idempotent of rank 1. Clearly $\mathbf{A}_k\mathbf{a}_j = \mathbf{0}$ for $j = 1, 2, \ldots, k$ and

$k = 1, 2, \ldots, n-1$. Hence $\mathbf{A}_{n-1}\mathbf{A} = \mathbf{A}_{n-1}(\mathbf{I}-\mathbf{P}) = \mathbf{0}$ and (2.2.7) follows.

Golub (1970) suggested the following least-squares procedure for computing $\mathbf{l}$. Write

$$\mathbf{X}\mathbf{l} = \begin{pmatrix} \mathbf{I}-\mathbf{P}' \\ \mathbf{e}' \end{pmatrix} \mathbf{l} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{2.2.10}$$

where $\mathbf{X}$ is $(n+1) \times n$. We may transform $\mathbf{X}$ to upper triangular form $\mathbf{T}$ using n Householder transformations assembled as a product in the orthogonal matrix $\mathbf{Q}$. Thus

$$\mathbf{Q}'\mathbf{X} = \begin{pmatrix} \mathbf{T} \\ \mathbf{0}' \end{pmatrix}. \tag{2.2.11}$$

The least-squares solution $\mathbf{l} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}_{n+1}$ may be found by solving the triangular system of equations:

$$\mathbf{T}\mathbf{l} = (\mathbf{I}_n, \mathbf{0})\mathbf{Q}'\mathbf{e}_{n+1}. \tag{2.2.12}$$

Note that $\mathbf{T}$ is $n \times n$, while $\mathbf{Q}$ is $(n+1) \times (n+1)$. Because of the special structure of $\mathbf{X}$ the residual sum of squares is zero.

Three other more naïve methods suggest themselves.

Let $(\mathbf{I}-\mathbf{P})_j$ be the $n \times n$ matrix $\mathbf{I}-\mathbf{P}$ with its jth column replaced by $\mathbf{e}$. Then $\mathbf{l}'$ may be found by solving the $n \times n$ system

$$\mathbf{l}'(\mathbf{I}-\mathbf{P})_j = \mathbf{e}_j', \tag{2.2.13}$$

for any $j = 1, 2, \ldots, n$. This works since

$$(\mathbf{I}-\mathbf{P})_j = (\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{e}_j\mathbf{e}_j') + \mathbf{e}\mathbf{e}_j' \tag{2.2.14}$$

is nonsingular. To prove this we need:

LEMMA 2.2 *Let the $n \times n$ matrix $\mathbf{A}$ have rank $n-1$. Then*

$$|\mathbf{A}+\mathbf{f}\mathbf{u}'| = \mathbf{u}'(\operatorname{adj} \mathbf{A})\mathbf{f}, \tag{2.2.15}$$

where $\mathbf{f}$ and $\mathbf{u}$ are any $n \times 1$ vectors and adj $\mathbf{A}$ *is the adjugate matrix of first cofactors transposed.*

Lemma 2.2 is proved by Bodewig (1959, p. 42). When $\mathbf{A} = \mathbf{I}-\mathbf{P}$ then adj $\mathbf{A} = k\mathbf{e}\mathbf{l}'$, where

$$k = \prod_{j=2}^{n} (1-\lambda_j) \neq 0; \tag{2.2.16}$$

the λ_j are the nonunit characteristic roots of $\mathbf{P}$. Bounds for k are given by Kielson and Styan, (1973). The representation for adj $\mathbf{A}$ as $k\mathbf{el}'$ is possible since adj $\mathbf{A}$ commutes with $\mathbf{A}$, and $\mathbf{A}$ has rank $n-1$. Hence

$$\begin{aligned} |(\mathbf{I}-\mathbf{P})_j| &= |\mathbf{I}-\mathbf{P}+(\mathbf{e}-\mathbf{e}_j+\mathbf{Pe}_j)\mathbf{e}'_j| = k\mathbf{e}'_j(\mathbf{el}')(\mathbf{e}-\mathbf{e}_j+\mathbf{Pe}_j) \\ &= \prod_{j=2}^{n}(1-\lambda_j) \neq 0. \end{aligned} \tag{2.2.17}$$

Two other naïve methods involve successive multiplication or powering. When the unit characteristic root of $\mathbf{P}$ is primitive ($|\lambda_j| < 1; j = 2, \ldots, n$) then (Kemeny, Snell, 1960, p. 128)

$$\mathbf{P}^r \to \mathbf{el}' \text{ as } r \to \infty. \tag{2.2.18}$$

The associated Markov chain is then called regular (Kemeny, Snell, 1960) or ergodic (Chung, 1960). If $\mathbf{u}'$ is any $1 \times n$ row vector such that $\mathbf{u}'\mathbf{e} = 1$ then we may form

$$\mathbf{u}'\mathbf{P}, (\mathbf{u}'\mathbf{P})\mathbf{P}, (\mathbf{u}'\mathbf{P}^2)\mathbf{P}, \ldots \tag{2.2.19}$$

which converges to $\mathbf{l}'$. Alternatively we may form

$$\mathbf{P}, \mathbf{P}^2, (\mathbf{P}^2)^2, (\mathbf{P}^4)^2, \ldots \tag{2.2.20}$$

which converges to $\mathbf{el}'$.

To summarize, we have considered these eight algorithms:

1) $\mathbf{l}'(\mathbf{I}-\mathbf{P}+\mathbf{ee}'_j\mathbf{P}) = \mathbf{e}'_j\mathbf{P}$,

2) $\mathbf{l}'(\mathbf{I}-\mathbf{P}+\mathbf{ee}'_j) = \mathbf{e}'_j$,

3) $\mathbf{l}' = \mathbf{e}'[\mathbf{I}-(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})^-]/\mathbf{e}'[\mathbf{I}-(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})^-]\mathbf{e}$,

4) rank reduction, cf. (2.2.5) through (2.2.9),

5) least squares, cf. (2.2.10) through (2.2.12),

6) $\mathbf{l}'[(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{e}_j\mathbf{e}'_j)+\mathbf{ee}'_j] = \mathbf{e}'_j$,

7) $\mathbf{u}'\mathbf{P}, (\mathbf{u}'\mathbf{P})\mathbf{P}, (\mathbf{u}'\mathbf{P}^2)\mathbf{P}, \ldots$,

8) $\mathbf{P}, \mathbf{P}^2, (\mathbf{P}^2)^2, (\mathbf{P}^4)^2, \ldots$.

Section 2.3 presents an error analysis of algorithms 1 and 2, while Section 2.4 gives computational comparisons between algorithms 1 through 5. We note that algorithm 1 involves the solution of $n-1$ simultaneous linear equations in $n-1$ unknowns, while algorithms 2 and 6 need one more equation and one more unknown. Algorithms 7 and 8 require respectively n^2 and n^3 operations per iteration and could be very slow, if the absolute value of the second root of $\mathbf{P}$ is close to 1. Algorithms 6–8 are, therefore, not considered further in this paper.

2.3 Error Analysis

A characteristic vector $\mathbf{l}$ satisfying (2.1.2) could be computed using inverse iteration (Wilkinson, 1965, p. 619). This approach would probably be taken by a numerical analyst, since direct solutions are usually unreliable [p. 316]. The danger occurs when the corresponding right-hand characteristic vector has a particular component which is small relative to any of the others [p. 627]. Here the right-hand characteristic vector $\mathbf{e}$ has equal components and it will turn out that using (2.1.3) is a numerically stable method for all $\mathbf{u}$ for which $\mathbf{u}'\mathbf{e}$ is not too small and $\mathbf{u}$ is not too large. The two choices of $\mathbf{e}_j'\mathbf{P}$ and $\mathbf{e}_j'$ for $\mathbf{u}'$ suggested in Section 2.1 are therefore ideal.

When (2.1.3) is solved using Gaussian elimination with pivoting, it can be shown (Wilkinson, 1963, p. 108) that the computed $\mathbf{l}$ satisfies

$$\mathbf{l}'(\mathbf{I}-\mathbf{P}+\mathbf{e}\mathbf{u}'+\mathbf{E}) = \mathbf{u}', \tag{2.3.1}$$

where the equivalent rounding error matrix $\mathbf{E}$ takes account of the error in forming $\mathbf{I}-\mathbf{P}+\mathbf{e}\mathbf{u}'$ and the error in solving the resulting equations. Various bounds can be obtained for $\mathbf{E}$, but here it is only necessary to realize that

$$\|\mathbf{E}\|_\infty \leqq k_1\varepsilon\|\mathbf{I}-\mathbf{P}+\mathbf{e}\mathbf{u}'\|_\infty \leqq k_1\varepsilon(2+\|\mathbf{u}'\|_\infty), \tag{2.3.2}$$

where $\|\mathbf{A}\|_\infty \equiv \max_i \Sigma_j|a_{ij}|$ is a suitable norm, ε is the relative precision of floating point computation, and k_1 is a factor dependent on the dimension of the matrix and the growth of elements in the decomposition. In common practice k_1 is quite small (p. 97), and for this reason Gaussian elimination with pivoting is recommended in the literature as a numerically stable algorithm.

Applying each side of (2.3.1) to $\mathbf{e}$ and using (2.1.1) gives

$$\mathbf{l}'\mathbf{e} = 1-\mathbf{l}'\mathbf{E}\mathbf{e}/\mathbf{u}'\mathbf{e} = 1-\varepsilon_1, \tag{2.3.3}$$

say, so that

$$\begin{aligned}\mathbf{l}'(\mathbf{I}-\mathbf{P}) &= (1-\mathbf{l}'\mathbf{e})\mathbf{u}'-\mathbf{l}'\mathbf{E}\\ &= \mathbf{l}'\mathbf{E}(\mathbf{e}\mathbf{u}'/\mathbf{u}'\mathbf{e}-\mathbf{I})\\ &= \mathbf{r}',\end{aligned} \tag{2.3.4}$$

say.

We see that small $\mathbf{u}'\mathbf{e}$ could cause large ε_1 and $\mathbf{r}$, while large $\mathbf{u}$ could cause a large $\mathbf{r}$. The choices of $\mathbf{e}_j'\mathbf{P}$ and $\mathbf{e}_j'$ for $\mathbf{u}'$ both give $\mathbf{u}'\mathbf{e} = 1$ and $\|\mathbf{u}'\|_\infty = 1$, so these are ideal, and we need only consider these further. For both these choices rearranging (2.3.4) gives

$$\mathbf{l}' = \mathbf{l}'[\mathbf{P}+\mathbf{E}(\mathbf{e}\mathbf{u}'-\mathbf{I})], \tag{2.3.5}$$

and using (2.3.2),

$$\|\mathbf{E}(\mathbf{e}\mathbf{u}'-I)\|_\infty \leqq 6k_1\varepsilon, \tag{2.3.6}$$

so that $\mathbf{l}'$ and 1 are a left-hand characteristic vector-root pair for a matrix very close to $\mathbf{P}$. From (2.3.3), using (2.3.2), we see that

$$|\varepsilon_1| \leqq 3k_1\varepsilon\|\mathbf{l}'\|_\infty, \tag{2.3.7}$$

so that $\mathbf{l}'\mathbf{e}$ will be roughly as close to unity as we can expect. If, as hoped, all elements of the computed $\mathbf{l}$ are nonnegative, then from (2.3.3) and (2.3.7),

$$\|\mathbf{l}'\|_\infty = \mathbf{l}'\mathbf{e} \leqq (1-3k_1\varepsilon)^{-1}; \tag{2.3.8}$$

but some ill-conditioned cases could give small negative elements in $\mathbf{l}$, thus destroying this bound. In fact, if there is another characteristic root of $\mathbf{P}$ very close to 1 then $\mathbf{l}$ can be strongly contaminated by the corresponding characteristic vector. Despite such possible ill-condition we still see from (2.3.5) and (2.3.6) that these are *numerically stable algorithms*, since the best we can ever hope to do is to compute the characteristic vector of a matrix very close to $\mathbf{P}$. To emphasize this we note that each element of the true $\mathbf{P}$ will usually have to be rounded for storage in the computer.

2.4 Computational Comparisons

To compare the computational speed and accuracy of the various algorithms we generated twenty 8×8 stochastic matrices and computed the associated stationary distribution using five of the algorithms described in Sections 2.1 and 2.2. We compared the average times (and associated standard errors) needed to compute $\mathbf{l}'$ to a certain accuracy and repeated the procedure with 20×20 and 40×40 matrices. This enabled us to compare the algorithms in terms of speed (cf. Table I).

As a measure of accuracy we considered the residual error

$$\rho = [(\hat{\mathbf{l}}'-\hat{\mathbf{l}}'\mathbf{P})(\hat{\mathbf{l}}-\mathbf{P}'\hat{\mathbf{l}})]^{1/2}, \tag{2.4.1}$$

where $\hat{\mathbf{l}}'$ was the computed solution. If $\hat{\mathbf{l}}'$ is the exact solution ρ would be zero. All calculations were performed in double precision arithmetic, using the FORTRAN IV, level G compiler, on the IBM 360/75 computer at McGill University.

The comparison of average times shows that the fastest method was our recommended algorithm (cf. Table I). The procedure we develop from the suggestion of Decell and Odell (1967), while comparable for small matrices, seems to take relatively more time for the larger matrices. The initial comparison of residual errors indicates our algorithm is the most accurate.

TABLE I

		Recommended Algorithm $\mathbf{l}'(\mathbf{I}-\mathbf{P}+\mathbf{e}\mathbf{e}_j'\mathbf{P}) = \mathbf{e}_j'\mathbf{P}$	Alternate to Recommended Algorithm $\mathbf{l}'(\mathbf{I}-\mathbf{P}+\mathbf{e}\mathbf{e}_j') = \mathbf{e}_j'$	Modified Decell and Odell (1967)	Rank Reduction (Albert) [1972, p. 37]	Least Squares (Golub, 1970)
Average Computer Time (Standard Errors in Parentheses)	8×8	0.70 (0.57)	0.75 (0.44)	*0.65* (0.49)	1.20 (0.70)	1.20 (0.70)
	20×20	*6.05* (0.60)	7.10 (1.07)	7.30 (1.53)	14.05 (1.43)	10.25 (1.48)
	40×40	*43.8* (2.4)	47.4 (4.0)	58.5 (5.7)	—	64.8 (4.2)
Average Residual Error (Standard Errors in Parentheses)	8×8	*1.4* (0.21)	3.2 (0.39)	*1.4* (0.16)	1.9 (0.26)	1.9 (0.26)
	20×20	*1.3* (0.27)	2.7 (0.88)	4.4 (0.34)	3.8 (0.48)	3.2 (0.40)
	40×40	*1.3* (0.12)	2.9 (0.52)	4.2 (0.22)	—	4.3 (0.51)

Times are in sixtieths of a second (IBM 360/75, Fortran IV, Level G, double precision). Residual error is $[(\hat{\mathbf{l}}'-\hat{\mathbf{l}}'\mathbf{P})(\hat{\mathbf{l}}-\mathbf{P}'\hat{\mathbf{l}})]^{1/2}$; all numbers of order 10^{-16}. Italicized numbers are optimal. Averages taken over 20 matrices.

To further assess numerical stability consider the $2n \times 2n$ matrix

$$Q = \begin{pmatrix} \mathbf{P}_1 & 0 \\ 0 & \mathbf{P}_2 \end{pmatrix}, \tag{2.4.2}$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are stochastic matrices with a simple root of unity. Let

$$\mathbf{Q}_r = \begin{pmatrix} \mathbf{P}_{1r} & \mathbf{E}_r \\ \mathbf{E}_r & \mathbf{P}_{2r} \end{pmatrix}, \tag{2.4.3}$$

where $\mathbf{E}_r$ is an $n \times n$ matrix of zeros except for the value 10^{-r} in the 1, 1 position. The matrices $\mathbf{P}_{1r}$ and $\mathbf{P}_{2r}$ are $\mathbf{P}_1$ and $\mathbf{P}_2$ scaled so that $\mathbf{Q}_r$ is stochastic. As the value of r increases the rank of $\mathbf{I}-\mathbf{Q}_r$ tends to $2n-2$, and it becomes increasingly more difficult to compute (or identify) a unique stationary distribution.

We applied three of the algorithms to such $\mathbf{Q}_r$ with $n = 10$ and found that the modified Decell and Odell algorithm broke down very quickly (cf. Table II), the residual error increasing almost exponentially with r. The two other algorithms shown seem remarkably accurate. An analysis of the solution vectors shows that for both algorithms the computed solution is essentially a linear combination of the stationary distributions associated with $\mathbf{P}_1$ and $\mathbf{P}_2$.

The algorithm we recommend is substantially faster than the others considered and is numerically stable; that is, it gives us as much accuracy as the condition of the problem and the precision of the computer allow.

TABLE II

Recommended Algorithm
$\mathbf{l}'(\mathbf{I}-P+\mathbf{e}\mathbf{e}_j'\mathbf{P}) = \mathbf{e}_j'\mathbf{P}$

r	l_1	$l_2 \ldots l_{19}$		l_{20}	Residual
1	0.0621	0.0617	0.0676	0.0479	10^{-16}
2	0.0571	0.0624	0.0684	0.0485	10^{-15}
3	0.0566	0.0625	0.0684	0.0485	10^{-16}
5	0.0566	0.0625	0.0684	0.0485	10^{-15}
7	0.0566	0.0625	0 0684	0.0485	10^{-15}
9	0.0566	0.0625	0.0684	0.0485	10^{-15}
11	0.0566	0.0624	0 0685	0 0486	10^{-15}
13	0.0558	0.0616	0.0692	0.0491	10^{-15}
14	0.0491	0.0524	0.0760	0.0539	10^{-15}
15	0.0231	0.0254	0.1030	0.0727	10^{-15}
16	0.0038	0.0041	0.1220	0.0866	10^{-15}
17	0.0004	0.0004	0.1260	0.0891	10^{-15}
20	0.0000	0.0000	0.1260	0.0893	10^{-15}

TABLE II (*continued*)

Modified
Decell and Odell

r	l_1	$l_2 \ldots l_{19}$		l_{20}	Residual
1	0.0621	0.0617	0.0676	0.0479	10^{-15}
2	0.0571	0.0624	0.0684	0.0489	10^{-14}
3	0.0566	0.0624	0.0684	0.0485	10^{-13}
5	0.0566	0.0625	0.0684	0.0485	10^{-11}
7	0.0566	0.0625	0.0684	0.0485	10^{-9}
9	0.0566	0.0625	0.0684	0.0485	10^{-7}
11	0.0566	0.0625	0.0684	0.0485	10^{-5}
13	0.0563	0.0621	0.0684	0.0489	10^{-3}
14	0.0547	0.0601	0.0728	0.0516	10^{-2}
15	0.0746	0.0677	0.0817	0.0580	10^{-1}
16	0.0886	0.0912	0.0781	0.0554	10^{-1}
17	0.0918	0.0920	0.0759	0.0538	10^{-1}
20	0.0921	0.0923	0.0759	0.0538	10^{-1}

Least Squares

r	l_1	$l_2 \ldots l_{19}$		l_{20}	Residual
1	0.0621	0.0617	0.0676	0.0479	10^{-15}
2	0.0571	0.0624	0.0684	0.0485	10^{-15}
3	0.0566	0.0625	0.0684	0.0485	10^{-15}
5	0.0566	0.0625	0.0684	0.0485	10^{-15}
7	0.0566	0.0625	0.0684	0.0485	10^{-15}
9	0.0566	0.0625	0.0684	0.0485	10^{-15}
11	0.0566	0.0625	0.0684	0.0485	10^{-15}
13	0.0569	0.0628	0.0681	0.0485	10^{-15}
14	0.0597	0.0659	0.0653	0.0463	10^{-15}
15	0.0854	0.0943	0.0392	0.0278	10^{-15}
16	0.0945	0.1040	0.0299	0.0212	10^{-15}
17	0.0914	0.1010	0.0330	0.0234	10^{-15}
20	0.0928	0.1020	0.0316	0.0224	10^{-15}

3. GENERALIZED INVERSES

Hunter (1969) proved that the fundamental matrix $\mathbf{Z} = (\mathbf{I}-\mathbf{P}+\mathbf{e}\mathbf{l}')^{-1}$ for a regular Markov chain with transition probability matrix $\mathbf{P}$ and stationary distribution $\mathbf{l}'$ is a nonsingular generalized inverse of $\mathbf{I}-\mathbf{P}$. We extend this result in various ways. Following Rao and Mitra (1971) we define $\mathbf{G}$ to be *a generalized inverse* (g-inverse) of $\mathbf{A}$ whenever

$$\mathbf{AGA} = \mathbf{A}. \tag{3.1}$$

THEOREM 3.1 *Let the $n \times n$ matrix $\mathbf{P}$ have a simple characteristic root of unity and satisfy*

$$\mathbf{Pe} = \mathbf{e} \text{ and } \mathbf{l}'\mathbf{P} = \mathbf{l}', \text{ with } \mathbf{l}'\mathbf{e} = 1, \tag{3.2}$$

where $\mathbf{l}$ is an $n \times 1$ vector, and $\mathbf{e}$ is the $n \times 1$ vector of ones. Then for any $n \times 1$ vector $\mathbf{u}$ such that $\mathbf{u}'\mathbf{e} \neq 0$, the matrix $\mathbf{I}-\mathbf{P}+\mathbf{eu}'$ is nonsingular and its inverse is a nonsingular generalized inverse of $\mathbf{I}-\mathbf{P}$.

Proof The nonsingularity of $\mathbf{I}-\mathbf{P}+\mathbf{eu}'$ follows directly from Lemma 2.1; moreover,

$$(\mathbf{I}-\mathbf{P}+\mathbf{eu}')(\mathbf{I}-\mathbf{P}+\mathbf{eu}')^{-1} = \mathbf{I}_n \tag{3.3}$$

implies

$$(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P}+\mathbf{eu}')^{-1} = \mathbf{I}_n-\mathbf{el}', \tag{3.4}$$

using (2.1.3). Postmultiplication of (3.4) by $\mathbf{I}-\mathbf{P}$ completes the proof.

Theorem 3.1 extends Hunter's (1969) result in two ways. First, the vector $\mathbf{l}'$ is replaced by any $\mathbf{u}'$ such that $\mathbf{u}'\mathbf{e} \neq 0$. Second, we only require $\mathbf{P}$ to have a simple root of unity. Hunter assumed $\mathbf{P}$ irreducible and the associated Markov chain regular (or ergodic).

Decell and Odell (1967) considered a generalized inverse $(\mathbf{I}-\mathbf{P})^-$ of $\mathbf{I}-\mathbf{P}$ with $(\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})^-$ symmetric [cf. (2.2.2)]. Our g-inverse $(\mathbf{I}-\mathbf{P}+\mathbf{eu}')^{-1}$ satisfies this symmetry condition if and only if $\mathbf{P}$ is doubly-stochastic, i.e., $\mathbf{l}' = \mathbf{e}'/n$. The other projector

$$(\mathbf{I}-\mathbf{P}+\mathbf{eu}')^{-1}(\mathbf{I}-\mathbf{P}) = \mathbf{I}-\mathbf{eu}'/\mathbf{u}'\mathbf{e}, \tag{3.5}$$

however, is symmetric if and only if $\mathbf{u}'$ is proportional to $\mathbf{e}'$.

A generalized inverse $\mathbf{G}$ of the matrix $\mathbf{A}$ is said to be *reflexive* (Rao, Mitra, 1971) whenever $\mathbf{G}$ has the same rank as $\mathbf{A}$. Thus

$$\mathbf{H} = (\mathbf{I}-\mathbf{P}+\mathbf{eu}')^{-1}-\mathbf{ew}' \tag{3.6}$$

is a reflexive g-inverse of $\mathbf{I}-\mathbf{P}$ if and only if $\mathbf{w}'\mathbf{e} = 1/\mathbf{u}'\mathbf{e}$. To see this note first that $(\mathbf{I}-\mathbf{P})\mathbf{H} = \mathbf{I}_n-\mathbf{el}'$, using (3.4), so $\mathbf{H}$ is a g-inverse of $\mathbf{I}-\mathbf{P}$. Then $\mathbf{H}$ is reflexive (Rao, Mitra, 1971, p. 28) if and only if $\mathbf{H}(\mathbf{I}-\mathbf{P})\mathbf{H} = \mathbf{H}$, which holds if and only if $\mathbf{w}'\mathbf{e} = 1/\mathbf{u}'\mathbf{e}$.

The generalized inverse $\mathbf{G}$ of the matrix $\mathbf{A}$ that has the same rank as $\mathbf{A}$ and for which the projectors $\mathbf{AG}$ and $\mathbf{GA}$ are symmetric is the unique Moore-

Penrose inverse $\mathbf{A}^+$. Rao and Mitra (1971) and Albert (1972) give numerous properties and applications. We find:

THEOREM 3.2 *Let the $n\times n$ matrix $\mathbf{I}-\mathbf{P}$ have rank $n-1$ and suppose*

$$\mathbf{Pe}=\mathbf{e} \text{ and } \mathbf{l}'\mathbf{P}=\mathbf{l}', \tag{3.7}$$

where $\mathbf{l}$ *is an* $n\times 1$ *vector, and* $\mathbf{e}$ *is the* $n\times 1$ *vector of ones. Then the Moore-Penrose inverse of* $\mathbf{I}-\mathbf{P}$ *is*

$$(I-P)^+=(\mathbf{I}-\mathbf{P}+\alpha\mathbf{le}')^{-1}-\alpha\mathbf{el}', \tag{3.8}$$

where

$$\alpha=(n\mathbf{l}'\mathbf{l})^{-1/2}. \tag{3.9}$$

Proof From Lemma 2.2 and (2.2.16), $|\mathbf{I}-\mathbf{P}+\alpha\mathbf{le}'|=k/\alpha$, with $k=\Pi_{j=2}^n(1-\lambda_j)\neq 0$, where $1=\lambda_1,\lambda_2,\ldots,\lambda_n$ are the characteristic roots of $\mathbf{P}$. Thus $\mathbf{I}-\mathbf{P}+\alpha\mathbf{le}'$ is nonsingular and using representations like (3.3) and (3.5) it follows that the two projectors

$$(\mathbf{I}-\mathbf{P})[(\mathbf{I}-\mathbf{P}+\alpha\mathbf{le}')^{-1}-\alpha\mathbf{el}']=\mathbf{I}_n-\mathbf{ll}'/\mathbf{l}'\mathbf{l}, \tag{3.10}$$

and

$$[(\mathbf{I}-\mathbf{P}+\alpha\mathbf{le}')^{-1}-\alpha\mathbf{el}'](\mathbf{I}-\mathbf{P})=\mathbf{I}_n-\mathbf{ee}'/n, \tag{3.11}$$

are both symmetric. Postmultiplying (3.10) by $\mathbf{I}-\mathbf{P}$ shows that (3.8) is a g-inverse of $\mathbf{I}-\mathbf{P}$ and premultiplying (3.10) by (3.8) shows that $\mathbf{I}-\mathbf{P}$ is a g-inverse of (3.8), which completes the proof.

Notice the reversed positions of $\mathbf{e}$ in (3.8) and (3.3). The assumption in Theorem 3.2 that $\mathbf{I}-\mathbf{P}$ have rank $n-1$ is weaker than the assumption in Theorem 3.1 that the unit characteristic root of $\mathbf{P}$ be simple. When $\mathbf{P}$ is a stochastic matrix, however, the two assumptions are equivalent (Marcus, Minc, 1964 p. 133). The special case of Theorem 3.2 with $\mathbf{l}=\mathbf{e}/n$ was studied in detail by Sharpe and Styan (1965a, 1965b).

Various extensions of Theorems 3.1 and 3.2 are obtained by Wachter (1973). We conclude our paper with one of these.

THEOREM 3.3 *Let the $n\times n$ matrix $\mathbf{A}$ have rank $n-1$ and suppose*

$$\mathbf{Af}=\mathbf{0} \text{ and } \mathbf{k}'\mathbf{A}=\mathbf{0}', \tag{3.12}$$

where $\mathbf{f}$ *and* $\mathbf{k}$ *are* $n\times 1$ *vectors. Then the Moore-Penrose inverse of* $\mathbf{A}$ *is*

$$\mathbf{A}^+=(\mathbf{A}+\gamma\mathbf{kf}')^{-1}-\gamma\mathbf{fk}';\quad \gamma=(\mathbf{f}'\mathbf{f}\cdot\mathbf{k}'\mathbf{k})^{-1/2}. \tag{3.13}$$

Acknowledgements

Much of the research of the second author was performed while visiting Stanford University in the summer of 1970. He gratefully acknowledges helpful conversations with T. W. Anderson, Gene H. Golub, Alberto Tubilla, and Arthur F. Veinott, Jr., and financial support from the National Science Foundation under Grant No. GS–2635 at Stanford University. Support from the National Research Council of Canada is also gratefully acknowledged.

References

Albert, A. (1972). *Regression and the Moore-Penrose Pseudoinverse*. Academic Press, New York.

Bodewig, E. (1959). *Matrix Calculus*. Second Edition. North-Holland, Amsterdam.

Brauer, A. (1952). "Limits for the characteristic roots of a matrix. IV: Applications to stochastic matrices." *Duke Mathematical Journal*, **19**, 75–91.

Chung, K. L. (1960). *Markov Chains with Stationary Transition Probabilities*. Springer-Verlag, Berlin.

Decell, H. P., Jr., and P. L. Odell, (1967). "On the fixed point probability vector of regular or ergodic transition matrices." *Journal of the American Statistical Association*, **62**, 600–602.

Golub, G. H. (1969). "Matrix decompositions and statistical calculations." *Statistical Computation*, R. C. Milton and J. A. Nelder, eds., Academic Press, New York, 365–397.

Golub, G. H. (1970). Personal communication.

Golub, G. H. and E. Seneta, (1973). "Computation of the stationary distribution of an infinite Markov matrix." *Bulletin of the Australian Mathematical Society*, **8**, 333–341.

Hunter, J. J. (1969). "On the moments of Markov renewal processes." *Advances in Applied Probability*, **1**, 188–210.

Keilson, J. and G. P. H. Styan, (1973). "Markov chains and *M*-matrices: inequalities and equalities." *Journal of Mathematical Analysis and Applications*, **41**, 439–459.

Kemeny, J. G. and J. L. Snell, (1960). *Finite Markov Chains*. Van Nostrand, Princeton.

Marcus, Marvin and Minc, Henryk (1964). *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, and Prindle, Weber and Schmidt, Boston.

Rao, C. Radhakrishna, and Mitra, Sujit Kumar (1971). *Generalized Inverse of Matrices and its Applications*. Wiley, New York.

Sharpe, G. E. and G. P. H. Styan (1965a). "Circuit duality and the general network inverse." *IEEE Transactions on Circuit Theory*, CT–12, 22–27.

Sharpe, G. E. and G. P. H. Styan (1965b). "A note on the general network inverse". *IEEE Transactions on Circuit Theory*, CT–12, 632–633.

Styan, G. P. H. (1964). "Selected Topics in Markov Chain Theory, with an Application to Marketing." *M.A. Thesis, Dept. Math. Statist.*, Columbia Univ., New York.

Tubilla, A. (1970). Personal communication.

Wachter, P. (1973). "Solving Certain Systems of Homogeneous Equations with Special Reference to Markov Chains." *M.Sc. Thesis. Dept. Math.*, McGill Univ., Montreal.

Wilkinson, J. H. (1963). *Rounding Errors in Algebraic Processes*. Notes on Applied Science No. 32. H.M.S.O., London.

Wilkinson, J. H. (1965). *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford.