BULL. AUSTRAL. MATH. SOC. 68A10, 60J99
VOL. 8 (1973), 333-341.

Computation of the stationary distribution of an infinite Markov matrix

G.H. Golub and E. Seneta

An algorithm is presented for computing the unique stationary distribution of an infinite stochastic matrix possessing at least one column whose elements are bounded away from zero. Elementwise convergence rate is discussed by means of two examples.

1. Introduction

For a denumerably infinite stochastic matrix $P = \{p_{ij}\}$, $i, j = 1, 2, \ldots$, a vector $\mathbf{x}$ satisfying

$$(1.1) \qquad \mathbf{x} \geq 0 , \quad \mathbf{x} \neq 0 , \quad \mathbf{x}'P = \mathbf{x}'$$

is called an invariant measure; any positive multiple of an invariant measure is one also. If an invariant measure satisfies, in addition,

$$(1.2) \qquad \mathbf{x}'\mathbf{1} \equiv \sum_{i=1}^{\infty} x_i = 1 ,$$

it is called a stationary distribution.

In this note we shall display an algorithm for computing a stationary distribution $\mathbf{x}$ (under conditions on P which ensure existence and

Received 5 December 1972. The work of the first author began whilst a visitor at the Australian National University, and it was in part supported by the Atomic Energy Commission and the National Science Foundation of the USA. This author is pleased to thank Dr M.R. Osborne for his kind hospitality.

https://doi.org/10.1017/S0004972700042623 Published online by Cambridge University Press

uniqueness) from successive finite matrix truncations of P . A similar algorithm when P is a finite matrix has been previously described (Styan, [5]).

It is now well known (Feller, [2]) that for an irreducible and recurrent (persistent) P an invariant measure always exists, and is unique, to positive multiples; and is elementwise strictly positive. In two previous papers (Seneta, [3], [4]) two algorithms were discussed which yielded pointwise convergence to such x , of vectors computed from the successive truncations of P , when x is normed so that a fixed element is unity. If P is in fact positive-recurrent, its invariant measure can be normed to satisfy (1.2), so that a (unique) stationary distribution exists, and it is in this form (of a stationary distribution) that the invariant measure is usually required to be computed, from the Markov chain context in which stochastic positive-recurrent P are important. This problem was touched on but not discussed to any extent in the two papers cited.

We shall not necessarily have present in this note the irreducibility of P , but work under the probabilistically restrictive, but classical assumption that P satisfies

$$\sup_j\left\{\inf_i p_{ij}\right\} > 0 \tag{1.3}$$

that is, there is at least one column of P , say the j^*-th, with positive elements, which are in addition uniformly bounded away from zero, that is, for at least one j , say $j = j^*$,

$$\inf_i p_{ij} > \delta(j) > 0 . \tag{1.4}$$

2. Markov matrices

Finite stochastic P with a positive column are classically known as Markov matrices (Bernšteĭn, [1]). The condition (1.3) is a natural way of extending this terminology to the infinite case, since, moreover as we shall now sketch, the implications are the same as in the finite case.

The positivity (alone) of the j^*-column implies that the index set of P contains a single essential class C of indices (that is, a single

https://doi.org/10.1017/S0004972700042623 Published online by Cambridge University Press

closed self-communicating class), which contains j^* and is therefore aperiodic. Indices outside C, if any, are inessential and lead to C. The index set C is in fact positive-recurrent under (1.3) for

$$p_{j^*j^*}^{(r)} = \sum_k p_{j^*k}^{(r-1)} p_{kj^*} \geq \inf_k p_{kj^*} \sum_k p_{j^*k}^{(r-1)} = \delta(j^*) > 0$$

where $P^r = \left\{p_{ij}^{(r)}\right\}$. A matrix P containing a single essential aperiodic class, C, which is in fact positive-recurrent, is sometimes called regular; for such P it is well-known that, elementwise, as $r \to \infty$, ergodicity obtains, that is,

$$P^r \to \mathbf{1} \cdot \mathrm{x}' \tag{2.1}$$

where x is the (unique) stationary distribution of P, and only those elements of $\mathrm{x} = \{x(i)\}$ are positive for which $i \in C$. In the present situation where (1.3) holds, it can be deduced that the elementwise approach to the limit in (2.1) is in fact (uniformly) geometric. This 'geometric ergodicity' testifies to the restrictiveness of condition (1.3).

For the sequel it is convenient, and results in no loss of generality, to take $j^* = 1$, so that

$$p_{i1} > \delta(1) > 0 , \quad i = 1, 2, \ldots . \tag{2.2}$$

If we denote by ${}_{(n)}P = \left\{{}_{(n)}p_{ij}\right\}$ the $(n \times n)$ northwest corner truncation of P, then ${}_{(n)}P$ is in general substochastic, and in virtue of (2.2) contains a single closed finite set of indices, ${}_{(n)}C$, which contains the index 1, and so is aperiodic.

3. The algorithm

Define the vector $\mathrm{y} = \{y(j)\} \geq \mathbf{0}$ by

$$y(j) = \begin{cases} \delta(j) & \text{if } j \text{ satisfies (1.4)}, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, by assumption, $\mathrm{y} \neq \mathbf{0}$, with at least first element positive. Focus attention on the following infinite system of equations, which is certainly satisfied by the unique stationary distribution corresponding to

https://doi.org/10.1017/S0004972700042623 Published online by Cambridge University Press

P :

(3.1) $$x'\left(I-(P-1\cdot y')\right) = y' ,$$

where x is a vector of unknowns; and on the corresponding $(n\times n)$ northwest truncated system for each $n = 1, 2, \ldots$,

(3.2) $${}_{(n)}z'\left({}_{(n)}I-\left({}_{(n)}P-{}_{(n)}1\cdot{}_{(n)}y'\right)\right) = {}_{(n)}y'$$

where ${}_{(n)}z'$ is a vector of unknowns.

It should be noted that the subtraction of ${}_{(n)}1 \cdot {}_{(n)}y'$ from ${}_{(n)}P$ does not alter the location of the zero and positive elements of ${}_{(n)}P$ and so does not change its essential structure; however ${}_{(n)}P - {}_{(n)}1 \cdot {}_{(n)}y'$ now has each row sum strictly less than unity, and by a well known property of such matrices the matrix $\left({}_{(n)}I-\left({}_{(n)}P-{}_{(n)}1\cdot{}_{(n)}y'\right)\right)$ has an inverse, and, furthermore,

$$\left({}_{(n)}I-\left({}_{(n)}P-{}_{(n)}1\cdot{}_{(n)}y'\right)\right)^{-1} = \sum_{k=0}^{\infty} \left({}_{(n)}P-{}_{(n)}1\cdot{}_{(n)}y'\right)^{k}$$

elementwise, so that the inverse has non-negative entries (and indeed at least one column, the first, strictly positive, in virtue of (2.2)). It thus follows

(3.3) $${}_{(n)}z' = {}_{(n)}y'\left({}_{(n)}I-\left({}_{(n)}P-{}_{(n)}1\cdot{}_{(n)}y'\right)\right)^{-1} \geq 0 , \neq 0 ,$$
$$\text{with } {}_{(n)}z' \geq {}_{(n)}y'$$

is the unique solution to (3.2), and, further, from (3.2), since

$${}_{(n)}z' - {}_{(n)}z' \cdot {}_{(n)}P + {}_{(n)}z' \cdot {}_{(n)}1 \cdot {}_{(n)}y' = {}_{(n)}y' ,$$

it follows that

(3.4) $${}_{(n)}z' \cdot {}_{(n)}1 - {}_{(n)}z' \cdot {}_{(n)}1 + {}_{(n)}z' \cdot {}_{(n)}1 \cdot {}_{(n)}y' \cdot {}_{(n)}1$$
$$\leq {}_{(n)}y' \cdot {}_{(n)}1$$

on account of the substochasticity of ${}_{(n)}P$. Hence, since ${}_{(n)}y' \geq 0$, $\neq 0$ it follows from (3.3) and (3.4) that

https://doi.org/10.1017/S0004972700042623 Published online by Cambridge University Press

(3.5) $$0 < \delta(1) \le {}_{(n)}z(1) \le {}_{(n)}\mathbf{z}' \cdot {}_{(n)}\mathbf{1} \le 1$$

where ${}_{(n)}\mathbf{z} = \{{}_{(n)}z(i)\}$; and from (3.2) that

(3.6) $${}_{(n)}\mathbf{z}' = {}_{(n)}\mathbf{z}'\left({}_{(n)}P - {}_{(n)}\mathbf{1} \cdot {}_{(n)}\mathbf{y}'\right) + {}_{(n)}\mathbf{y}' .$$

Now since

$${}_{(n)}\mathbf{z}' = {}_{(n)}\mathbf{y}' \sum_{k=0}^{\infty} \left({}_{(n)}P - {}_{(n)}\mathbf{1} \cdot {}_{(n)}\mathbf{y}'\right)^k ,$$

it follows that

(3.7) $${}_{(n+1)}\mathbf{z}' \ge {}_{(n)}\mathbf{z}'$$

(if we extend, for the present instance only, the definition of ${}_{(n)}\mathbf{z}$ by putting ${}_{(n)}z(i) = 0$ for $i > n$). Thus we know that the limit

$$z^*(i) = \lim_{n\to\infty} {}_{(n)}z(i)$$

exists for each $i = 1, 2, \ldots$, by the boundedness of (3.5). If we put $\mathbf{z}^* = \{z^*(i)\}$, (3.5) and (3.6) give, by Fatou's Lemma,

(3.8) $$\begin{cases} 0 < \delta(1) \le \mathbf{z}^{*\prime}\mathbf{1} \le 1 , \\ \mathbf{z}^{*\prime} \ge \mathbf{z}^{*\prime}(P - \mathbf{1} \cdot \mathbf{y}') + \mathbf{y}' , \end{cases}$$

which implies

$$\mathbf{z}^{*\prime} \ge \mathbf{z}^{*\prime}P , \quad \mathbf{z}^* \ge \mathbf{0} , \quad \ne \mathbf{0} .$$

Now in fact, equality must hold at all entries, for otherwise,

$$\mathbf{z}^{*\prime}\mathbf{1} > \mathbf{z}^{*\prime}\mathbf{1}$$

by stochasticity of P . Thus

$$\mathbf{z}^{*\prime} = \mathbf{z}^{*\prime}P$$

and from (3.8),

$$\mathbf{z}^{*\prime}\mathbf{1} = 1 .$$

Thus $\mathbf{z}^*$ is the unique stationary distribution corresponding to P .

Thus to summarize: *The successive solutions* ${}_{(n)}\mathbf{z}$, $n = 1, 2, \ldots$

https://doi.org/10.1017/S0004972700042623 Published online by Cambridge University Press

for the finite systems (3.2) *converge elementwise to the unique stationary distribution corresponding to* P . *Moreover, from* (3.7), *the elementwise convergence is monotone increasing in* ${}_{(n)}\mathbf{z}$, *thus providing a steadily improving bound for the required limit vector.*

4. Convergence rate

It appears that little, in general, can be said about the convergence rate. This is borne out by the following simple example. Let $\mathbf{p} = \{p_i\}$ be a probability vector with all entries positive, that is, $p_i > 0$, $\sum_{i=1}^{\infty} p_i = 1$. The infinite matrix

$$P = \mathbf{1} \cdot \mathbf{p}' \tag{4.1}$$

clearly satisfies (1.4), and has unique stationary distribution $\mathbf{p}$. If we indicate with a subscript n the usual truncations, then

$$ {}_{(n)}P = {}_{(n)}\mathbf{1} \cdot {}_{(n)}\mathbf{p}' \tag{4.2}$$

and corresponding to its Perron-Frobenius eigenvalue, ${}_{(n)}\mathbf{p}' \cdot \mathbf{1}$, has left and right positive eigenvectors respectively ${}_{(n)}\mathbf{p}'$, ${}_{(n)}\mathbf{1}$. It follows that

$$ {}_{(n)}P^k = \begin{cases} \left({}_{(n)}\mathbf{p}' \cdot {}_{(n)}\mathbf{1}\right)^{k-1} {}_{(n)}\mathbf{1} \cdot {}_{(n)}\mathbf{p}' , & k \geq 1 , \\ I & , \ k = 0 . \end{cases} \tag{4.3}$$

Now a permissible choice of $\mathbf{y}$ is $\delta\mathbf{p}$, where $0 < \delta < 1$, in which case (3.3) becomes

$$\begin{aligned} {}_{(n)}\mathbf{z}' &= \delta\, {}_{(n)}\mathbf{p}' \left[{}_{(n)}I - (1-\delta)\, {}_{(n)}\mathbf{1} \cdot {}_{(n)}\mathbf{p}'\right]^{-1} \\ &= \delta\, {}_{(n)}\mathbf{p}' \sum_{k=0}^{\infty} (1-\delta)^k \left({}_{(n)}\mathbf{1} \cdot {}_{(n)}\mathbf{p}'\right)^k \end{aligned}$$

and using (4.3),

https://doi.org/10.1017/S0004972700042623 Published online by Cambridge University Press

$$
{}_{(n)}\mathbf{z}' = \left[\delta \sum_{k=0}^{\infty} \{(1-\delta)({}_{(n)}\mathbf{p}' \cdot {}_{(n)}\mathbf{1})\}^k\right]\mathbf{p}'_n
$$
$$
= \frac{\delta}{1-(1-\delta)({}_{(n)}\mathbf{p}' \cdot {}_{(n)}\mathbf{1})} \mathbf{p}'_n \, .
$$

Therefore

$$
{}_{(n)}\mathbf{p}' - {}_{(n)}\mathbf{z}' = \frac{(1-\delta)(1-{}_{(n)}\mathbf{p}' \cdot {}_{(n)}\mathbf{1})}{1-(1-\delta)({}_{(n)}\mathbf{p}' \cdot {}_{(n)}\mathbf{1})} \, {}_{(n)}\mathbf{p}' \, ,
$$

and we notice that, since ${}_{(n)}\mathbf{p}$ already coincides with the first n elements of the stationary distribution $\mathbf{p}$, the *rate* of pointwise convergence is that of

$$
(4.4) \qquad 1 - \sum_{i=1}^{n} p_i
$$

to zero as $n \to \infty$.

Since we are at liberty to choose the $\{p_i\}$, within the constraint $p_i > 0$ all i, $\sum_i p_i = 1$, we can arrange to make the convergence of this quantity to zero quite slow, for example, if we choose $p_i = \text{const } i^{-(1+\gamma)}$, $\gamma > 0$, then (4.4) is $O(n^{-\gamma})$ as $n \to \infty$.

It may be relevant to note, that for this rather specialized example, one of the approximation techniques described in Seneta [3], that of finding successive left Perron-Frobenius eigenvectors of ${}_{(n)}P$ and norming always so that, for example, the first element is unity, "settles down" immediately to the elements of the stationary distribution similarly normed, for ${}_{(n)}\mathbf{p}/p_1$ coincides with the first n elements of $\mathbf{p}/p_1$. However, it is also known by example (Example (1) in the paper just cited) that the eigenvector convergence for this method can be slow also; and in any case the 'convergence radius' (reciprocal of the Perron-Frobenius eigenvalue)

$$
[{}_{(n)}\mathbf{p}'\mathbf{1}]^{-1} \downarrow 1
$$

again at rate (4.4).

https://doi.org/10.1017/S0004972700042623 Published online by Cambridge University Press

We conclude with another simple example. If $P = \{p_{ij}\}$ is given by $p_{i1} = a$, $p_{i,i+1} = 1 - a$, $i = 1, 2, \ldots$, $0 < a < 1$, and $p_{ij} = 0$ otherwise, and we take $y = \{y(j)\}$ to be defined by $y(1) = (1-\gamma)a$, $y(j) = 0$ otherwise, where $0 < \gamma < 1$, straightforward calculations give

$$ {}_{(n)}z(i) = C(n)(1-a)^i \ , \quad i = 1, 2, \ldots, n $$

where

$$ C(n) = \frac{a}{1-a}\left\{\frac{1-\gamma}{1-\gamma+\gamma(1-a)^n}\right\} . $$

The difference between the required i-th component and its approximations obtained from the n-th truncation is thus

$$ x(i) - {}_{(n)}z(i) = a(1-a)^{i-1}\left\{\frac{\gamma(1-a)^n}{1-\gamma+\gamma(1-a)^n}\right\} , $$

so that the pointwise convergence rate is geometric and independent of γ .

References

[1] С.Н. Бернштейн [S.N. Bernšteĭn], Теория вероятностей, изд. 4-е [*Probability theory*, 4th edition] (Gostehizdat, Moscow, Leningrad, 1946).

[2] William Feller, *An introduction to probability theory and its applications*, Volume 1, 3rd edition (John Wiley & Sons, New York, London, Sydney, 1968).

[3] E. Seneta, "Finite approximations to infinite non-negative matrices", *Proc. Cambridge Philos. Soc.* 63 (1967), 983-992.

[4] E. Seneta, "Finite approximations to infinite non-negative matrices, II: refinements and applications", *Proc. Cambridge Philos. Soc.* 64 (1968), 465-470.

https://doi.org/10.1017/S0004972700042623 Published online by Cambridge University Press

[5] G.P.H. Styan, Personal communication (1970).

Department of Computer Science,
Stanford University,
California,
USA;

Department of Statistics,
Princeton University,
New Jersey,
USA,
and
Department of Statistics,
School of General Studies,
Australian National University,
Canberra, ACT.

https://doi.org/10.1017/S0004972700042623 Published online by Cambridge University Press