Trends in Mathematics
Information Center for Mathematical Sciences
Volume 7, Number 2, December, 2004, Pages 169–175

LIMIT STATES OF STATIONARY MARKOV CHAINS

KEONHEE LEE AND YONG ZHANG

Abstract. We give the formula of the limits states of stationary Markov chains.

1. Introduction

Markov chains have become a standard topic in probability and a useful tool in many applications. See [1] P342-427 and references there.

One of the most important results in Markov chains is the existence of an invariant probability distribution of the irreducible Markov Chain. This invariant probability distribution corresponds to the limit theorem about irreducible stochastic matrices. In this paper, we will generalize the limit theorem to any stochastic matrices, not just irreducible ones.

The notion of stochastic matrix comes from Markov Chain. An experiment has k distinct states $s_1, s_2, \cdots, s_k$. Suppose we know that the probability of the experimental outcomes s_j is always $p_{i,j}$ when starting from the state s_i. Then this transition matrix $P = (p_{i,j})_{1\leq i,j\leq k}$ has the following property:

$$\sum_j p_{i,j} = 1, \ \forall \ i. \tag{1.1}$$

From this simple physics consideration, we have the following definition.

Definition 1.1. *A $k \times k$ non-negative matrix $P = (p_{i,j})$ is called* **stochastic** *if (1.1) is satisfied. A non-negative vector $v = (v_1, \cdots, v_k)$ is called a* **probability vector** *if $\sum_i v_i = 1$.*

Definition 1.2. *A square, say $k\times k$, stochastic matrix $P = (p_{i,j})$ is called* **positive** *(and write as $P > 0$) if $p_{i,j} > 0$ for all i, j. The notations are the same for vectors. A stochastic matrix P is called* **aperiodic** *is there exists an integer $N \geq 1$ such that $P^N > 0$. A stochastic matrix P is called* **irreducible** *if for any pair (i, j), there exists an integer $n = n(i, j) \geq 1$ such that $p^{(n)}_{i,j} > 0$ where $p^{(n)}_{i,j}$ is the (i, j)-element of P^n*

The following result is well known. For a proof, we refer to [3].

Theorem 1.1. *If a stochastic matrix $P \geq 0$ is irreducible, then 1 is a simple eigenvalue with a positive eigenvector. And 1 is the only eigenvalue with a non-negative eigenvector.*

Remark 1.1. *If a stochastic matrix P is irreducible, from Theorem 1.1, there exists a unique* **left** *(resp.* **right***)* **probability eigenvector** *of* 1*. We call it as the left (right resp.) probability eigenvector of P.*

Now we will state two classical limit theorems about stochastic matrices whose proofs are in [3].

Theorem 1.2. *(Limit theorem for aperiodic stochastic matrix). Suppose that P is an $k \times k$ aperiodic stochastic matrix. Then we have*

$$\lim_{n\to+\infty} P^n = e^T p,$$

where $e = \overbrace{(1, \cdots, 1)}^{k}$, p is the unique left probability eigenvector of P. And the (p, P) Markov shift is mixing.

Remark 1.2. *In the case that P is aperiodic, the i-coordinate of p, p_i is the probability that the system will, in the limit, be the state s_i. And it is independent of the initial states of the experiment. That is, for any probability vector q, we have*

$$qP^n \to qP^\infty = qe^T p = p.$$

Theorem 1.3. *(Limit theorem for irreducible stochastic matrix). Suppose that P is an irreducible $k \times k$ stochastic matrix. Then we have*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=0}^{n-1} P^i = e^T p,$$

where $e = \overbrace{(1, \cdots, 1)}^{k}$, p is the unique left probability eigenvector of P. And the (p, P) Markov shift is ergodic.

2. Main Theorems

The goal of this paper is to generalize the previous two limit theorems to any stochastic matrices.

At first we will introduce the notion of quasi-stochastic matrix which will be used in the proof of our results. A non-negative $k \times k$ matrix $Q = (q_{i,j})$ is called **quasi-stochastic** if either $k = 1$ and $Q = (q) < 1$ or $k \geq 2$ and Q is irreducible which satisfies for all i,

$$\sum_j q_{i,j} \leq 1$$

and for some $1 \leq i_0 \leq k$,

$$\sum_j q_{i_0,j} < 1.$$

Proposition 2.1. *Let Q be a quasi-stochastic matrix. Then the spectral radius of Q, $\rho(Q)$, is strictly less than* 1*.*

For proving the result, we need the classical Perron-Frobenious theorem (for its proof and many applications, we refer to [2]):

Theorem 2.1. *(Perron-Frobenious Theorem) Let A be a non-negative $k \times k$ matrix, $\rho = \rho(A)$ its spectral radius. Then the following hold:*

(i) ρ is an eigenvalue of A with a non-negative eigenvector.

(ii) If A is aperiodic, then the absolute value of any other eigenvalue is strictly less that ρ.

(iii) If A is irreducible, then ρ is a simple eigenvalue of A with a strictly positive eigenvector.

Proof of Proposition 2.1. From the Perron-Frobenius theorem, there exists a strictly positive probability left eigenvector q of ρ. From $qQ = \rho q$, we have

$$\rho = \sum_j \rho q_j = \sum_{i,j} q_i q_{i,j} = \sum_i q_i \sum_j q_{i,j} < \sum_i q_i = 1.$$

□

Fix any stochastic matrix $P = (p_{i,j})_{1 \le i,j \le k}$ for $k \ge 2$ and let $\mathcal{S} = \{s_1, \cdots, s_k\}$ be a set of k states. We introduce a directed graph on $\mathcal{S}$ by $s_i \to s_j$ if and only if $p_{i,j} > 0$. The directed graph would have several connected components in general. For the Markov chain on each connected component is independent, we will study each component separately. Thus we may assume that the directed graph is connected.

First we introduce an equivalent relation " $\sim$ " on $\mathcal{S}$ by the direction $\to$. We say that $s_i \sim s_j$ if and only if there exist $i_1, \cdots, i_n$ such that

$$s_i \to s_{i_1} \to s_{i_2} \to \cdots \to s_{i_n} \to s_j$$

and there exist $j_1, \cdots, j_m$ such that

$$s_j \to s_{j_1} \to s_{j_2} \to \cdots \to s_{j_m} \to s_i.$$

It is easy to see that $\sim$ is an equivalent relation on $\mathcal{S}$. An equivalent class of $\sim$ is called a **block**. We denote the blocks by $B_1, \cdots, B_l$. Now we have **direction** among these blocks defined by $B_i \to B_j$ if and only if there exist some $s_{i_0} \in B_i$ and some $s_{j_0} \in B_j$ with $s_{i_0} \to s_{j_0}$. It is easy to see that with the direction, the set of blocks, $\{B_1, \cdots, B_l\}$, becomes a connected directed graph without cycle. Consequently, the set of blocks $\{B_1, \cdots, B_l\}$ becomes a connected directed tree with the direction " $\to$ ".

We say that a block B is in **the first level** if there no direction from B to any other blocks. We say that a block B which is not in the first level is in **the second level** if it has and only has directions from it to some blocks in the first level. A block B which is not in the first two level is said to be in **the third level** if it has and only has directions from it to some blocks in the first two levels. Similarly, we can introduce the notion of i-th level for $i \in \mathbb{N}$.

Next, we classify the blocks into levels. For this, we rearrange the blocks as following: First we put the blocks of the first level, and then the blocks of the second level, and so on.

Suppose that $\{B_1, \cdots, B_l\}$ has r levels, and the number of blocks of the i-th level is $n(i)$ for each $i = 1, \cdots, r$. Let $\{B_{i,1}, \cdots, B_{i,n(i)}\}$ be the set of blocks of the i-th level. Let $\{t_1, \cdots, t_k\}$ be the arrangement of $\{s_1, \cdots, s_k\}$ corresponding to the previous rearrangement of the blocks. We will also denote the corresponding stochastic matrix by P. Then P has the following diagonal form:

$$\begin{pmatrix} \begin{pmatrix} A_{1,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_{1,n(1)} \end{pmatrix} & 0 & 0 & 0 \\ \vdots & \begin{pmatrix} A_{2,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_{2,n(2)} \end{pmatrix} & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ \vdots & \vdots & \vdots & \begin{pmatrix} A_{r,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_{r,n(r)} \end{pmatrix} \end{pmatrix},$$

where $A_{i,j}$ is the corresponding transition matrix of the block $B_{i,j}$. For $i = 1$, that is, when the block $B_{1,j}$ is in the first level, $A_{1,j}$ is an irreducible stochastic matrix. For $i \geq 2$, that is, the block $B_{i,j}$ is not in the first level, $A_{i,j}$ is a quasi-stochastic matrix. By Theorem 1.3, for any $1 \leq j \leq n(1)$,

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} A^n_{1,j} = A^\infty_j = e^T_{1,j} p_{1,j},$$

where $e_{1,j} = \overbrace{(1, \cdots, 1)}^{n_{1,j}}$, $p_{1,j}$ is the unique left probability eigenvector of $A_{1,j}$, and $n_{1,j}$ is the number of states in the block $B_{1,j}$.

Denote by D the matrix

$$\begin{pmatrix} \begin{pmatrix} A_{2,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_{2,n(2)} \end{pmatrix} & 0 & 0 \\ \vdots & \ddots & 0 \\ \vdots & \vdots & \begin{pmatrix} A_{l,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_{l,n(l)} \end{pmatrix} \end{pmatrix}.$$

We have the decomposition of P by

$$\begin{pmatrix} \begin{pmatrix} A_{1,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_{1,n(1)} \end{pmatrix} & 0 \\ C_1 \quad \cdots \quad C_{n(1)} & D \end{pmatrix}.$$

We also decompose P^n by

$$\begin{pmatrix} \begin{pmatrix} A_{1,1}^n & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_{1,n(1)}^n \end{pmatrix} & 0 \\ C_1^{(n)} \quad \cdots \quad C_{n(1)}^{(n)} & D^n \end{pmatrix}.$$

¿From these preparing, we can state our theorems as following.

Theorem 2.2. *Let P be any stochastic matrix. $p_{(1,j)}$ and $e_{(1,j)}$ are as before. Then we have*

$$\lim_{n\to+\infty} D^n = 0,$$

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=0}^{N-1} A_{1,i}^n = e_{(1,i)}^T p_{(1,i)},$$

and

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=0}^{N-1} C_i^{(n)} = (1-D)^{-1} C_i e_{(1,i)}^T p_{(1,i)}$$

for $1 \le i \le n(1)$.

Theorem 2.3. *Moreover, if $A_{1,i}$ are aperiodic for all $1 \le i \le n(1)$, we have*

$$\lim_{n\to+\infty} C_i^{(n)} = (1-D)^{-1} C_i e_{(1,i)}^T p_{(1,i)}$$

for $1 \le i \le n(1)$.

Corollary 2.1. *If there is only one block B in the first level, and the corresponding transition matrix, A, is aperiodic, we have*

$$\lim_{n\to+\infty} C^n = e_2^{\perp} p,$$

where $e_2 = \overbrace{(1,\cdots,1)}^{k-m}$, p is the unique left eigenvector of A, and m is the number of states in the block B.

For simplicity, we will prove Theorem 2.2. The proof of Theorem 2.1 is similar to that of Theorem 2.2.

The Proof of Theorem 2.2. It is easy to see that every diagonal matrix $A_{i,j}$ in B is quasi-stochastic for $i \ge 2$, $1 \le j \le n(i)$. By Proposition 2.1, the spectral radius of $A_{i,j}$, $\rho(A_{i,j})$, is less than 1. So the spectral radius of D, $\rho(D)$, is less than 1. Thus, there exist constant $M > 0$ and $0 \le \lambda < 1$ such that

$$\|D^n\| \le M\lambda^n$$

for all $n \ge 1$. Here $\| \ \|$ is the operator norm of matrix.

For simplicity, we will denote $A_{1,i}$ by A_i for $1 \le i \le n(1)$. By computation, we have that

$$C_i^{(n)} = C_i A_i^{n-1} + DC_i A_i^{n-2} + \cdots + D^{n-2} C_i A_i + D^{n-1} C_i$$

for all $n \geq 1$ and $i = 1, 2$.

For all $1 \leq i \leq n(1)$, let $A_i^\infty = e_{(1,i)}^T p_{(1,i)}$. By Theorem 1.3, we have

$$\lim_n A_i^n = A_i^\infty.$$

Thus, $\{A_i^n\}$ is bounded. There exists $K > 0$ such that $\|A_i^n\| \leq K$ for all $1 \leq n \leq \infty$, $1 \leq i \leq n(1)$.

Fix any $\epsilon > 0$ and choose $N_1 > 0$ with $\lambda^{N_1} \leq \epsilon$ and

$$\| \sum_{n \geq N_1+1} D^n C A_i^\infty \| \leq \epsilon$$

for $1 \leq i \leq n(1)$. Choose $N_2 > 0$ ($N_2 \geq N_1$) such that

$$\|A_i^n - A_i^\infty\| \leq \epsilon$$

holds for any $n \geq N_2$ and $1 \leq i \leq n(1)$.

Then we have

$$\begin{aligned}
& \|C_i^{(n)} - (C_i A_i^\infty + D C_i A_i^\infty + \cdots + D^{n-1} C_i A_i^\infty)\| \\
= & \|(C_i A_i^{n-1} + B C_i A_i^{n-2} + \cdots + D^{n-N_2-1} C_i A_i^{N_2}) \\
& +(D^{n-N_2} C_i A_i^{N_2-1} + \cdots + D^{n-1} C_i) \\
& -((C_i A_i^\infty + D C_i A_i^\infty + \cdots + D^{n-N_2-1} C_i A_i^\infty) \\
& +(D^{n-N_2} C_i A_i^\infty + \cdots + D^{n-1} C_i A_i^\infty))\| \\
\leq & (1 + M\lambda + \cdots + M\lambda^{n-N_2-1})\|C_i\|\epsilon \\
& +2K\|C_i\|(M\lambda^{n-N_2} + M\lambda^{n-N_2+1} + \cdots + M\lambda^{n-1}) \\
\leq & (1 + \frac{M\lambda}{1-\lambda})\|C_i\|\epsilon + 2K\|C_i\|\frac{M\lambda^{N_1}}{1-\lambda} \\
\leq & K_1\epsilon,
\end{aligned}$$

for $n \geq 2N_2$.

Consequently, we get

$$\begin{aligned}
& \|C_i^{(n)} - (1-D)^{-1} C_i e_{(1,i)}^T p_{(1,i)}\| \\
= & \|C_i^{(n)} - \sum_{j=0}^{+\infty} D^j C_i A_i^\infty\| \\
\leq & \|C_i^{(n)} - \sum_{j=0}^{n-1} D^j C_i A_i^\infty\| + \|\sum_{j=n}^{+\infty} D^j C_i A_i^\infty\| \\
\leq & (K_1 + 1)\epsilon
\end{aligned}$$

for any $n \geq 2N_2$.

For $\epsilon > 0$ is arbitrary, we have proved that for all $1 \leq i \leq n(1)$,

$$\lim_{n \to +\infty} C_i^{(n)} = (1-D)^{-1} C_i A_i^\infty,$$

and so completes the proof. □

Remark 2.1. *The dimension of eigenspace of* 1 *of any stochastic matrix is the number of the blocks in the first level. And every probability eigenvector of* 1 *is some convex combination of the unique left probability eigenvectors* p_i *corresponding to the block* B_i *of the first level. So if* 1 *is a simple eigenvalue if and only if* P *is connected and there is only one block in the first level.*

Remark 2.2. *Suppose that* P *is connected, then there exists a positive probability left eigenvector of* P *if and only if* P *is irreducible. When* P *is not connected, then there exists a positive probability left eigenvector of* P *if and only if the transition matrix of every connected component is irreducible.*

References

[1] Feller, W., An introduction to probability theory and its applications, Vol 1, Third edition, 1968, John Wiley and Sons, Inc. New York.

[2] MacCluer C. R., The Many Proofs and Applications of Perron's Theorem, *SIAM Review*, 2000, 42(3): 487-498.

[3] Walters, P., An introduction to ergodic theory, Grad. Text in Math., Spinger-Verlag (1981).

Keonhee Lee: Department of Mathematics, Chungnam National University, Daejeon, 305-764, Korea

E-mail address: khlee@math.cnu.ac.kr

Yong Zhang: Department of Mathematics, Chungnam National University, Daejeon, 305-764, Korea; Department of Mathematics, Suzhou University, Suzhou, China

E-mail address: yzhang@math.cnu.ac.kr; yongzhang@suda.edu.cn