Abstract. In this paper, we attempt to answer the three questions about the invariant probability distribution for stochastic matrices: (1) does every stochastic matrix have an invariant probability distribution?; (2) is the invariant probability distribution unique?; and (3) when can we conclude that the power of a stochastic matrix converges? To answer these questions, we present the Perron-Frobenius Theorem about matrices with positive entries.

1. Background of Probability and Markov Property

Let $A$ and $B$ be two events. The conditional probability of $A$ given $B$, denoted by $Pr\{A|B\}$, is defined by

$$Pr\{A|B\} = \frac{Pr\{A \cap B\}}{Pr\{B\}}.$$ 

Let $X_n$ be a discrete-time stochastic process, by which we mean a sequence of random variables indexed by time $n = 1, 2, \cdots$. Suppose that the process takes values in the set $S = \{1, \cdots, N\}$. We call the possible values of $X_n$ the states of the process. In this paper, we restrict our attention to an important class of processes, called Markov chain. Roughly speaking, a Markov chain is a process with the property that given the present state, the future and past states are independent. We state the following definition:

**Definition 1.1.** A Markov chain is a sequence of random variables $X_1, X_2, X_3, \cdots$, such that

$$Pr\{X_{n+1} = x|X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} = Pr\{X_{n+1} = x|X_n = x_n\}.$$ 

We also assume another important property, called time-homogeneity. This property states that the transition probabilities do not depend on the time of the process.

Formally, we have the following definition

*Date: July 29, 2011.*
Definition 1.2. A *time-homogenous* Markov chain is a Markov chain that satisfies
\[ Pr\{X_{n+1} = j|X_n = i\} = Pr\{X_2 = j|X_1 = i\} \]
for all \( n \geq 1 \).

Thus, we can define *transition probabilities* \( p(i, j) \) and \( p_n(i, j) \) by
\[ p(i, j) = Pr\{X_{k+1} = j|X_k = i\}, \]
and
\[ p_n(i, j) = Pr\{X_{n+k} = j|X_k = i\}. \]

2. The existence of invariant probability distribution

A (right) *stochastic matrix* \( Q \) is a square matrix each of whose rows consists of non-negative real entries \( Q(i, j) \) with each row summing up to 1, i.e., \( \sum_j Q(i, j) = 1 \).

Given a Markov chain with transition probabilities \( p_n(i, j) \), we can define its \( (n) \)th *transition matrix*, denoted by \( P_n \), to be the stochastic matrix with entries
\[ P_n(i, j) = p_n(i, j), \]
and we write \( P = P_1 \). Note that
\[ p_n(i, j) = \sum_k p(i, k)p_{n-1}(k, j), \]
which follows from the definition of a time-homogenous Markov chain. Using induction, one can show \( P_n = P^n \).

A row vector \( \vec{v} \) is a *probability vector* if all the components are non-negative and sum to 1. A probability vector \( \vec{\pi} \) is an *invariant probability distribution* for stochastic matrix \( P \) if \( \vec{\pi}P = \vec{\pi} \). In other words, an invariant probability distribution of \( P \) is a left eigenvector of \( P \) with eigenvalue 1.

**Lemma 2.1.** Any stochastic matrix has at least one invariant probability distribution.

*Proof.* We can check that for any stochastic matrix \( P \), the vector
\[ \vec{\pi}_0 = \begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} \]
is always a right eigenvalue of \( P \) with eigenvalue 1.

Let
\[ P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \]
with $\sum_j p_{ij} = 1$. Then,

$$P\pi_0 = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \sum_j p_{1j} \\ \vdots \\ \sum_j p_{nj} \end{pmatrix} = \frac{1}{n} \pi_0$$

Thus, at least one left eigenvector with eigenvalue one exists. \(\square\)

Generally speaking, the invariant probability distribution is not unique. For example, consider a stochastic matrix

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$ For any probability vector $\vec{v} = (x, 1-x)$ with any real number $x$, we have

$$\vec{v}P = (x, 1-x) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = (x, 1-x) = \vec{v}.$$

3. The Uniqueness of Invariant Probability Distribution

The Perron-Frobenius Theorem ensures that for a matrix with strictly positive entries, the invariant probability distribution is always unique. Before we state the theorem, we give the following definition.

**Definition 3.1.** For vectors $\vec{v} = (v_1, \cdots, v_n)$ and $\vec{u} = (u_1, \cdots, u_n)$, we have $\vec{v} \geq \vec{u}$ if $v_i \geq u_i$ for each $i$, and $\vec{v} > \vec{u}$ if $v_i > u_i$ for each $i$.

**Theorem 3.2** (Perron-Frobenius Theorem). Let $P$ be an $n \times n$ matrix with positive entries. Then $P$ has a dominant eigenvalue $\alpha$ such that

1. $\alpha > 0$, and its associated eigenvector has all positive entries;
2. Any other eigenvalue $k$ of $P$ satisfies $|k| < \alpha$;
3. $\alpha$ is simple;
4. $P$ has no other eigenvector with all non-negative entries.

We complete details of the proof as outlined in [1].

**Proof.** Let $P = (p_{ij})$ be a $n \times n$ matrix with $p_{ij} > 0$. If $\vec{v} \geq \vec{0}$, and $\vec{v} \neq \vec{0}$, then it is easy to show that $P\vec{v} > \vec{0}$. Define $G := \{\lambda > 0 : P\vec{v} \geq \lambda\vec{v} \text{ for } \vec{v} \in [0, \infty)^n \text{ and } \vec{v} \neq \vec{0}\}$. Let $\alpha$ be the maximum in $G$. Then we have: $P\vec{v} \geq \alpha\vec{v}$ for some $\vec{v} \in [0, \infty)^n$.

We claim that actually we have $P\vec{v} = \alpha\vec{v}$ for some $\vec{v}$ with positive entries. Suppose not, then $P\vec{v} - \alpha\vec{v} > \vec{0}$. Given $p_{ij} > 0$, we have $P(P\vec{v}) - \alpha(P\vec{v}) > \vec{0}$. Thus, there exists $\alpha' > \alpha > 0$ such that $P(P\vec{v}) - \alpha'(P\vec{v}) \geq \vec{0}$. 


Since we proved \( P\tilde{v} > \tilde{0} \), then \( P\tilde{v} \in [0, \infty)^n \). Thus, \( \alpha' \in G \), contradicting that \( \alpha \) is the maximum in \( G \). Thus \( P\tilde{v} = \alpha\tilde{v} \). Thus we can write \( \tilde{v} = P\tilde{v}/\alpha \). Since \( P\tilde{v} > \tilde{0} \), and \( \alpha > 0 \), we have \( \tilde{v} > \tilde{0} \). This proves (1).

Next, we prove that there is a unique \( \tilde{v} \) such that \( P\tilde{v} = \alpha\tilde{v} \). If not, suppose \( P\tilde{w} = \alpha\tilde{w} \) for \( \tilde{w} \) distinct from any constant multiple of \( \tilde{v} \). By (1), we know that \( \tilde{w} > \tilde{0} \). Then there exists \( c \in \mathbb{R} \) such that \( \tilde{w} + c\tilde{v} \geq \tilde{0} \), \( \tilde{w} + c\tilde{v} \neq \tilde{0} \), and \( \tilde{w} + c\tilde{v} \) has zero entries.

Since \( P\tilde{w} = \alpha\tilde{w} \), and \( cP\tilde{v} = P(c\tilde{v}) = c\alpha\tilde{v} = \alpha(c\tilde{v}) \), we have \( P\tilde{w} + P(c\tilde{v}) = \alpha\tilde{w} + \alpha(c\tilde{v}) \). That is, \( P(\tilde{w} + c\tilde{v}) = \alpha(\tilde{w} + c\tilde{v}) \). Now since \( \tilde{w} + c\tilde{v} \geq \tilde{0} \), we have \( P(\tilde{w} + c\tilde{v}) > \tilde{0} \). Thus, \( \alpha(\tilde{w} + c\tilde{v}) > \tilde{0} \). In particular, \( \tilde{w} + c\tilde{v} > \tilde{0} \) given \( \alpha > 0 \), which is contradiction.

Now suppose that \( P\tilde{w} = kw \) for \( k \neq \lambda \). Then \( |k||w| = |kw| = |P\tilde{w}| \leq |P||\tilde{w}| = P|\tilde{w}| \), where \( |\tilde{w}| = (|w_1|, \ldots, |w_n|) \). Therefore, \( |k| \in G \), and hence \( |k| \leq \alpha \). This proves (2).

To prove (3), let \( B \) be any \((n-1) \times (n-1)\) submatrix of \( P \). We claim that all the eigenvalues of \( B \) have absolute value strictly less than \( \alpha \). Our claim rests upon the following lemma:

**Lemma 3.3.** If \( 0 \leq B \leq P \), the every eigenvalue \( \beta \) of \( B \) satisfies \( |\beta| < \alpha \).

The proof of this lemma follows the same reasoning as the proof of (1) and (2). We leave the proof of this lemma as a simple exercise to the readers. The hint is that given \( 0 \leq B \leq P \) and \( B\tilde{z} = \beta\tilde{z} \) for some \( \tilde{z} \neq \tilde{0} \), we have \( |\beta| |\tilde{z}| \leq B |\tilde{z}| \leq P |\tilde{z}| \). Hence, \( |\beta| \in G \) and \( |\beta| \leq \alpha \). For the details of the proof of Lemma 3.3, readers should consult [3].

Now consider \( f(\lambda) = \text{det}(\lambda I - P) \). Expanding \( \text{det}(\lambda I - P) \) along its \( i^{th} \) row, we have \( \frac{\partial}{\partial \lambda} \text{det}(\lambda I - P) = \text{det}(\lambda I - P_i) \), where \( P_i \) is the matrix obtained by eliminating the \( i^{th} \) row and \( i^{th} \) column of \( P \). Thus, by chain rule, \( f'(\lambda) = \frac{\partial}{\partial \lambda} \text{det}(\lambda I - P) = \sum_{i=1}^{n} \text{det}(\lambda I - P_i) \). By Lemma 3.3, we see that each of the matrices \( \alpha I - P_i \) has strictly positive determinant. That is, \( f'(\alpha) > 0 \), and therefore \( \alpha \) is simple. This proves (3).

Now we apply (1),(2), and (3) to \( PT \), which has the same eigenvalues as \( P \). Then there exists a non-negative and non-trivial \( \tilde{w} \) such that \( P^T\tilde{w} = \lambda\tilde{w} \). Suppose \( \tilde{h} \) is another non-negative eigenvector of \( P \) with eigenvalue \( \lambda' \neq \lambda \), that is, \( P\tilde{h} = \lambda'\tilde{h} \). Then \( \lambda'\langle \tilde{w}, \tilde{h} \rangle = \langle P^T\tilde{w}, \tilde{h} \rangle = \langle P^T\tilde{w}, \tilde{h} \rangle = \lambda\langle \tilde{w}, \tilde{h} \rangle \). Since \( \lambda' \neq \lambda \), we can only have \( \langle \tilde{w}, \tilde{h} \rangle = 0 \), which is impossible, if both \( \tilde{w} \) and \( \tilde{h} \) have all non-negative entries.

Thus, the invariant probability vector is unique given the stochastic matrix has positive entries.

4. **The convergence of stochastic matrices**

The Perron-Frobenius Theorem about stochastic matrices with positive entries does not cover all matrices with an invariant probability distribution.
For example, consider
\[ P = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}. \]
We observe that \( P^2 = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \) with all positive entries. Since the eigenvalues of \( P^n \) is the \( n^{th} \) power of the eigenvalues of \( P \), and the eigenvectors of \( P^n \) are the same as those of \( P \), we can easily conclude the following:

**Proposition 4.1.** If \( P \) is a stochastic matrix such that for some \( n \), \( P^n \) has all strictly positive entries. Then \( P \) satisfies the conclusions of the Perron-Frobenius Theorem.

We now attempt to characterise such \( P \).

We say that two states \( i \) and \( j \) of Markov Chain communicate with each other if there exist \( m, n \geq 0 \) such that \( p_m(i, j) > 0 \), and \( p_n(j, i) > 0 \). In this case, we write \( (i \leftrightarrow j) \).

**Lemma 4.2.** If \( (i \leftrightarrow j) \) and \( (i \leftrightarrow k) \), then \( (i \leftrightarrow k) \).

We follow the proof presented in [1]

**Proof.** If \( p_{m_1}(i, j) > 0 \), and \( p_{m_2}(i, j) > 0 \), then
\[
\begin{align*}
p_{m_1+m_2}(i, k) &= \Pr\{X_{m_1+m_2} = k | X_0 = i\} \\
&\geq \Pr\{X_{m_1+m_2} = k, X_{m_1} = j | X_0 = i\} \\
&= \Pr\{X_{m_1} = j | X_0 = i\} \Pr\{X_{m_1+m_2} = k | X_0 = j\} \\
&= p_{m_1}(i, j)p_{m_2}(j, k) > 0
\end{align*}
\]

\( \square \)

**Remark 4.3.** Lemma 4.2 gives transitivity, so that \((\leftrightarrow)\) is an equivalence relation. This equivalence relation partitions the state space into disjoint sets called communication classes.

If there is only one communication class, then the chain is called irreducible. This is the case when for any \( i \) and \( j \), there exists \( n = n(i, j) \) with \( p_n(i, j) > 0 \). Suppose that \( P \) is a matrix for an irreducible Markov Chain. We define the period of state \( i \), denoted \( d = d(i) \), to be the greatest common divisor of \( J_i = \{ n \geq 0 : p_n(i, i) > 0 \} \). Note that \( J_i \) is closed under addition; that is, if \( p_m(i) > 0 \) and \( p_n(i) > 0 \), then \( p_{m+n}(i) \geq p_m(i)p_n(i) > 0 \). We call an irreducible matrix \( P \) aperiodic if \( d = 1 \).

**Lemma 4.4.** If \( P \) is irreducible and aperiodic, then there exists \( M > 0 \), such that for all \( n > M \), \( P^n \) has all entries strictly positive.

We present the proof of Lemma 4.4 cited in [1]

**Proof.** Since \( P \) is irreducible, for any \( (i, j) \), there exists \( m(i, j) \) so that \( p_{m(i,j)}(i,j) > 0 \). Since \( P \) is aperiodic, there exists \( M(i) \) such that for all \( n > M(i) \), \( p_n(i, i) > 0 \). Thus, for any \( n \geq M(i) \), \( p_{n+m(i,j)}(i,j) \geq p_n(i,j)p_{m(i,j)}(i,j) > 0 \). Since the state space is finite, we can consider the maximum value \( M \) of \( M(i) + m(i,j) \) over all \( (i, j) \). Then for any \( n \geq M \) and for all \( (i, j) \), \( p_n(i, j) > 0 \). \( \square \)

Given Lemma 4.4, we deduce the following theorem:
Theorem 4.5. If $P$ is a stochastic matrix for an irreducible and aperiodic Markov Chain, then

1. there exists a unique invariant probability vector such that $\pi P = \pi$;
2. $\pi$ can be computed by $\lim_{n \to \infty} p_n(i,j) = \pi_{ij}$, and if $\pi$ is any initial probability vector, $\lim_{n \to \infty} \pi P^n = \pi$.

We complete details of the proof as presented in [4].

Proof. We have proved (1) in the proof of the Perron-Frobenius Theorem. For (2), it suffices to consider $P$ with positive entries. If $P$ contains zero entries, Lemma 4.4 tells us that we can raise $P$ to a power $n$ such that $P^n$ has all positive entries. Consider the $(i,j)$ entry of $P^{t+1} = PP^t$, that is, $p_{(t+1)}(i,j) = \sum_k p(i,k)p_{(t)}(k,j)$. Let $m^{(1)}_j := \min_k p_{(t)}(i,j)$, and $M^{(1)}_j := \max_k p_{(t)}(i,j)$, so that $0 \leq m^{(1)}_j \leq M^{(1)}_j \leq 1$. Then

$$m^{(t+1)}_j = \min_k \sum p(i,k)p^{(t)}(k,j) \geq \min_k \sum p(i,k)p^{(t)}(i,j)$$

$$= \min_k p^{(t)}(i,j) \sum p(i,k) = m^{(t)}_j = \max_k p^{(t)}(i,j)$$

Thus, the sequence $\{m^{(1)}_j, m^{(2)}_j, \ldots \}$ is non-decreasing. Similarly, the sequence $\{M^{(1)}_j, M^{(2)}_j, \ldots \}$ is non-increasing. Since $\{m^{(1)}_j, m^{(2)}_j, \ldots \}$ is bounded by $M^{(1)}_j$, and $\{M^{(1)}_j, M^{(2)}_j, \ldots \}$ is bounded by $m^{(1)}_j$, there exist $m_j$ and $M_j$ such that $\lim_{t \to \infty} m^{(t)}_j = m_j$ and $\lim_{t \to \infty} M^{(t)}_j = M_j$, and $m_j \leq M_j$.

We now prove that $m_j = M_j$. We compute:

$$M^{(t+1)}_j - m^{(t+1)}_j = \max_k \sum p(i,k)p^{(t)}(k,j) - \min_k \sum p(i,k)p^{(t)}(k,j)$$

$$= \max_k \sum p(i,k)p^{(t)}(k,j) - \min_k \sum p(i,k)p^{(t)}(k,j)$$

$$\leq \max_k \sum p(i,k)p^{(t)}(k,j) - \min_k \sum p(i,k)p^{(t)}(k,j)$$

$$\leq m^{(t+1)}_j - m^{(t+1)}_j, \ast$$

where $\sum_k (p(i,k) - p(l,k))^+$ denotes the sum of positive terms $p(i,k) - p(l,k) > 0$, and $\sum_k (p(i,k) - p(l,k))^-$ denotes the sum of negative terms $p(i,k) - p(l,k) < 0$.

If we define

$$\sum_k (p(i,k) - p(l,k)) = \sum_k (p(i,k) - p(l,k))^-, \ast$$

and similarly

$$\sum_k (p(i,k) - p(l,k)) = \sum_k (p(i,k) - p(l,k))^+, \ast$$

then
\[
\sum_k (p(i, k) - p(l, k))^- = \sum_k (p(i, k) - p(l, k)) \\
= \sum_k p(i, k) - \sum_k p(l, k) \\
= (1 - \sum_k p(i, k)) - (1 - \sum_k p(l, k)) \\
= \sum_k (p(l, k) - p(i, k)) \\
= -\sum_k ((p(i, k) - p(l, k))^+ \\
\]

Thus, (\ref{eqn:8}) becomes
\[
0 \leq M_j^{(t+1)} - m_j^{(t+1)} \leq (M_j^{(t)} - m_j^{(t)})\max_{k,l} (p(i, k) - p(l, k))^+.
\]

If \( \max_{k,l} \sum_k (p(i, k) - p(l, k)) = 0 \), we have \( M_j^{(t+1)} = m_j^{(t+1)} \).

Otherwise, let \( r \) be the number of terms in \( k \) such that \( p(i, k) - p(l, k) > 0 \), and \( s \) be the number of terms in \( k \) such that \( p(i, k) - p(l, k) < 0 \). Then \( r \geq 1 \). Let \( n := r + s \geq 1 \). Denote \( \delta := \min_{i,j} p(ij) > 0 \). Thus we can write:
\[
\sum_k (p(i, k) - p(l, k))^+ = \sum_k^+ p(i, k) - \sum_k^+ p(l, k) \\
= 1 - \sum_k p(k, j) - \sum_k^+ p(l, k) \\
\leq 1 - s\delta - r\delta \\
= 1 - n\delta
\]

Thus, (\ref{eqn:8}) becomes:
\[
0 \leq M_j^{(t+1)} - m_j^{(t+1)} \leq (1 - \delta)(M_j^{(t)} - m_j^{(t)}) \leq (1 - \delta)^t (M_j^{(1)} - m_j^{(1)}) \longrightarrow 0.
\]
as \( t \longrightarrow \infty \). Denote \( \pi_j = M_j = m_j > 0 \). Given \( m_j^{(t)} \leq p(t)(i, j) \leq M_j(t) \), we have
\[
\lim_{n \rightarrow \infty} p(t)(ij) = \lim_{n \rightarrow \infty} m_j^{(t)} = \lim_{n \rightarrow \infty} M_j^{(t)} = \pi_j,
\]
and thus
\[
\lim_{t \rightarrow \infty} P^t = \begin{pmatrix}
\pi_1 \\
\vdots \\
\pi_\ell
\end{pmatrix}
\]

Moreover, for any probability vector \( \vec{v} \) with \( \sum_i v_i = 1 \), we have
\[
\lim_{t \rightarrow \infty} \vec{v}P^t = \begin{pmatrix}
v_1 & \cdots & v_n
\end{pmatrix} \begin{pmatrix}
\pi_1 & \pi_2 & \cdots & \pi_n \\
\pi_1 & \pi_2 & \cdots & \pi_n \\
\vdots & \vdots & \ddots & \vdots \\
\pi_1 & \pi_2 & \cdots & \pi_n
\end{pmatrix} \\
= \begin{pmatrix}
\pi_1 \sum_i v_i & \cdots & \pi_n \sum_i v_i \\
\pi_1 & \cdots & \pi_n
\end{pmatrix} = \vec{\pi}
\]
Acknowledgments. I would like to thank my mentor Marcelo Alvisio for introducing this topic to me. I would also like to thank James Buchanan for supervising my REU project. I also like to thank Professor Peter May for holding the REU program for Mathematics lovers.

REFERENCES


http://www.math.nus.edu.sg/ matsr/ProbII/Lec8.pdf


http://www.math.unl.edu/ bdeng1/Teaching/math428/Lecture Notes/PFTheorem.pdf