Let \( \{X_n\}_{n \geq 0} \) be a discrete-time stochastic process with countable state-space \( E \). If, for all integers \( n \geq 0 \) and all states \( i_0, i_1, \ldots, i_{n-1}, i, j \) in \( E \),

\[
P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i),
\]

whenever both sides are well-defined, this stochastic process is called a Markov Chain. It is called a homogeneous Markov Chain (abbreviated as HMC) if in addition, the right hand side in the above equation is independent of \( n \).

Transition probabilities for a homogeneous Markov Chain are defined as

\[
p_{ij} = P(X_{n+1} = j | X_n = i), \text{ where } i, j \in E.
\]

The transition matrix of the Markov Chain has its \((i,j)^{th}\) element as \( p_{ij} \). Thus,

\[
[P]_{ij} = p_{ij}
\]

The matrix dimension of \( P \) is finite or infinite depending upon cardinality of \( E \).

The random variable \( X_0 \) denotes the start of the Markov Chain and represents the initial state. The corresponding distribution \( \nu_0(i_0) = P(X_0 = i_0) \) is called the initial distribution.

**Lemma 1** Consider a Markov chain with \( r > 1 \) states having an \( r \times r \) transition matrix \( P = [p_{ij}] \) all of whose entries are positive, i.e. \( p_{ij} > 0 \) for all \( i, j \), with the minimum entry being \( d \). Let \( u \) be a \( r \times 1 \) vector with non-negative entries, the largest of which is \( M_0 \) and the smallest of which is \( m_0 \). Also, let \( M_1 \) and \( m_1 \) be the maximum and minimum elements in \( Pu \). Then,

\[
M_1 \leq dm_0 + (1 - d)M_0, \quad \text{and} \quad m_1 \geq dM_0 + (1 - d)m_0,
\]

and hence that

\[
M_1 - m_1 \leq (1 - 2d)(M_0 - m_0).
\]
Proof Each entry in the vector $Pu$ is a weighted average of the entries in $u$. The largest weighted average that could be obtained in the present case would occur if all but one of the entries of $u$ have value $M_0$ and one entry has value $m_0$, and this one small entry is weighted by the smallest possible weight, namely $d$. In this case, the weighted average would equal
\[ dm_0 + (1-d)M_0. \]
Similarly, the smallest possible weighted average equals:
\[ dM_0 + (1-d)m_0. \]
Thus,
\[ M_1 - m_1 \leq (dm_0 + (1-d)M_0) - (dM_0 + (1-d)m_0) = (1-2d)(M_0-m_0). \]

**Lemma 2** In the setup of Lemma 1, define $M_n$ and $m_n$ as the maximum and minimum elements of $P^n u$. Then $\lim_{n \to \infty} M_n$ and $\lim_{n \to \infty} m_n$ both exist and are in fact equal.

**Proof** The vector $P^n u$ is obtained from the vector $P^{n-1} u$ by multiplying on the left by the matrix $P$. Hence each component of $P^n u$ is an average of the components of $P^{n-1} u$. Thus,
\[ M_0 \geq M_1 \geq M_2 \cdots \]
and
\[ m_0 \leq m_1 \leq m_2 \cdots. \]
Each of the above sequences is monotone and bounded:
\[ m_0 \leq m_n \leq M_n \leq M_0. \]
Hence each of these sequences will have a limit as $n \to \infty$.

Let $\lim_{n \to \infty} M_n = M$ and $\lim_{n \to \infty} m_n = m$. We know that $m \leq M$. We shall prove that $M - m = 0$, which will be the case if $M_n - m_n$ tends to 0. Recall that $d$ is the smallest element of $P$ and, since all entries of $P$ are strictly positive, we have $d > 0$. By our lemma
\[ M_n - m_n \leq (1-2d)(M_{n-1} - m_{n-1}), \]
from which it follows that
\[ M_n - m_n \leq (1-2d)^n(M_0 - m_0). \]
Since $r \geq 2$, we must have $d \leq 1/2$, so $0 \leq 1 - 2d < 1$, so $M_n - m_n \to 0$ as $n \to \infty$. Since every component of $P^n u$ lies between $m_n$ and $M_n$, each component must approach the same number $M = m$. \[ \square \]

**Theorem 3** If $P = [p_{ij}]$ is an $r \times r$ transition matrix all of whose entries are positive, then
\[ \lim_{n \to \infty} P^n = W \]
where $W$ is some $r \times r$ matrix with identical row vectors. Also, each element of $W$ is positive.
**Proof** Lemma 2 implies that if we denote the common limit $M = m = \alpha$ then:

$$\lim_{n \to \infty} P^n u = \alpha 1.$$

In particular, suppose we choose $u = e_j$ where $e_j$ is the $r \times 1$ vector with its $j$-th component equal to 1 and all other components equaling 0. Then we obtain an $\alpha_j$ such that $\lim_{n \to \infty} P^n e_j = \alpha_j 1$. Repeating this for each $j = 1, \ldots, r$ we obtain:

$$\lim_{n \to \infty} P^n [e_1 : \ldots : e_r] = [\alpha_1 1 : \ldots : \alpha_r 1]$$

which implies that $\lim_{n \to \infty} P^n = W$ with each row of $W$ being the vector $\alpha^T = (\alpha_1, \ldots, \alpha_r)$.

It remains to show that all entries in $W$ are strictly positive. Note that $P e_j$ is the $j$-th column of $P$, and this column has all entries strictly positive. The minimum component of the vector $P u$ was defined to be $m_1$, hence $m_1 > 0$. Since $m_1 \leq m$, we have $m > 0$. Note finally that this value of $m$ is just the $j$-th component of $\alpha$, so all components of $\alpha$ are strictly positive. □

**Theorem 4** Suppose we relax the condition for all entries of $P$ being positive to the condition that there exists an integer $N$ such that all entries in $P^N$ are positive. Then, the result in Theorem 3 will still hold for such a $P$.

**Proof** We can directly apply Theorem 3 to the strictly positive matrix $P^N$ and obtain $M_{nN} - m_{nN} \to 0$ as $n \to \infty$, where $M_{nN}$ and $m_{nN}$ are defined analogously as the maximum and minimum elements of $P^{Nn} u$. However, by our earlier inequality the difference $M_n - m_n$ can never increase. Hence, if we know that the differences obtained by looking at every $N$-th time tend to 0, then the entire sequence must also tend to 0. □

To interpret the above result in terms of the stationary distribution of a finite Markov Chain, note that the vector $\alpha$ in Theorem 3 satisfies $\sum_{i=1}^r \alpha_i = 1$ since:

$$1 = \lim_{n \to \infty} P^n 1 = W 1$$

and each element of $W 1$ is $\sum_{i=1}^r \alpha_i$. Finally, observe that $\alpha$ (i.e. each row of $W$) is precisely the stationary distribution of $P$ by noting that if $\pi$ denotes a stationary distribution of $P$, then $\pi$ must coincide with $\alpha$:

$$\pi^T = \pi^T P = \pi^T P^n = \pi^T \lim_{n \to \infty} P^n = \pi^T W = \pi^T [\alpha_1 1 : \ldots : \alpha_r 1] = \alpha^T,$$

where we used the fact $\pi^T 1 = 1$ in the last equality. This also shows that such a chain must have a unique stationary distribution.