INVARIANT PROBABILITY DISTRIBUTIONS

BOTAO WU

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.

Contents

1. Background of Prabability and Markov Property 1
2. The existence of invariant probability distribution 2
3. The uniqueness of invariant probability distribution 3
4. The convergence of stochastic matrices 4
Acknowledgments 8
References 8

1. Background of Prabability 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

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

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 probablity 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

$$\vec{\pi_0} = \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix}$$

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 ,

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

Thus, at least one *left* eigenvector with eigenvalue one exists.

□

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 probablity vector $\vec{v} = \begin{pmatrix} x & 1-x \end{pmatrix}$ with any real number x, we have

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

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 α such that*

(1) $\alpha > 0$, and its associated eigenvector has all positive entries;
(2) Any other eigenvalue k of P satisfies $|k| < \alpha$;
(3) α 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 $G := \{\lambda > 0 : P\vec{v} \geq \lambda\vec{v} for\ \vec{v} \in [0, \infty)^n\ \ and\ \vec{v} \neq \vec{0}\}$. Let α 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 $P(P\vec{v}) - \alpha\vec{v} > \vec{0}$. Thus, there exists $\alpha' > \alpha > 0$ such that $P(P\vec{v}) - \alpha'(P\vec{v}) \geq \vec{0}$.

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

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

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

Now suppose that $P\vec{w} = k\vec{w}$ for $k \neq \lambda$. Then $|k||\vec{w}| = |k\vec{w}| = |P\vec{w}| \leq |P||\vec{w}| = P|\vec{w}|$, where $|\vec{w}| = (|w_1|, ..., |w_n|)$. 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 α. Our claim rests upon the following lemma:

Lemma 3.3. *If $0 \leq B \leq P$, the every eigenvalue β 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 $B\vec{z} = \beta\vec{z}$ for some $\vec{z} \neq \vec{0}$, we have $|\beta||\vec{z}| \leq B|\vec{z}| \leq P|\vec{z}|$. Hence, $|\beta| \in G$ and $|\beta| \leq \alpha$. For the details of the proof of Lemma 3.3, readers should consult [3].

Now consider $f(\lambda) = det(\lambda I - P)$. Expanding $det(\lambda I - P)$ along its ith row, we have $\frac{\partial}{\partial\lambda} det(\lambda I - P) = det(\lambda I - P_i)$, where P_i is the matrix obtained by eliminating the ith row and i^{th} column of P. Thus, by chain rule, $f'(\lambda) = \frac{d}{d\lambda} det(\lambda I - P) = \sum_{i=1}^{n} det(\lambda I - P_i)$. 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 α is simple. This proves (3).

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

$$\begin{aligned} p_{m_1+m_2}(i,k) &= Pr\{X_{m_1+m_2} = k | X_0 = i\} \\ &\geq Pr\{X_{m_1+m_2} = k, X_{m_1} = j | X_0 = i\} \\ &= Pr\{X_{m_1} = j | X_0 = i\} Pr\{X_{m_1+m_2} = k | X_0 = j\} \\ &= p_{m_1}(i,j) p_{m_2}(j,k) > 0 \end{aligned}$$

□

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 $p_m(ii) > 0$ and $p_n(ii) > 0$, then $p_{m+n}(ii) \geq p_m(ii) p_n(ii) > 0$. 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 irruducible, for any (i,j), there exists $m(i,j)$ so that $p_{m(i,j)}(ij) > 0$. 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 $M(i) + m(ij)$ over all (i,j). Then for any $n \geq M$ and for all (i,j), $p_n(i,j) > 0$. □

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 $\vec{\pi}P = \vec{\pi}$;
(2) $\vec{\pi}$ can be computed by $\lim_{n\to\infty} p_n(i,j) = \pi_j$, and if $\vec{v}$ is any initial probability vector, $\lim_{n\to\infty} \vec{v}P^n = \vec{\pi}$

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

Proof. We have proved (1) in the proof of the 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 $m_j^{(t)} := min_i p_{(t)}(i,j)$, and $M_j^{(t)} := max_i p^{(t)}(i,j)$, so that $0 \le m_j^{(t)} \le M_j^{(t)} \le 1$. Then

$$m_j^{(t+1)} = min_i \sum_k p(i,k)p^{(t)}(k,j) \ge min_i \sum_k p(i,k)p^{(t)}(i,j)$$

$$= min_i p^{(t)}(i,j) \sum_k p(i,k) = m_j^{(t)} \cdot 1 = m_j^{(t)}.$$

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

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

$$\begin{aligned}
M_j^{(t+1)} - m_j^{(t+1)} &= max_i \sum_k p(i,k)p^{(t)}(k,j) - min_l \sum_k p(l,k)p^{(t)}(k,j) \\
&= max_{i,l} \sum_k p^{(t)}(k,j)(p(i,k) - p(l,k)) \\
&= max_{i,l}[\sum_k p^{(t)}(k,j)(p(i,k) - p(l,k))^+ + \sum_k p(k,j)(p(i,k) - p(l,k))^-] \\
&\le max_{i,l}[M_j^{(t)} \sum_k (p(i,k) - p(l,k))^+ + m_i^{(t)} \sum_k (p(i,k) - p(l,k))^-], (*)
\end{aligned}$$

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

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

and similarly

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

then

$$\begin{aligned}
\sum_k (p(i,k)-p(l,k))^- &= \sum_k^- (p(i,k)-p(l,k)) \\
&= \sum_k^- p(i,k) - \sum_k^- p(l,k) \\
&= (1-\sum_k^+ p(i,k) - (1-\sum_k^+ p(l,k)) \\
&= \sum_k^+ (p(l,k)-p(i,k)) \\
&= -\sum_k ((p(i,k)-p(l,k))^+
\end{aligned}$$

Thus, $(*)$ becomes

$$0 \le M_j^{(t+1)} - m_j^{(t+1)} \le (M_j^{(t)} - m_j^{(t)}) max_{j,l} \sum_k (p(i,k)-p(l,k))^+.$$

If $max_{j,l} \sum_k (p(i,k)-p(l,k)) = 0$, 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 $r \ge 1$. Let $n := r + s \ge 1$. Denote $\delta := min_{i,j} p(ij) > 0$. Thus we can write:

$$\begin{aligned}
\textstyle\sum_k (p(i,k)-p(l,k))^+ &= \textstyle\sum_k^+ p(i,k) - \sum_k^+ p(l,k)) \\
&= 1 - \textstyle\sum_k^- p(k,j) - \sum_k^+ p(l,k)) \\
&\le 1 - s\delta - r\delta \\
&= 1 - n\delta
\end{aligned}.$$

Thus, $(*)$ becomes:

$$0 \le M_j^{(t+1)} - m_j^{(t+1)} \le (1-\delta)(M_j^{(t)} - m_j^{(t)}) \le (1-\delta)^t (M_j^{(1)} - m_j^{(1)}) \longrightarrow 0.$$

as $t \longrightarrow \infty$. Denote $\pi_j = M_j = m_j > 0$. Given $m_j^{(t)} \le p_{(t)}(i,j) \le M_j(t)$, we have

$$\lim_{n\to\infty} p^{(t)}(ij) = \lim_{n\to\infty} m_j^{(t)} = \lim_{n\to\infty} M_j^{(t)} = \pi_j,$$

and thus

$$\lim_{t\to\infty} P^t = \begin{pmatrix} \vec{\pi} \\ \vdots \\ \vec{\pi} \end{pmatrix}.$$

Moreover, for any probability vector $\vec{v}$ with $\sum_i v_i = 1$, we have

$$\begin{aligned}
\lim_{t\to\infty} \vec{v}P^t &= \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}. \\
&= \begin{pmatrix} \pi_1 \sum_i v_i & \cdots & \pi_n \sum_i v_i \end{pmatrix} \\
&= \begin{pmatrix} \pi_1 & \cdots & \pi_n \end{pmatrix} = \vec{\pi}
\end{aligned}$$

□

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

[1] Lawler, G.F.(1995). *Introduction to Stochastic Processes.* USA: Chapman & Hall.

[2] Department of Mathematics of National University of Singapore. (2011). *Lecture 8. Perron-Frobenius theorem.* Retrieved from:
http://www.math.nus.edu.sg/ matsr/ProbII/Lec8.pdf

[3] Sternburg, S.(2011). *The Perron-Frobenius theorem.* Retrieved Jul. 15, 2011, from Harvard University, Department of Mathematics Web site:
http://www.math.harvard.edu/library/sternberg/slides/1180912pf.pdf.

[4] Deng, B.(2011) *Proof of the Perron-Frobenius Theorem.* Retrieved Jul. 12, 2011, from University of Nebraska-Lincoln, Department of Mathematics Web site:
http://www.math.unl.edu/ bdeng1/Teaching/math428/Lecture Notes/PFTheorem.pdf