Discrete-State Markov Chains

(A) Basic Concepts, Transition Probabilities
(B) Classification of States and Chains
(C) Invariant Distributions and Large-Time Behavior
(D) Finite-State Markov Chains

References


(A) Basic Concepts, Transition Probabilities

Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space. Let \(S\) be a finite or countably infinite set. We will specify it to be

\[ S = \{1, 2, \ldots, N\} \text{ for some } N \in \mathbb{N} \]

or

\[ S = \mathbb{N} = \{1, 2, \ldots\}. \]

We consider a discrete-time stochastic process

\[ X_0, X_1, \ldots, X_n, \ldots\]

where each \(X_n : \Omega \to S\) is a random variable describing the state of an underlying stochastic system at time \(n\).
We call $S$ the state space, and each $i \in S$ a state. It is a discrete space, meaning that it is finite or co-countably infinite.

Two general questions:

1. Determine the joint distributions
   \[
   P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)
   \]
   for any $n \geq 0$ and $i_0, i_1, \ldots, i_n \in S$. Note that
   \[
   P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)
   = P(X_0 = i_0) \cdot P(X_1 = i_1 | X_0 = i_0) \cdot P(X_2 = i_2 | X_0 = i_0, X_1 = i_1) \cdot \ldots \cdot P(X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}).
   \]

2. Determine the large-time behavior of the distribution for $X_n$:
   \[
   P(X_n = j) \quad j \in S, \quad n \gg 1.
   \]

Definition (1) A stochastic process $\{X_n\}_{n=0}^{\infty}$ is a Markov process, if
   \[
   P(X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n)
   \]
   for any $n \geq 0$ and any $i_0, i_1, \ldots, i_{n+1} \in S$.

Definition (2) A stochastic process $\{X_n\}_{n=0}^{\infty}$ is time-homogeneous, if
   \[
   P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i)
   \]
   for any $n \geq 0$ and any $i, j \in S$. 
Assumption

Unless otherwise stated, we assume always that the discrete-time stochastic process \( \{X_n\}_{n=0}^{\infty} \) is a time-homogeneous Markov chain with a discrete state space \( S = \{1, \ldots, N\} \) or \( S = \mathbb{N} \).

The following two identities are very useful:

1. \[
P(X_{n+1}=i_{n+1}, \ldots, X_{n+k}=i_{n+k} \mid X_n=i_n, \ldots, X_1=i_1) = P(X_{n+1}=i_{n+1}, \ldots, X_{n+k}=i_{n+k} \mid n_{k+1}=i_{k+1})
\]
   where \( 0 < n_1 < \ldots < n_k < n_{k+1} < \ldots < n_m \), and \( i_1 \ldots i_m \in S \).

2. \[
P(X_{n+1}\in B_1, \ldots, X_{n+k}\in B_k \mid X_n=i_n, \ldots, X_1=i_1) = P(X_{n+1}\in B_1, \ldots, X_{n+k}\in B_k \mid X_n=i_n)
\]
   \[
   = P(X_{n+1}\in B_1, \ldots, X_{n+k}\in B_k \mid X_0=i_0)
\]
   where \( n, k \) are positive integers, \( A_1 \ldots A_m \) and \( B_1 \ldots B_k \) are all subsets of \( S \), and \( i_0 \in S \).

(They can be \( S \) or \( \mathbb{N} \) or other subsets.)

We prove these at the end of this set of notes.

With the Markovian property, we can rewrite the joint distribution

\[
P(X_0=i_0, X_1=i_1, \ldots, X_n=i_n) = P(X_0=i_0) P(X_1=i_1 \mid X_0=i_0) \cdots P(X_n=i_n \mid X_0=i_0 \cdots X_{n-1}=i_{n-1})
\]

\[
= P(X_0=i_0) P(X_1=i_1 \mid X_0=i_0) \cdots P(X_n=i_n \mid X_0=i_0 \cdots X_{n-1}=i_{n-1})
\]
Definition Let $i, j \in S$. The transition probability is
\[ p(i, j) = P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i) \]
for any $n \geq 0$.

The joint distribution can finally be written as
\[ P(X_0 = i_0, \ldots, X_n = i_n) = P(X_0 = i_0) P(X_1 = i_1 \mid X_0 = i_0) \cdots P(X_n = i_n \mid X_{n-1} = i_{n-1}) = P(X_0 = i_0) P(i_0, i_1) \cdots P(i_{n-1}, i_n). \]
Thus, the transition probabilities $p(i, j)$ ($i, j \in S$), together with the distribution of the initial random variable $X_0$, $P(X_0 = i_0)$ ($i_0 \in S$), determine completely the joint distributions.

Definition The transition matrix is $P = (p(i, j))_{i, j \in S}$.
Here, $i, j \in S$ means that $i, j = 1, \ldots, N$ for $S = \{S_1, \ldots, S_N\}$ and $i_j = 1, 2, \ldots, \text{ for } S = N$.

The transition matrix $P$ is a stochastic matrix, i.e., it satisfies the following:

1. All entries $p(i, j) \geq 0$ for $i, j \in S$;
2. $\sum_{j \in S} p(i, j) = 1$ (each row sums to 1) for $i \in S$.

Note: If $|S| < \infty$, i.e., $S$ is finite, the transition matrix is the usual, square, stochastic matrix of finitely many rows and columns. Here, we allow infinite matrices.
(i) For a (finite) stochastic matrix $P = (p(i,j))_{i,j=1}^{n}$ we observe that 1 is an eigenvalue of $P$ and $[1; 1]$ is a corresponding eigenvector. This is true for an infinite stochastic matrix.

(ii) It is easy to check that the product of two stochastic matrices is still a stochastic matrix.

**Definition.** Let $n \geq 0$ be an integer and $i,j \in S$. The $n$-step transition probability is 

$$p_n(i,j) = P(X_n = j \mid X_0 = i).$$

For $n = 0$, $p_0(i,j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \end{cases} = \delta_{ij}$.

For $n = 1$, $p_1(i,j) = \delta_{ij}$. The $n$-step transition matrix is $P_n = (p_n(i,j))_{i,j \in S}$.

A probability vector $\alpha = (\alpha(i))_{i \in S}$ is a vector of nonnegative components and all summed up to 1. For each $n \geq 0$, we define the distribution (vector) $\pi_n$ for $X_n$ by 

$$\pi_n(i) = P(X_n = i), \quad i \in S.$$ 

Call $\alpha$ the initial distribution of the chain $\{X_n\}_{n=0}^{\infty}$.

**Proposition.**

1. For any integers $n \geq 0$ and any $i, j \in S$, 

$$p_n(i,j) = P(X_{n+1} = j \mid X_n = i).$$
(2) The Chapman-Kolmogorov equation
\[ P_{n+m}(i,j) = \sum_{k \in S} P_n(i,k) P_m(k,j) \]
\[ \forall i,j \in S, \forall n,m \geq 0: \text{integers}. \]
In particular, this means
\[ P_n P_m = P_{n+m}. \quad (\text{Matrix}) \]
Also,
\[ P^n = P^n \]
i.e., the \[ P_{n}(i,j) \] is the \( (i,j) \)-entry of \( P^n \). This also implies that \( P_n \) is a stochastic matrix.

(3) For any integer \( n \geq 0 \),
\[ \pi_n = \pi_0 P^n = \pi_0 P^n. \]
which is
\[ P(X_n = j) = \sum_{i \in S} \pi_0(i) P_n(i,j) \quad \forall j \in S. \]
Also,
\[ \pi_{n+1} = \pi_n P. \]

We end this part by providing a brief proof of the identities in page 3.

To prove (1), we first show
\[ P(X_{n+k+m} = j \mid X_n = i_k, \ldots, X_1 = i_1) = P(X_{n+k+m} = j \mid X_n = i_k) \]
where \( n \geq 1, n_1 < n_2 < \cdots < n_k \), and all \( i_1, \ldots, i_k \) \( \in S \).

By induction on \( m \geq 1 \).

\( m=1 \). Let \( I \) denote the set of "missing indices": \( I \in \{0, 1, \ldots, n_k\} \) such that \( \{n_1, n_2, \ldots, n_k\} \).

If \( I = \emptyset \), then (X) is true for \( m=1 \) by the definition of a Markov chain. Assume \( I \neq \emptyset \). Then, using the elementary but useful identity
\[ P(A \cap B | C) = P(A | B \cap C) P(B | C), \]

and denote \( I = \{ i_1, \ldots, i_S \} \), we have
\[ P(X_{n_k+1} = j | X_{n_k} = i_k, \ldots, X_{n_l} = i_l) \]
\[ = \sum_{j, \ldots, j_S \in \mathcal{S}} P(X_{n_k+1} = j, X_{t_1} = i_k, \ldots, X_{t_S} = j_S | X_{n_k} = i_k, \ldots, X_{n_l} = i_l) \]
\[ = \sum_{j, \ldots, j_S \in \mathcal{S}} P(X_{n_k+1} = j | X_{n_k} = i_k, \ldots, X_{n_l} = i_l) \cdot P(X_{t_1} = i_k, \ldots, X_{t_S} = j_S | X_{n_k} = i_k, \ldots, X_{n_l} = i_l) \]
\[ = \sum_{j, \ldots, j_S \in \mathcal{S}} P(X_{n_k+1} = j | X_{n_k} = i_k) \cdot \sum_{j, \ldots, j_S \in \mathcal{S}} P(X_{t_1} = i_k, \ldots, X_{t_S} = j_S | X_{n_k} = i_k, \ldots, X_{n_l} = i_l) \]
\[ = P(X_{n_k+1} = j | X_{n_k} = i_k) \sum_{j, \ldots, j_S \in \mathcal{S}} P(X_{t_1} = i_k, \ldots, X_{t_S} = j_S | X_{n_k} = i_k, \ldots, X_{n_l} = i_l) \]
\[ = P(X_{n_k+1} = j | X_{n_k} = i_k). \]

So, (x) is true for \( m = 1 \).

Suppose (x) is true for any \( m \geq 1 \). Then,
\[ P(X_{n_k+m+1} = j | X_{n_k+m} = i_k, \ldots, X_{n_l} = i_l) \]
\[ = \sum_{j, \ldots, j_S \in \mathcal{S}} P(X_{n_k+m+1} = j, X_{n_k+m} = i_k, \ldots, X_{n_l} = i_l) \]
\[ = \sum_{j, \ldots, j_S \in \mathcal{S}} P(X_{n_k+m+1} = j | X_{n_k+m} = i_k, \ldots, X_{n_l} = i_l) \cdot P(X_{n_k+m} = i_k, \ldots, X_{n_l} = i_l) \]
\[ = \sum_{j, \ldots, j_S \in \mathcal{S}} P(X_{n_k+m+1} = j | X_{n_k+m} = i_k) \cdot P(X_{n_k+m} = i_k, \ldots, X_{n_l} = i_l) \]
\[ = \sum_{j, \ldots, j_S \in \mathcal{S}} P(X_{n_k+m+1} = j | X_{n_k+m} = i_k) \cdot P(Y_{n_k} = i_k). \]
Thus, (x) is true for m+1. Hence it is true for all m≥1.

Now, we prove the first identity on page 3.

\[ P(X_{m} \cap \cdots \cap X_{k+1} | X_{k} = c_{k}, \cdots, X_{1} = c_{1}) \]

\[ = \frac{P(X_{m} \cap \cdots \cap X_{k+1} = c_{k+1}, X_{k} = c_{k}, \cdots, X_{1} = c_{1})}{P(X_{k} = c_{k}, \cdots, X_{1} = c_{1})} \]

\[ = \frac{P(X_{m} = c_{m} | X_{m-1} = c_{m-1}) \cdots P(X_{k+1} = c_{k+1} | X_{k} = c_{k})}{P(X_{k} = c_{k})} \]

\[ = P(X_{m} = c_{m}, \cdots, X_{k+1} = c_{k+1} | P(X_{k} = c_{k})). \]

This completes the proof of the 1st identity on page 3.

For the second identity on page 3, we first show that

\[ P(X_{n} = c_{n}, \cdots, X_{n+1} = c_{n+1} | X_{0} = A_{1}, X_{1} = A_{2}, \cdots, X_{n-1} = A_{n-1}, X_{n} = c_{n}) \]

\[ = P(X_{n} = c_{n}, \cdots, X_{n+1} = c_{n+1} | X_{n} = c_{n}) \]

where \( n \leq n_{1} \leq \cdots \leq n_{k}, c_{n}, c_{n}, \cdots, c_{n+1} \in S \), and all \( A_{1}, \cdots, A_{m} \) are subsets of \( S \). This is proved by using the following property: If \( D_{1} \) are disjoint, and \( P(C | D_{i}) = p \) is independent of \( c \), then

\[ P(C | \cup D_{i}) = p. \]
We now have
\[ P(X_{n+1} \in B_1, \ldots, X_{n+k} \in B_k \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = i_0) \]
\[ = \sum_{i_1 \in B_1, \ldots, i_k \in B_k} P(X_{n+1} = i_1, \ldots, X_{n+k} = i_k \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = i_0) \]
\[ = \sum_{i_1 \in B_1, \ldots, i_k \in B_k} P(X_{n+1} = i_1, \ldots, X_{n+k} = i_k \mid X_n = i_0) \]
\[ = P(X_{n+1} \in B_1, \ldots, X_{n+k} \in B_k \mid X_n = i_0). \]

The proof is complete.

(13) Classification of States and Chains

Given a state \(i \in S\). The chain may visit this state infinitely many times, or otherwise only finitely many times, probabilistically. These two distinguished situations can be described using different notions.

**Definition.** A state \(i \in S\) is **recurrent** if
\[ P(X_n = i \text{ for some } n \geq 1 \mid X_0 = i) = 1, \]
A state \(i \in S\) is **transient** if
\[ P(X_n = i \text{ for some } n \geq 1 \mid X_0 = i) < 1. \]

So, any state is either recurrent or transient, but not both.
A useful concept is the hitting time.

**Definition.** The hitting time for a state $i \in S$ is

$$T_i = \min \{ n \geq 1 : X_n = i \}$$

if the minimum exists. Otherwise, we define $T_i = \infty$.

Clearly, the event $\{X_n = i \text{ for some } n \geq 1\} = \{T_i < \infty\}$.

Hence, we conclude:

- $i \in S$ is recurrent $\iff P(T_i < \infty | X_0 = i) = 1$.
- $i \in S$ is transient $\iff P(T_i < \infty | X_0 = i) < 1$.

The following result connects the hitting times with transition probabilities.

**Proposition.** We have for any $n \geq 1$ and $i, j \in S$ that

$$P_n(i, j) = \sum_{k=1}^{n} P(T_j = k \mid X_0 = i) \cdot P_{n-k}(j, j).$$

See [HPS] (Section 1.4.1).

The first passage times are also useful quantities.

**Definition.** Let $i, j \in S$. Denote for any $n \geq 1$

$$F_n(i, j) = P(X_{n+1} = j, \ldots, X_{n+k} = j, X_n = i \mid X_0 = i)$$

as probability that the first visit to state $j$, starting from state $i$, takes place at the $n$th step.
Denote 
\[ f_{i,j} = \sum_{n=1}^{\infty} f_{n}(i,j) \]
= probability that the chain ever visits state \( j \) starting from state \( i \).

We have
\[ f_{n}(i,j) = \Pr(T_j = n \mid X_0 = i), \]
\[ f(i,j) = \Pr(T_j < \infty \mid X_0 = i), \]
= probability of ever returning to \( j \) starting from \( i \),
\[ 1 - f(i,j) = \Pr(T_j = \infty \mid X_0 = i) \]
= probability of never returning to \( j \) starting from \( i \).

Therefore
\[ i \in S \text{ is recurrent } \iff f(i,i) = 1, \]
\[ i \in S \text{ is transient } \iff f(i,i) < 1. \]

A more intuitive quantity is the number of visits to a state \( i \).

**Definition.** Let \( i \in S \). Define
\[ N(i) = \sum_{n=1}^{\infty} 1_i(X_n) \]
= number of times the chain visits state \( i \) (after the initial time).
where \( 1_i = 1_{\{i\}} \) is the indicator function of \( \{i\} \).

We have clearly that
\[ \{N(i) > 1\} = \{T_j < \infty\} \quad \forall j \in S.\]
In particular,

\[ P(M(i) \geq 1 \mid X_0 = i) = P(T_j < \infty \mid X_0 = i) \quad \forall i, j \in S. \]

Hence,

- \( i \in S \) is recurrent \( \iff \) \( P(M(i) \geq 1 \mid X_0 = i) = 1 \),
- \( i \in S \) is transient \( \iff \) \( P(M(i) \geq 1 \mid X_0 = i) < 1 \).

Note also that for any \( i \in S \),

\[ \{N(i) = \infty\} \equiv \{X_n = i \text{ for infinitely many } n\} \]

The expected number of visits to a state \( j \in S \), \( X_0 = i \), starting from state \( i \), is

\[ E(\mathbf{1}_j(X_n) \mid X_0 = i) = E(X_n = j \mid X_0 = i) = p_{n(i,j)} \]

for any \( n \geq 0 \).

Thus,

\[ E(N(j) \mid X_0 = i) = \sum_{n=1}^{\infty} E(\mathbf{1}_j(X_n) \mid X_0 = i) = \sum_{n=1}^{\infty} p_{n(i,j)} \]

The following result characterizes the fundamental difference between a recurrent state and a transient state.

**Theorem**

1. A state \( i \in S \) is recurrent \( \iff \) \( \sum_{n=1}^{\infty} p_{n(i,i)} = \infty \)

\( \iff \)

\[ P(N(i) = \infty \mid X_0 = i) = 1. \]

In this case, for any \( j \in S \),
\[ P(N(i) = \infty \mid X_0 = k) = P(T_i < \infty \mid X_0 = k) = f(k, i) \]

If \( f(k, i) = 0 \) then all \( P_n(k, i) = 0 \) \( (n = 1, 2, \ldots) \)
If \( f(k, i) > 0 \) then \( \sum_{n=1}^{\infty} P_n(k, i) = \infty \).

(2) A state \( i \in S \) is transient

\[ \iff \sum_{n=1}^{\infty} p_n(i, i) < \infty \]
\[ \iff P(N(i) = \infty \mid X_0 = i) = 0 \]

In this case, for any \( k \in S \),

\[ P(N(i) = \infty \mid X_0 = k) = 0, \]

\[ \sum_{n=1}^{\infty} p_n(k, i) = \frac{f(k, i)}{1 - e^{-f(k, i)}} \]

which is finite.

Note that the theorem particularly indicates that

\[ P(N(i) = \infty \mid X_0 = i) = \begin{cases} 1 & \text{if } i \text{ is recurrent,} \\ 0 & \text{if } i \text{ is transient}. \end{cases} \]

There are at least two different proofs of the theorem. One of them is based on the generating functions

\[ P_{ij}(s) = \sum_{n=0}^{\infty} s^n p_n(i, j) \]
\[ F_{ij}(s) = \sum_{n=0}^{\infty} s^n P_n(i, j) \]

where \( f_0(i, j) = 0, \ i, j \in S \), and \( s \in \mathbb{C} \). They satisfy

\[ P_{ij}(s) = P_{ij} + F_{ij}(s) P_{jj}(s), \ \forall i, j \in S, \forall s \in \mathbb{C}. \]
This, together with Abel's Theorem on the convergence of power series at the end points of the interval of convergence interval, will leads the desired result. See [GS].

A different proof presented in [HP5] is more of a probabilistic proof, using some relations between the hitting times and total number of visiting times, e.g. the formula in Proposition on page 10, and also the expected number of visits.

**Corollary** If \( i \in S \) is a transient state, then
\[
\lim_{n \to \infty} p_n(k,i) = 0 \quad \forall k \in S.
\]

**Corollary** If \( S \) is finite, there there exists at least one recurrent state.

**Definition.** The mean recurrence time \( \mu_i \) of a state \( i \in S \) is
\[
\mu_i = \mathbb{E}(T_i \mid X_0 = i) = \begin{cases} 
\sum_{n=1}^{\infty} n f_{pi}(i) & \text{if } i \text{ is recurrent,} \\
\infty & \text{if } i \text{ is transient.}
\end{cases}
\]

Note that \( \mu_i \) can be infinite even if it is a recurrent state. Note also that, if \( i \) is transient, then
\[
P(T_i = \infty \mid X_0 = i) = 1 - P(T_i < \infty \mid X_0 = i) > 0.
\]

Hence, \( \mathbb{E}(T_i \mid X_0 = i) = \infty \).
Definition A recurrent state \( i \in S \) is a null recurrent state if \( \mu_i = \infty \), and a positive recurrent state if \( \mu_i < \infty \).

Theorem Let \( i \in S \) be a recurrent state.

1. If \( i \) is null recurrent, then
   \[ \lim_{n \to \infty} \mathbb{P}_n(i,i) = 0 \]
   In this case, \( \lim_{n \to \infty} \mathbb{P}_n(j,i) = 0 \quad \forall j \in S \).

2. If \( i \) is positive recurrent, then
   \[ \lim_{n \to \infty} \frac{N_n(i)}{n} = \mathbb{P}(T_i < \infty) \quad \text{with probability 1}, \]
   where \( N_n(i) = \sum_{k=1}^{n} 1(X_k = i) \), and
   \[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}_k(j,i) = \mathbb{P}(T_i < \infty | X_0 = j) \quad \forall j \in S. \]
   (When \( j = i \), the right-hand side is just \( \frac{1}{\mu_i} \).

Note that Part (2) above improves the result in a theorem on page 12.

Corollary If \( S \) is finite, then any recurrent state must be positive recurrent.

We now consider states that communicate with each other.

Definition A state \( i \in S \) leads to a state \( j \in S \) if
\[ \mathbb{P}(T_j < \infty | X_0 = i) > 0. \]
We write \( i \rightarrow j \).

Proposition \( i \rightarrow j \iff \exists n \geq 1 \) such that \( \mathbb{P}_n(i,j) > 0 \).

Theorem (1) If \( i \in S \) is recurrent and \( i \rightarrow j \in S \), then \( j \) is also recurrent. Moreover, \( \mathbb{P}(T_j < \infty | X_0 = i) = \mathbb{P}(T_i < \infty | X_0 = j) = 1 \), and \( j \rightarrow i \). (resp. positive recurrent)

(2) If \( i \in S \) is null recurrent (positive recurrent) and \( i \rightarrow j \), then \( j \) is also null recurrent (resp. positive recurrent), and \( j \rightarrow i \).
See [HPS] for a proof of this theorem.

By the previous theorem, we immediately have: \( \forall i,j \in S \) (possible)

If \( i \leftrightarrow j \), then:
1. \( i \) is recurrent \( \iff \) \( j \) is.
2. \( i \) is transient \( \iff \) \( j \) is.

Moreover, \( i \) is null-rec. (possible)

Definition Two states \( i, j \in S \) communicate with each other, if \( i \rightarrow j \) and \( j \rightarrow i \), written \( i \leftrightarrow j \).

Proposition The relation \( \leftrightarrow \) is an equivalent relation, i.e., it satisfies the following:

1. \( i \leftrightarrow i \) \( \forall i \in S \),
2. \( i \leftrightarrow j \Rightarrow j \leftrightarrow i \) \( \forall i, j \in S \),
3. \( i \leftrightarrow j, j \leftrightarrow k \Rightarrow i \leftrightarrow k \) \( \forall i, j, k \in S \).

Definition The equivalent relation \( \leftrightarrow \) classifies \( S \) into disjoint equivalent classes, called communication classes.

In each communication class, any two states communicate with each other. States in different communication classes will not communicate with each other. But, one way "communication" is still possible. For instance, if \( i \in S \) is transient and \( j \in S \) is recurrent, then it is possible that \( i \rightarrow j \). But, \( j \rightarrow i \) is impossible.

Moreover, a non-empty subset of states \( C \subseteq S \) is closed if
\[
P(i,j) = 0 \ \forall i \in C \ \forall j \in S \setminus C.
\]
Proposition If $S$ is finite, then any recurrent state is positive recurrent.

Proof. Use Part (b) of Theorem 1 on page 15. Consider a recurrent communication class.

We now describe another characteristic property of a state, other than recurrences and transience.

Definition The period $d(i)$ of a state $i \in S$ is
\[ d(i) = \text{g.c.d.} \{ n \geq 0 : P^n(i, i) > 0 \} \]
where g.c.d. = greatest common divisor.

If $d(i) > 1$, call $i$ a periodic state.

If $d(i) = 1$, call $i$ an aperiodic state.

Clearly, $d(i) = \min \{ n \geq 1 : P^n(i, i) > 0 \}$.

Hence, \[ P(k+i) > 0 \iff d(k) = 1 \]

Definition A state $i \in S$ is ergodic if it is positively recurrent and aperiodic.

Theorem If $i \in S$ is aperiodic, then $\lim_{n \to \infty} P^n(i, i) = 1$, and
\[ \lim_{n \to \infty} P^n(j, i) = \frac{f(j, i)}{d(i)} \forall j \in S. \]

In particular, $\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} P_k(i, i) = \frac{\pi_i(i)}{d(i)}$ $\forall j \in S$.

Proposition Let $i, j \in S$. If $i \rightarrow j$ then $d(i) = d(j)$.

Therefore, we can define the period of a communication class of states as that of any state in the class.

Assume $S$ is finite.

Theorem Let $i, j \in S$ be such that $i \rightarrow j$. Then
\begin{align*}
(1) d(i) &= d(j) \\
(2) i \text{ is transient} &\iff j \text{ is transient} \\
(3) i \text{ is recurrent} &\iff j \text{ is recurrent} \\
(4) i \text{ is null recurrent} &\iff j \text{ is null recurrent}
\end{align*}

Definition A non-empty subset of states $C \subseteq S$ is closed if
\[ P(c, j) = 0 \quad \forall c \in C, \forall j \in S \setminus C. \]

Proposition Let $\emptyset \neq C \subseteq S$. If $C$ is closed, then $\forall i \not\in C$ there exists $n \geq 1$ such that
\[ P_n(c, j) = 0 \quad \forall c \in C, \forall j \in S \setminus C, \forall n \geq 1. \]
**Definition** A non-empty subset of states, $C \subseteq S$, is irreducible, if $i \rightarrow j$ for any $i, j \in C$.

If $S$ is itself irreducible, then the chain, $\{X_n\}_{n=0}^{\infty}$, is irreducible.

Clearly, any communication class of states is irreducible. If there exists only one communication class, then the chain is irreducible.

**Decomposition Theorem** The state space $S$ is the union of disjoint subsets of states,

$$S = T \cup \left( \bigcup_{i=1}^{n} R_i \right),$$

where $T$ is the set of all transient states, and each $R_i$ is a recurrent communication class. There are finitely many or countably infinitely many $R_i$'s.

**Note.** Among those recurrent classes, some of them can be null, i.e., states in such a class are null recurrent. Some of them can be positive.

For a finite $S$, there exists at least one $R_i$.

In general, $T = \emptyset$ is possible. If $T \neq \emptyset$, $T$ itself may not be a communication class. But, we have

$$T = \bigcup_{j} T_j,$$

with each $T_j$ a communication class of transient states, and $T \cap T_j = \emptyset$ if $i \neq j$. 
If the chain starts with a state in some recurrent class $R_j$, then the chain will stay in $R_j$. This means each recurrent class defines itself a (sub)state space.

If the chain starts at a transient state then the chain may never leave the set of all transient states $T$, or it may eventually get into a recurrent class and stays there forever. If $S$ is finite, then the first situation will not occur.

(C) Invariant Distributions and Large-Time Behavior

**Definition** A probability vector $\pi = (\pi(i)), i \in S$ is an invariant (or stationary) distribution (for the transition matrix $P$) if $\pi = \pi P$, i.e., $\forall i \in S$

$$\pi(j) = \sum_{i \in S} \pi(i) p(i,j)$$

**Remarks**

(1) Motivation. If for some initial distribution $\pi_0$, we have

$$\lim_{n \to \infty} \pi_n = \lim_{n \to \infty} \pi_0 P^n = \pi$$

for some probability vector $\pi$, where $\pi_n = \pi_0 P^n$ is the distribution of $X_n$, then

$$\pi = \lim_{n \to \infty} \pi_{n+1} = \lim_{n \to \infty} \pi_0 P^{n+1} = \lim_{n \to \infty} (\pi_0 P^n) P = \pi P$$

Hence the limit $\pi$ is invariant.
(2) \( \pi = \pi P \) means that \( \pi \) is a left eigenvector of \( P \) corresponding to the eigenvalue 1 of \( P \). Recall that \( \left[ \begin{array}{c} 1 \\ \end{array} \right] \) is a (right) eigenvector of \( P \) corresponding to the eigenvalue 1.

Proposition. Let \( \pi \) be an invariant distribution. Then
\[
\lim_{n \to \infty} \pi_0 P^n = \pi
\]
for any probability vector \( \pi_0 \), if and only if
\[
\lim_{n \to \infty} P^n = \left\lfloor \begin{array}{c} \pi \\ \end{array} \right\rfloor
\]
i.e.,
\[
\lim_{n \to \infty} P^n(i,j) = \pi(j) \quad \forall i, j \in S.
\]
In this case, the invariant distribution is unique. Moreover,
\[
\lim_{n \to \infty} P(X_n = i) = \pi(i) \quad \forall i \in S.
\]

Proof. Suppose \( \pi_0 P^n \to \pi \) as \( n \to \infty \) for any probability vector \( \pi_0 \). Let \( i, j \in S \). Choose \( \pi_0 = (\pi_0(k))_{k \in S} \) by \( \pi_0(k) = \delta_{k,i} \). Then, \( \pi_0 P^n \to \pi \) implies
\[
\sum_{k \in S} \pi_0(k) P^n(i,k,j) \to \pi(j) \quad \forall j \in S.
\]
But \( \sum_{k \in S} \pi_0(k) P^n(i,k,j) = P^n(i,j) \). Hence, \( P^n(i,j) \to \pi(i) \) as \( n \to \infty \) for any \( j \).

Conversely, assume that \( P^n(i,j) \to \pi(i) \) \( \forall i, j \in S \).

If \( \pi_0 = (\pi_0(k))_{k \in S} \) is any probability vector, then
\[
(\pi_0 P^n)(i) = \sum_{k \in S} \pi_0(k) P^n(i,k) \to \sum_{k \in S} \pi_0(k) \pi(i) = \pi(i) \sum_{k \in S} \pi_0(k) = \pi(i) \quad \forall i \in S.
\]
[Note that the convergence here can be justified.]
If \( \hat{\pi} \) is also an invariant distribution, then
\[
\hat{\pi} = \hat{\pi} P = \hat{\pi} P^2 = \ldots = \hat{\pi} P^n \rightarrow \pi.
\]
Hence, \( \hat{\pi} = \pi \). Moreover,
\[
P (X_n = i) = \pi_n (i) = (\pi_0 P^n) (i) \rightarrow \pi(i) \quad \forall i \in S.
\]

**Theorem** Let \( \pi \) be an invariant distribution. If \( i \in S \) is a transition state or a null recurrent state, then \( \pi(i) = 0 \).

This follows from the fact that \( \lim_{n \to \infty} p_{ij} (n) = 0 \) and \( \pi(i) = (\pi P^n) (i) = \sum_{j \in S} \pi(j) p_{ij} (n) \to 0 \).

**Definition** A stochastic matrix \( P \) and a probability vector of order \( |S| \) (i.e., \( \pi \in \mathbb{R}^{|S|} \)) satisfy the detailed balance if
\[
\pi(i) p(i, j) = \pi(j) p(j, i) \quad \forall i \in S.
\]

**Theorem** If \( \pi = (\pi(i))_{i \in S} \) is a probability vector, and the transition matrix \( P \) and \( \pi \) satisfy the detailed balance, then \( \pi \) is an invariant distribution.

**Proof** We have for any \( j \in S \) that
\[
(\pi P)(j) = \sum_{i \in S} \pi(i) p(i, j) = \sum_{i \in S} \pi(j) p(j, i) = \pi(j) \sum_{i \in S} p(j, i) = \pi(j).
\]

**Theorem** An irreducible chain has an invariant distribution \( \pi \) \( \iff \) all the states are positively recurrent. In this case, \( \pi \) is the unique invariant distribution given by
\[
\pi(i) = \frac{1}{\xi(i)} > 0 \quad \forall i \in S.
\]

(In particular, \( \sum_{i \in S} \pi(i) = 1 \).
Theorem For an irreducible aperiodic chain, we have that

$$\lim_{n \to \infty} P^n(i, j) = \frac{1}{\pi_i} \quad \forall i, j \in S.$$  

If, in addition, the chain is positive recurrent, then the vector $\pi \in \mathbb{R}^{|S|}$, defined by $\pi(i) = \frac{1}{\pi_i} \quad \forall i \in S$, is the unique invariant distribution, and

$$\lim_{n \to \infty} P^n(i) = \pi(i) \quad \forall i \in S$$

where $P^n(i) = P(X_n = i)$.

Note that the convergence $P^n(i) \to \pi(i) = \frac{1}{\pi_i}$ is independent of an initial distribution $\pi_0$:

$$P^n(i) = P(X_n = i) = \sum_{j \in S} P(X_0 = j) P(X_{n-1} = i | X_0 = j)$$

$$= \sum_{j \in S} P(X_0 = j) P_{n-1}(j, i)$$

$$\to \sum_{j \in S} P(X_0 = j) \frac{1}{\pi_i}$$

$$= \pi(i) \sum_{j \in S} P(X_0 = j)$$

$$= \pi(i).$$

Theorem Suppose the chain is irreducible, recurrent, and periodic with period $d > 1$. Let $i, j \in S$. Then there exists an integer $r$ such that $0 < r < d$ such that

$$P^n(i, j) = 0 \quad \text{if } n \not\equiv r \pmod{d}$$

and

$$\lim_{n \to \infty} P^{md+r}(i, j') = d(i, j').$$
(1) Finite-State Markov Chains

Assume in this part $|S|<\infty$. In particular, $S=\{1, 2, \ldots, N\}$ for some integer $N \geq 1$.

**Theorem.** Let $P = (p_{ij})$ now be a stochastic matrix. Suppose:

1. $1$ is a simple eigenvalue of $P$ and all other eigenvalues of $P$ have absolute values less than $1$.
2. There exists a left eigenvector of $P$ corresponding to the eigenvalue $1$ that has nonnegative components.

Then $P$ has a unique invariant distribution $\pi \geq 0$ and for any initial distribution $\pi_0$,\[ \lim_{n \to \infty} \pi_0 P^n = \pi, \text{ i.e. } \lim_{n \to \infty} \pi_n = \pi, \]
where $\pi_n$ is the distribution of $X_n (n \geq 0)$.

**Proof.** We have the Jordan decomposition $P = Q^{-1} D Q$, where $D = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$, $\alpha = \begin{bmatrix} a_1 \\ \vdots \\ \alpha_n \end{bmatrix}$ (the row vectors),

$Q^T = \begin{bmatrix} I & b_1 & \cdots & b_N \end{bmatrix}$ ($I = \begin{bmatrix} 1 \\ \vdots \end{bmatrix}$ is an $N \times N$ identity matrix, $b_j$ is a column vector).

From $D Q^{-1} = Q^{-1} P$, we can get $a_i = a_1 P$

$D^n = \begin{bmatrix} 1 & \cdots & 0 \\ 0 & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$, $M$ consists of Jordan blocks. But, all eigenvalues of $P$ have absolute values $\leq 1$, except $1$, which is simple. So, $M^n \to 0$. Thus, $P^n = Q D^n Q^{-1} \to \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$

If necessary, normalize $a_1$ so that $a_1 > 0$. Thus, $\frac{a_1}{a_1} = 1$.

Thus $\pi := a_1$ is an invariant distribution. It is unique and $\pi_n \to \pi$ by a previous proposition.
Theorem. If $S$ is finite, and there exists $n \geq 1$ such that $p_n(i, j) > 0$ for all $i, j \in S$, then (1) and (2) in the above theorem holds true, and hence the conclusions of the above theorem follow.

Proof. This follows from the Perron-Frobenius Theorem. \[ \square \]

Note that, for a finite-state chain, there exists at least one recurrent state, and any recurrent state is positively recurrent. Therefore, if a finite-state chain is irreducible, all states are positively recurrent.

Theorem. If a finite-state chain is irreducible and aperiodic, then $\exists n \geq 1$ such that $p_n(i, j) > 0$ $\forall i, j \in S$. Hence, $\exists$ invariant distribution $\pi > 0$ determined by $\pi(i) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} p_k(i) = \pi$ $\forall i \in S$ and satisfies $\lim_{n \to \infty} \mathbb{P}(X_n = i) = \pi$ a initial distribution $\pi_0$.

Proof. Let $i, j \in S$. Since $P$ is irreducible, $\exists m = m(i, j)$ s.t. $p_m(i, j) > 0$. Since $P$ is aperiodic, $\exists M(i)$ for $i \in S$ such that $p_n(i, i) > 0$ if $n \geq M(i)$. Hence, if $n > M(i)$, then $p_{n+m}(i, j) = p_n(i, i)p_m(i, j) > 0$.

Let $M$ be the maximum value of $M(i) + m(i, j)$ over all $i, j \in S$. The maximum exists, since $S$ is finite.
Then $P^n(i,j) \to 0$ if $n \to \infty$, $\forall i,j \in S$. Then apply the previous theorem. □

**Theorem 1.** If the state $S$ is finite and $P$ is symmetric ($P(i,j) = P(j,i)$), $\forall i,j \in S$, and irreducible, then the invariant distribution is uniform: $\pi(i) = \frac{1}{|S|} \forall i \in S$.

**Brief discussion on the case of a reducible and/or periodic finite-state chain.**

Suppose the state $S$ is finite. If the chain is reducible, then there exist disjoint recurrent communication classes $R_1, \ldots, R_k$. Each $R_k$ is irreducible and (positively) recurrent. Thus, for each $R_k$, there exists an invariant distribution $\pi^{(k)}$ which is concentrated on $R_k$, i.e., $\pi^{(k)}(i) = 0$ if $i \notin R_k$. The eigenvalue $1$ of the transition matrix $P$ has $r$ linearly independent eigenvectors, i.e., no more than $r$ eigenvectors of the eigenvalue $1$ are linearly independent. Let $\pi^{(k)}$ be the eigenvalue $1$ of the transition matrix $P$ has $r$ linearly independent eigenvectors, i.e., no more than $r$ eigenvectors of the eigenvalue $1$ are linearly independent. Let $\pi^{(k)}$ be the eigenvalue $1$ of the transition matrix $P$ has $r$ linearly independent eigenvectors, i.e., no more than $r$ eigenvectors of the eigenvalue $1$ are linearly independent. Then $\pi^{(k)}(i) = 0$ if $i \notin R_k$.

If $i \in S$ is a transient state, then the chain starting at $i$ eventually ends up in a recurrent class, i.e.,

$$\lim_{n \to \infty} P^n(i,j) = 0 \quad \text{if } j \text{ is transient.}$$

Let $\alpha_x(i)$ be the probability that the chain starting at $i$ eventually ends up in $R_k$. Then

$$\lim_{n \to \infty} P^n(i,j) = \alpha_x(i) \pi^{(k)}(j) \quad \forall j \in R_k.$$

The limit $\pi_0 P^n$ exists but may depend on $\pi_0$. 
Suppose now the transition matrix $P$ for a finite-state chain $\mathcal{S}$ is irreducible but periodic with period $d > 1$. Then $P$ has exactly $d$ eigenvalues with absolute values equal to 1. These are $d$ roots of the equation $z^d = 1$. They and except 1, the others are complex but not real numbers. Each of such eigenvalues is simple.

In this case, there exists a unique invariant distribution $\pi$. For any initial distribution $\pi_0$,

$$ \lim_{n \to \infty} \frac{1}{d} \sum_{j=1}^{d} (P^{nd} \pi_0) = \pi. $$