1.1 Definition and characterization

In this section we give the definition of a Markov Chain and several related concepts.

A \textit{stochastic process} is an indexed collection of random variables $X = \{X(t) : t \in T\}$ or $X = \{X_t : t \in T\}$ with $T$ being the index set, all of them defined on the same probability space $(\Omega, \mathcal{F}, P)$. $t$ often represents time, in which case $X(t)$ or $X_t$ is the state at time $t$, which models the time variation of a particular quantity. If $T = \mathbb{R}$ or $T$ is an interval of $\mathbb{R}$, then $X$ is a continuous-time random process. If $T$ is a countable set, then $X$ is a discrete-time process. For example, if $T = \mathbb{N}$, then $X = \{X_0, X_1, X_2, \cdots\}$. For $T = \mathbb{N}$ or $T = \mathbb{Z}$ we may alternatively write $X = \{X_n\}_{n \geq 0}$ or $X = \{X_n\}_{n \in \mathbb{Z}}$, respectively.

Suppose $X_k \in \mathcal{X}$ for all $k$. The set $\mathcal{X}$ is called \textit{state space}.

\textbf{Definition 1.1. (Markov Chain)} A discrete-time stochastic process $X = \{X_0, X_1, X_2, \cdots\}$ with a countable state space $\mathcal{X}$ is a Markov Chain if

$$ P(X_{k+1} = j | X_k = i, X_{k-1} = i_{k-1}, \cdots, X_0 = i_0) = P(X_{k+1} = j | X_k = i) \quad \forall i, j, i_{k-1}, \cdots, i_0 \text{ and } \forall k \geq 0. $$

In words, a Markov Chain is a stochastic process such that given the value of the current state, the distribution of the next state is independent of the past. (1.1) is called \textit{Markov property}.

\textbf{Note:} (1.1) requires that the conditional probabilities in both sides are well-defined, i.e., $P(X_k = i, X_{k-1} = i_{k-1}, \cdots, X_0 = i_0) > 0$.

A Markov Chain is time-homogeneous if the transition probabilities $P(X_{k+1} = j | X_k = i)$ do not depend on $k$, in which case we denote them as $P_{ij}$. We will assume finite state time-homogeneous Markov Chains from now on. The matrix $P = \{P_{ij}\}_{i,j \in \mathcal{X}}$ is called the \textit{transition matrix} of the Markov Chain $X$. $P$ satisfies the properties that $P_{ij} \geq 0$ for all $i, j \in \mathcal{X}$ and $\sum_{j \in \mathcal{X}} P_{ij} = 1$ for all $i \in \mathcal{X}$, i.e. the entries in each row of $P$ sum to one. A matrix with this property is called \textit{stochastic} (or sometimes row-stochastic).

Let $p(k)$ be a row vector with the $i$th entry being $p_i(k) \triangleq P(X_k = i)$, the probability that the value of the state at time $k$ equals $i$. Then,

$$ p(k) = p(k-1)P. $$

(1.2)

The above equation can be easily checked by noting that $p_j(k) = P(X_k = j) = \sum_{i \in \mathcal{X}} p_i(k-1)P_{ij}$, where the last step is by the law of total probability. By iterating Equation (1.2) $k$ times, we obtain:

$$ p(k) = p(0)P^k. $$

(1.3)
We call $P^k$ the \textit{k-step transition matrix}. The interpretation of $P^k$ is that its entries correspond to the $k$-step transition probabilities $P_{ij}^{(k)} = P(X_{n+k} = j | X_n = i)$ for any $n$. That is, $P_{ij}^{(k)}$ is the probability that we start from state $i$ (at time $n$) and arrive at state $j$ in $k$ steps. From (1.3), we see that the state probabilities at any time $k$ are completely characterized by $P$ and $p(0)$. In fact, the distribution or probability law of a discrete-time homogeneous Markov Chain is determined by its initial distribution and its transition matrix since for any $k \geq 0$ and any $i_0, i_1, \ldots, i_k$:

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k) = p_{i_0}(0)P_{i_0 i_1} \cdots P_{i_{k-1} i_k}.$$ 

Finally, we note here that for any $i, j$ and $n, m \geq 0$ the \textbf{Chapman-Kolmogorov equations} hold:

$$P_{ij}^{(n+m)} = \sum_{k \in \mathcal{X}} P_{ik}^{(n)} P_{kj}^{(m)}.$$ 

These equations provide an alternative justification of the fact that the $k$-step transition matrix $P^{(k)} = [P_{ij}^{(k)}]$ coincides with $P^k$ (i.e., $P$ raised to the power $k$). To see this, use the Chapman-Kolmogorov equations to write $P^{(k)} = P^{(k-1)} P^{(1)} = P^{(k-1)} P$ and iterate this relationship to obtain $P^{(k)} = P^k$.

\textbf{Note}: Due to $P^{(k)} = P^k$ as a consequence of the Chapman-Kolmogorov equations, sometimes in the literature (1.3) (and by extension (1.2)) is called Chapman-Kolmogorov equation.

Moreover, a Markov Chain can be represented by a \textit{transition graph}.

\textbf{Example}. For a Markov Chain with transition matrix $P = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$, the corresponding transition graph is drawn in Figure 1.1.

![Figure 1.1: An example of a Markov Chain transition graph.](Image)

From the example, we see that the transition graph is a directed graph and an arrow from node $i$ to $j$ exists when $P_{ij} > 0$.

Lastly, many homogeneous Markov Chains have a natural representation in terms of a recurrence equation driven by white noise, which is more concretely explained by the following theorem.

\textbf{Theorem 1.2}. Let $\{\xi_n\}_{n \geq 1}$ be an i.i.d. sequence of random variables with values in some arbitrary space $E$. Let $X$ be a countable space and $f : \mathcal{X} \times E \to \mathcal{X}$ be some function. Let also $X_0$ be a random variable with values in $\mathcal{X}$, independent of $\{\xi_n\}_{n \geq 1}$. Then the recurrence $X_{n+1} = f(X_n, \xi_{n+1})$ defines a homogeneous Markov Chain.

\textbf{Example}. (1-D random walk on $\mathbb{Z}$) Suppose $X_{n+1} = X_n + \xi_{n+1}$, with $X_0$ being an integer and $\{\xi_n\}$ is a sequence of independent Rademacher random variables, independent of $X_0$. That is,

$$\xi_n = \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}.$$ 

More generally, $\xi_n$ can take the value $+1$ with probability $p$ and the value $-1$ with probability $1 - p$. Then $X = \{X_n\}_{n \geq 0}$ is a homogeneous Markov Chain by the previous theorem and is called \textit{random walk on $\mathbb{Z}$}.

\textbf{Note}: Not all homogeneous Markov Chains are naturally described by this model (although Discrete-Time Markov Chains (DTMC) that are homogeneous can always be represented \textit{at least distributionwise} by a stochastic recurrence equation $X_{n+1} = f(X_n, \xi_{n+1})$, where $\{\xi_n\}_{n \geq 1}$ is an i.i.d. sequence, independent of $X_0$).
1.2 Communicating class, transience and recurrence, irreducibility

In this section we discuss some properties of the states in a Markov Chain.

A state \( j \) is reachable (or accessible) from state \( i \) if there exists a path from \( i \) to \( j \) in the transition graph \( i \rightarrow i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow j \). This implies that there exists an integer \( n_{ij} \geq 1 \) such that \( P_{ij}^{(n_{ij})} > 0 \). A node, by definition, is said to be always reachable from itself in a Markov Chain.

**Definition 1.3.** (Communicating Class) A set \( C \) of states is a communicating class if for all \( i, j \in C \), \( j \) is reachable from \( i \) and \( i \) is reachable from \( j \), i.e., every pair of states in \( C \) communicates with each other. A communicating class \( C \) is maximal, i.e., adding a state to \( C \) would mean that not all states in \( C \) are reachable from each other anymore.

**Definition 1.4.** (Transience and Recurrence) A state \( i \) is called transient if starting at state \( i \) there is a nonzero probability of never returning to \( i \). A state \( i \) is called recurrent if it is not transient.

![Figure 1.2: An example of a Markov Chain with three communicating classes](image)

**Example.** Figure 1.2 shows a Markov Chain with three communicating classes: \( \{1\} \), \( \{2, 3, 4\} \) and \( \{5\} \). State 1 is transient. State 5 is an absorbing state since it is impossible to leave it.

Recurrence and transience are class properties, i.e. they either hold or do not hold for all states in a communicating class. In the previous example, all states in \( \{2, 3, 4\} \) are transient.

**Definition 1.5.** (Irreducibility) If a Markov Chain has a single communicating class, we say that it is irreducible.

### 1.2.1 Stationary and limiting distributions

**Theorem 1.6.** A finite state irreducible Markov Chain has a unique stationary measure (otherwise known as a stationary or invariant or equilibrium probability distribution). This stationary measure, \( \pi \), is a row probability vector such that

\[ \pi = \pi P. \]  

(1.4)

Theorem (1.6) implies that if \( X_0 \) is drawn from \( \pi \) (we denote this by \( X_0 \sim \pi \)), then \( X_n \sim \pi \) for all \( n \).

We note that because \( P \) is a stochastic matrix, it always has an eigenvalue equal to 1. This can be easily seen by noting that \( P1 = 1 \), where 1 is the all-ones vector of appropriate dimension. Therefore, 1 is an eigenvalue of \( P \) and by extension of \( P^T \), i.e., of the transpose of \( P \). This implies that \( \pi \) is the normalized left eigenvector of \( P \) corresponding to eigenvalue 1.

**Example.** Figure 1.3 represents a two-state Markov Chain with probability matrix \( P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \). Is this Chain irreducible?
**Answer:** No. The Markov Chain has more than one communicating class. Also, if it were irreducible then there would exist a unique stationary distribution. However, for the given $P$, we have that $\pi = \pi P$ for all $\pi$ (since $P$ is the identity matrix).

**Definition 1.7.** *(Limiting Distribution)* The limiting distribution of a Markov Chain is:

$$\lim_{k \to \infty} p(k) = \lim_{k \to \infty} p(0)P^k.$$ 

The limiting distribution may not exist.

![Figure 1.3: An example of a Markov Chain with two states](image)

**Example.** Figure 1.4 represents a two-state Markov Chain with probability matrix $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Suppose that $p(0) = (1 \ 0)$, i.e. the Markov Chain starts in state 1 with probability 1. Running the $p(k)$ recursion, we obtain:

$$p(k) = \begin{cases} (0 \ 1), & \text{if } k \text{ is odd} \\ (1 \ 0), & \text{if } k \text{ is even} \end{cases}.$$ 

Thus we see that the sequence $\{p(k)\}$ oscillates and therefore, it does not have a limiting distribution.

By the definition of the limiting distribution, the associated initial distribution $p(0)$ is important. As we see above, for the initial distribution $p(0) = (1 \ 0)$, the limiting distribution does not exist. However, if $p(0) = (1/2 \ \ 1/2)$, then $p(k) = (1/2 \ \ 1/2)$ for any $k$. Thus, in this particular case, a limiting distribution exists. Additionally, since in the long-run the Markov Chain will spend half of the time in each state, it is trivial to see that the stationary distribution $\pi$ for this chain is $(1/2 \ \ 1/2)$ (can be also derived by solving $\pi = \pi P$ and by taking into account that $\sum_{i \in X} \pi_i = 1$). Therefore, for $p(0) = (1/2 \ \ 1/2)$:

$$\lim_{k \to \infty} p(k) = \pi.$$ 

The limiting distributions (if they exist) are always a subset of the stationary distributions. Therefore, any limiting distribution is a stationary distribution. The converse is not true in general.

**Definition 1.8.** *(Regular transition matrix)* A transition matrix $P$ is regular (or primitive) if there exists a positive integer $k$ such that $P^{(k)}_{ij} > 0 \ \forall \ i, j$. Recall here that $P^{(k)}_{ij}$ is the $(i, j)$th element of $P^k$, which coincides with $P^k$. 

![Figure 1.4: An example of a Markov Chain with two states and no limiting distribution for $p(0) \neq (1/2 \ \ 1/2)$](image)
Regularity is stronger than irreducibility.

*Example.* If \( P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \), then for any integer \( k \) any two entries of \( P^k \) are 0. To see this, note that for all \( n \in \mathbb{N} \),

\[
P^{2n} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\quad \text{and} \quad
P^{2n+1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = P.
\]

Thus this Markov Chain is not regular, but it is irreducible.

If \( P \) is regular, then there exists a unique stationary distribution.