Existence and Uniqueness of Stationary Probability Vector for Stochastic Matrices using Farkas' Lemma

Abstract

The purpose of this note is to provide a simple proof of existence of stationary probability vectors (fixed points) for stochastic matrices using Farkas' lemma. This result as well as the uniqueness of stationary probability vectors also holds for a certain subclass of quasi-stochastic matrices.

1. Introduction: In this note we provide a new and simple proof of the existence of stationary probability vectors (fixed points) for stochastic matrices as defined for instance in section 6 of Chapter 5 of the book by Margalit and Rabinoff (2017), using Farkas' lemma-a very simple proof of the latter, using purely combinatorial arguments, is available in topic 3 of Lahiri (2022). The usual approaches to the proof rely either on Brouwer's fixed point theorem (for instance the first theorem in section 24.6 of https://python.quantecon.org/finite_markov.html), or the contraction mapping theorem- the latter being the proof in Lalley (undated). The proof that there exists a unique stationary probability vector for regular stochastic matrices is an easy consequence of our main result. Both results hold for a subclass of quasi-stochastic matrices (i.e., non-negative matrices with row sums less than or equal to one) which satisfy the property that either there is a square submatrix in the upper left-hand corner with all row sums equal to one or there is a square submatrix in the lower right-hand corner with all row sums equal to one.

We thus provide an alternative and considerably easier route to the proofs of theorems in section 7 of chapter 4 in Kemeny, Snell and Thompson (1974) concerning convergence results for stochastic matrices.

2. The model and the main existence result: An n-**dimensional probability row vector** is a row vector u in $\mathbb{R}^n_+$ such that $\sum_{j=1}^{n} u_j = 1$.

An n×n matrix P (of real numbers) is said to be **a non-negative matrix** if all entries in P are non-negative, i.e., if p_{ij} denotes the $(i,j)^{th}$ entry of P then for all $i,j\in\{1,\ldots,n\}$, $p_{ij} \geq 0$.

An n×n non-negative matrix P is said to be a **stochastic matrix** if all row sums of P are equal to one, i.e. for all $i\in\{1,\ldots,n\}$: $\sum_{j=1}^{n} p_{ij} =1$.

An n-dimensional probability row vector is said to be a **stationary probability vector** for an n×n matrix P, if uP = u.

Given a n n×n matrix P and $i,j\in\{1,\ldots,n\}$, let P_i denote the i^{th} row of P and $P^{(j)}$ denote the j^{th} column of j.

The simple proof of the following theorem uses Farkas' lemma.

Theorem 1: Let P be a n×n stochastic matrix. Then there exists a stationary probability vector for P.

Proof: Let A = P-I, where I is the n×n identity matrix, and towards a contradiction suppose there does not exist any probability row vector u such that uA = 0.

Then the system u[A|e(n)] = [0|1] has no non-negative solution, where

(a) [A|e(n)] is the n×(n+1) matrix whose first n columns are the columns of the matrix A and its $(n+1)^{st}$ column e(n) is the n-dimensional column vector all whose entries are 1; and

(b) [0|1] is the (n+1)-dimensional row vector whose first 'n' entries are 0 and the last entry is 1.

Thus, by Farkas' Lemma there exists an n- dimensional column vector z and a real number α such that $Az + \alpha e(n) \geq 0$ and $\alpha < 0$.

Thus, Pz-z = Az >> 0, i.e., $P_i z > z_i$ for all i = 1,…,n, where P_i is the i^{th} row of P.

Let $z_h = \max\{z_i | i = 1,\ldots,n\}$.

Thus $P_h z > z_h$ which implies that a convex combination of the set of real numbers $\{z_i|i = 1,\ldots,n\}$ is greater than its maximum which is not possible.

Thus, it must be the case that there exists a probability row vector u such that uP = u. Q.E.D.

An n×n non-negative matrix P is said to be a **quasi- stochastic matrix** if all row sums of P are "positive" and "less" than or equal to one, i.e. for all $i\in\{1,\ldots,n\}$: $0 < \sum_{j=1}^{n} p_{ij} \leq 1$.

Note: Theorem 1 is not valid if we replace "stochastic matrix" by "quasi-stochastic matrix" as is apparent if we choose P = $\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$. That however does not imply that it is never true for quasi-stochastic matrices. If we let P = $\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$, then the probability vector (0,1) is a stationary probability vector $\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$. In fact, quasi-stochastic matrices of the form $\begin{bmatrix} P_1 & Q \\ R & P_2 \end{bmatrix}$, where P_1 and P_2 are square sub-matrices where either all row sums of P_1 are equal to one or all row sums of P_2 are equal to one, will have at least one stationary probability vector. This follows easily by adapting the proof of Theorem 1 to this context. If all row sums of P_1 are equal to one then Q has to be the zero matrix and if all row sums of P_2 are equal to one then R has to be the zero matrix.

3. Some properties of products of Quasi-Stochastic and Stochastic Matrices: Given any quasi-stochastic matrix P and an n-dimensional probability row vector u, [uP = u] implies [$u_i = 0$ whenever $\sum_{j=1}^{n} p_{ij} < 1$]. The reason is as follows:

If e(n) is the n-dimensional column vector all whose entries are 1, then uP = u implies uPe(n) = ue(n) = 1 and hence $\sum_{i=1}^{n} u_i \sum_{j=1}^{n} p_{ij} = 1$. Thus, $\sum_{i=1}^{n} u_i \left(\sum_{j=1}^{n} p_{ij} - 1\right) = 0$, which given $\sum_{j=1}^{n} p_{ij} \leq 1$ for all $i\in\{1,\ldots,n\}$ implies $u_i = 0$ if $\sum_{j=1}^{n} p_{ij} < 1$.

Note that if P and Q are quasi-stochastic matrices of size n×n then PQ is a non-negative matrix whose i^{th} row is P_iQ and j^{th} entry in row i of PQ is $P_iQ^{(j)}$.

Hence sum of terms in the i^{th} row of PQ is $P_i \sum_{j=1}^{n} Q^{(j)}$.

Since Q is a quasi-stochastic matrix each coordinate of $\sum_{j=1}^{n} Q^{(j)}$ is positive and less than or equal to one and since P quasi-stochastic matrix $P_i \sum_{j=1}^{n} Q^{(j)}$ is positive and less than or equal to one.

Thus, PQ is a quasi-stochastic matrix.

In particular P^r (i.e. P multiplied by itself r times for some positive integer r) is a quasi-stochastic matrix for all $r \in \mathbb{N}$. Clearly, $P^1 = P$.

Note the difference between P^r and $P^{(j)}$, where the latter denotes the j^{th} column of P.

If x is a <u>non-negative</u> row vectors and Q is a quasi-stochastic matrix, then $\sum_{j=1}^{n} xQ^{(j)} = \sum_{j=1}^{n} \sum_{i=1}^{n} x_i q_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i q_{ij} = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} q_{ij} \leq \sum_{i=1}^{n} x_i$.

If P and Q are stochastic matrices of size n×n then PQ is a non-negative matrix whose i^{th} row is P_iQ and j^{th} entry in row i of PQ is $P_iQ^{(j)}$.

Hence sum of terms in the i^{th} row of PQ is $P_i \sum_{j=1}^{n} Q^{(j)}$.

Since Q is a stochastic matrix implies $\sum_{j=1}^{n} Q^{(j)} = e(n)$ and since P stochastic matrix $P_i \sum_{j=1}^{n} Q^{(j)} = P_ie(n) = 1$.

Thus, PQ is a stochastic matrix.

In particular P^r (i.e. P multiplied by itself r times for some positive integer r) is a stochastic matrix for all $r \in \mathbb{N}$. Clearly, $P^1 = P$.

If x is a row vectors and Q is a stochastic matrix, then $\sum_{j=1}^{n} xQ^{(j)} = \sum_{j=1}^{n} \sum_{i=1}^{n} x_i q_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i q_{ij} = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} q_{ij} = \sum_{i=1}^{n} x_i$.

Note that if Q is a stochastic matrix, the above holds for all (and not just non-negative) row vectors which has as many co-ordinates as the number of rows in Q.

Thus, the sum of the coordinates of xQ^r, where Q^r is Q multiplied by itself r times, is equal to the sum of the coordinates of x for all positive integers r.

Given $x,y \in \mathbb{R}^n$, the **Manhattan distance** between x and y denoted $||x-y||_{MD}$ is defined as $\sum_{i=1}^{n} |x_i - y_i|$.

If x and y are n-dimensional non-negative row vectors and P is a quasi-stochastic matrix, then $||xP-yP||_{MD} = \sum_{j=1}^{n} |(x-y)P^{(j)}| = \sum_{j=1}^{n} |\sum_{i=}^{n} (x_i - y_i)p_{ij}| \leq \sum_{j=1}^{n} \sum_{i=1}^{n} |x_i - y_i| p_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - y_i| p_{ij} = \sum_{i=1}^{n} |x_i - y_i| \sum_{j=1}^{n} p_{ij} \leq \sum_{i=1}^{n} |x_i - y_i| = ||x-y||_{MD}$ since $0 < \sum_{j=1}^{n} p_{ij} \leq 1$ for all $i \in N$.

Thus, for all $r \in \mathbb{N}$ and n-dimensional non-negative row vectors x, y, $||x-y||_{MD} \geq ||xP^r-yP^r||_{MD} \geq ||xP^{r+1}-yP^{r+1}||_{MD}$.

4. Regular Quasi-Stochastic Matrices: Results in this section lead to the second theorem in section 24.6 of https://python.quantecon.org/finite_markov.html.

A quasi-stochastic matrix P is said to be **regular** if for some positive integer r, all entries of P^r are positive.

For $i,j \in \{1,\ldots,n\}$ and any positive integer r, let $p_{ij}^{[r]}$ denote the j^{th} entry in the i^{th} row of P^r, P_i^r denote the i^{th} row of P^r and $P^{r(,j)}$ denote the j^{th} column of P^r.

The proof of the following- intended to be a slight generalization of Lalley (undated)- closely follows page 10 of Lalley's notes on "Markov Chains: Basic Theory" available at: http://galton.uchicago.edu/~lalley/Courses/312/MarkovChains.pdf .

Lemma 1: If P is a regular quasi-stochastic matrix with all entries of P^K strictly positive for some positive integer K, then there exists $\varepsilon > 0$ with $0 < 1-n\varepsilon < 1$ such that for $x,y \in \{u|\ u$ is an n-dimensional row vector satisfying $\sum_{j=1}^{n} u_j = 1\}$, then $||xP^K-yP^K||_{MD} \leq (1-n\varepsilon)||x-y||_{MD}$. Thus, for all positive integers m, $||xP^{mK}-yP^{mK}||_{MD} \leq (1-n\varepsilon)^m||x-y||_{MD}$.

Proof: Suppose P is a regular quasi-stochastic matrix with all entries of P^K strictly positive for some positive integer K and let $x,y \in \{u|\ u$ is an n-dimensional row vector satisfying $\sum_{j=1}^{n} u_j = 1\}$.

Since all entries of P^K are positive, there exists $\varepsilon > 0$, such that $p_{ij}^{[K]} > \varepsilon$ for all $i,j \in \{1,\ldots,n\}$.

Since P^K is also a quasi-stochastic matrix, it must be the case that $0 < \sum_{j=1}^{n} p_{ij}^{[K]} \leq 1$.

Thus, $1 \geq \sum_{j=1}^{n} p_{ij}^{[K]} > n\varepsilon$ for all $j \in \{1,\ldots,n\}$.

Thus, $1 > 1-n\varepsilon > 0$.

For $i,j \in \{1,\ldots,n\}$, let $q_{ij} = \frac{p_{ij}^{[K]}-\varepsilon}{1-n\varepsilon}$. Since $p_{ij}^{[K]} > \varepsilon$ for all $i,j \in \{1,\ldots,n\}$ and $1-n\varepsilon > 0$, it must be the case that $q_{ij} > 0$ for all $i,j \in \{1,\ldots,n\}$. Further, for all $i \in \{1,\ldots,n\}$, $0 < \sum_{j=1}^{n} q_{ij} = \sum_{j=1}^{n} \frac{p_{ij}^{[K]}-\varepsilon}{1-n\varepsilon} \leq 1$.

$\sum_{j=1}^{n} |xP^{K(j)} - yP^{K(j)}| = \sum_{j=1}^{n} |\sum_{i=1}^{n}(x_i - y_i)p_{ij}^{[K]}| = \sum_{j=1}^{n} |\sum_{i=1}^{n}(x_i - y_i)[q_{ij}(1-n\varepsilon) + \varepsilon]| = \sum_{j=1}^{n} |(1-n\varepsilon)\sum_{i=1}^{n}(x_i - y_i)q_{ij} + \varepsilon\sum_{i=1}^{n}(x_i - y_i)| = \sum_{j=1}^{n} |(1-n\varepsilon)\sum_{i=1}^{n}(x_i - y_i)q_{ij}|$, since $\varepsilon\sum_{i=1}^{n}(x_i - y_i) = 0$.

Thus, $\sum_{j=1}^{n} |xP^{K(j)} - yP^{K(j)}| = (1-n\varepsilon)\sum_{j=1}^{n} |\sum_{i=1}^{n}(x_i - y_i)q_{ij}| \leq (1-n\varepsilon)\sum_{j=1}^{n}\sum_{i=1}^{n} |(x_i - y_i)q_{ij}| = (1-n\varepsilon)\sum_{j=1}^{n}\sum_{i=1}^{n} |(x_i - y_i)|q_{ij}$, since $q_{ij} > 0$ for all $i,j \in \{1,\ldots,n\}$.

Hence $\sum_{j=1}^{n} |xP^{K(j)} - yP^{K(j)}| \leq (1-n\varepsilon)\sum_{i=1}^{n}\sum_{j=1}^{n} |(x_i - y_i)|q_{ij} = (1-n\varepsilon)\sum_{i=1}^{n} |(x_i - y_i)|\sum_{j=1}^{n} q_{ij} \leq (1-n\varepsilon)\sum_{i=1}^{n} |(x_i - y_i)|$, since $0 < \sum_{j=1}^{n} q_{ij} \leq 1$ for all $i \in \{1,\ldots,n\}$.

Thus, $||xP^K - yP^K||_{MD} \leq (1-n\varepsilon)||x-y||_{MD}$ where $0 < 1-n\varepsilon < 1$.

Thus, for all positive integers m, $||xP^{(m+1)K} - yP^{(m+1)K}||_{MD} \leq (1-n\varepsilon)||xP^{mK} - yP^{mK}||_{MD}$ and hence $||xP^{mK} - yP^{mK}||_{MD} \leq (1-n\varepsilon)^m||x-y||_{MD}$, where $0 < 1-n\varepsilon < 1$. Q.E.D.

The following result is an immediate consequence of Theorem 1 and Lemma 1.

Theorem 2: If P is a regular stochastic matrix, then {u| u is an n-dimensional row vector satisfying $\sum_{j=1}^{n} u_j = 1$ and uP = u} is a singleton with all coordinates strictly positive.

Proof: Suppose P is a regular stochastic matrix with all entries of P^K being strictly positive for some positive integer K.

By Theorem 1, {u| u is an n-dimensional row vector satisfying $\sum_{j=1}^{n} u_j = 1$ and uP = u} is non-empty.

Further uP = u implies $uP^r = u$ for all positive integers r.

Towards a contradiction suppose, u and v are any two distinct (i.e., $u \neq v$) n-dimensional row vectors satisfying uP = u, vP = v, $\sum_{j=1}^{n} u_j = \sum_{j=1}^{n} v_j = 1$.

Thus $uP^K = u$ and $vP^K = v$.

By lemma 1, there exists $\varepsilon > 0$ with $0 < 1-n\varepsilon < 1$ such that $||uP^K - vP^K||_{MD} \leq (1-n\varepsilon)||u-v||_{MD}$.

Since $1 > 1-n\varepsilon > 0$, $u \neq v$ implies $||uP^K - vP^K||_{MD} < ||u-v||_{MD}$, contradicting $uP^K = u$ and $vP^K = v$.

Thus, {u| u is an n-dimensional row vector satisfying $\sum_{j=1}^{n} u_j = 1$ and uP = u} must be a singleton and by Theorem 1 it follows that there is a unique non-zero n-dimensional row vector u satisfying uP = u and this u is an n-dimensional probability row vector.

Since all entries of P^K are strictly positive, $u_j = uP^{K(j)} > 0$ for all $j \in \{1,\ldots,n\}$ and hence u is a strictly positive probability row vector. Q.E.D.

Note: It is easy to see that the above result holds for quasi-stochastic matrices of the form $\begin{bmatrix} P_1 & Q \\ R & P_2 \end{bmatrix}$, where P_1 and P_2 are square sub-matrices where either all row sums of P_1 are equal to one *and* P_1 is regular or all row sums of P_2 are equal to one *and* P_2 is regular.

5. Conclusion: The purpose of this short research communication on methodology is purely pedagogical. The generalizations of the results to quasi-stochastic matrices are immediate, and hence the originality of the paper is not in the results but in the proofs. This short note would allow the teaching finite Markov chains in an undergraduate course or MBA course on mathematical economics or operations research or stochastic processes. Applications of results presented here are well known and already exist in Kemeny, Snell and Thompson (1974). Needless duplication of effort would lead one to miss the point that this paper- if referred to at all- needs to be read along with material similar to what is already there in Kemeny, Snell and Thompson (1974). At the cost of sounding a little blunt though realistic, it may be noted that adding needless references, apart from sounding pedantic, would unnecessarily scare away many potential readers who may find this paper useful.

Acknowledgment: Many thanks to Dr. Subhadip Chakraborty for comments and corrections and much gratitude to Prof. Manjunatha Prasada for valuable discussions at the expense of his "peace of mind".

References

1. Kemeny, J.G., J. L. Snell and G.L. Thompson (1974): Introduction to Finite Mathematics (Third Edition). Prentice-Hall, Inc., Englewood Cliffs, N.J.

2. Lahiri, S.(2020): The essential appendix on linear programming. Available at: https://www.academia.edu/44541645/The_essential_appendix_on_Linear_Programming

3. Lalley,S. (undated): "Markov Chains: Basic Theory" available at: http://galton.uchicago.edu/~lalley/Courses/312/MarkovChains.pdf

4. Margalit, D. and J. Rabinoff (2017): Interactive Linear Algebra. Georgia Institute of Technology. Available at: https://textbooks.math.gatech.edu/ila/ila.pdf