CHAPTER II

BASIC CONCEPTS OF MARKOV CHAINS

§ 2.1 Definition of a Markov process and a Markov chain. We recall that for a finite stochastic process we have a tree and a tree measure and a sequence of outcome functions $f_n$, $n = 0, 1, 2, \ldots$. The domain of $f_n$ is the tree $T_n$ and the range is the set $U_n$ of possible outcomes for the $n$-th experiment. The value of $f_n$ is $s_j$ if the outcome of the $n$-th experiment is $s_j$ (see § 1.9). In the following definitions, whenever a conditional probability $\Pr[q|p]$ occurs, it is assumed that $p$ is not logically false. The reader may find it convenient from time to time to refer to the summary of basic notations and quantities at the end of the book.

A finite stochastic process is an independent process if

(I) *For any statement* $p$ *whose truth value depends only on the outcomes before the* $n$-*th,*

\[ \Pr[f_n = s_j | p] = \Pr[f_n = s_j]. \]

For such a process the knowledge of the outcome of any preceding experiment does not affect our predictions for the next experiment. For a Markov process we weaken this to allow the knowledge of the immediate past to influence these predictions.

2.1.1 Definition. A *finite Markov process is a finite stochastic process* such that

(II) *For any statement* $p$ *whose truth value depends only on the outcomes before the* $n$-*th,*

\[ \Pr[f_n = s_j | (f_{n-1} = s_i) \land p] = \Pr[f_n = s_j | f_{n-1} = s_i]. \]

(*It is assumed that* $f_{n-1} = s_i$ *and* $p$ *are consistent.*)

We shall refer to condition II as the Markov property. For a Markov process, knowing the outcome of the last experiment we can neglect any other information we have about the past in predicting the future. It is important to realize that this is the case only if we
This condition says essentially that, given the present, the past and future are independent of each other. This more symmetric definition suggests in turn that a Markov process should remain a Markov process if it is observed in reverse order. That the latter is true is seen from the following theorem. (We shall not prove this theorem.)

2.1.5 Theorem. Given a Markov process let \( p \) be any statement whose truth value depends only on experiments after the \( n \)-th experiment. Then

\[
\Pr[f_n = s_j | (f_{n+1} = s_t) \land p] = \Pr[f_n = s_j | f_{n+1} = s_t].
\]

Since a Markov process observed in reverse order remains a Markov process, it might be suspected that the same is true for a Markov chain. This would be the case if the “backward transition probabilities,” \( p^*_{ij}(n) = \Pr[f_n = s_j | f_{n+1} = s_t] \), were independent of \( n \). These probabilities may be found as follows:

\[
p^*_{ij}(n) = \frac{\Pr[f_n = s_j \land f_{n+1} = s_t]}{\Pr[f_{n+1} = s_t]} = \frac{\Pr[f_{n+1} = s_t | f_n = s_j] \cdot \Pr[f_n = s_j]}{\Pr[f_{n+1} = s_t]} = \frac{p_{ji} \cdot \Pr[f_n = s_j]}{\Pr[f_{n+1} = s_t]}.
\]

These transition probabilities would be independent of \( n \) only if the probability of being in a particular state at time \( n \) was independent of \( n \). This is certainly not the case in general. For example, if the system is started in state \( s_1 \) with probability 1, then the probability that it is there on the next step is \( p_{11} \). Thus, in general, \( \Pr[f_0 = s_1] \neq \Pr[f_1 = s_1] \). Thus a Markov chain looked at in reverse order will be a Markov process, but in general its transition probabilities will depend on time and hence it will not be a Markov chain. We will return to this problem in § 5.3.

§2.2 Examples. In this section we shall give several simple examples of Markov chains which will be used in future work for illustrative purposes. The first five examples relate to what is normally called a “random walk.” We imagine a particle which moves in a straight line in unit steps. Each step is one unit to the right with probability \( p \) or one unit to the left with probability \( q \). It moves until it reaches one of two extreme points which are called “boundary points.” The possibilities for its behavior at these points determine
several different kinds of Markov chains. The states are the possible positions. We take the case of 5 states, states $s_1$ and $s_5$ being the "boundary" states, and $s_2$, $s_3$, $s_4$ the "interior states."

\[
\begin{array}{c|ccccc}
   & s_1 & s_2 & s_3 & s_4 & s_5 \\
\hline
s_1 & 1 & 0 & 0 & 0 & 0 \\
s_2 & q & 0 & p & 0 & 0 \\
s_3 & 0 & q & 0 & p & 0 \\
s_4 & 0 & 0 & q & 0 & p \\
s_5 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\]

**Example 1**

Assume that if the process reaches state $s_1$ or $s_5$ it remains there from that time on. In this case the transition matrix is given by

\[
P = s_3 \begin{pmatrix}
   s_1 & s_2 & s_3 & s_4 & s_5 \\
   1 & 0 & 0 & 0 & 0 \\
   q & 0 & p & 0 & 0 \\
   0 & q & 0 & p & 0 \\
   0 & 0 & q & 0 & p \\
   0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}.
\]

(1)

**Example 2**

Assume now that the particle is "reflected" when it reaches a boundary point and returns to the point from which it came. Thus if it ever hits $s_1$ it goes on the next step back to $s_2$. If it hits $s_5$ it goes on the next step back to $s_4$. The matrix of transition probabilities becomes in this case

\[
P = s_3 \begin{pmatrix}
   s_1 & s_2 & s_3 & s_4 & s_5 \\
   0 & 1 & 0 & 0 & 0 \\
   q & 0 & p & 0 & 0 \\
   0 & q & 0 & p & 0 \\
   0 & 0 & q & 0 & p \\
   0 & 0 & 0 & 1 & 0 \\
\end{pmatrix}.
\]

(2)

**Example 3**

As a third possibility we assume that whenever the particle hits one of the boundary states, it goes directly to the center state $s_3$. We may think of this as the process of Example 1 started at state $s_3$ and...
repeated each time the boundary is reached. The transition matrix is

$$
P = \begin{pmatrix}
s_1 & s_2 & s_3 & s_4 & s_5 \\
s_1 & 0 & 0 & 1 & 0 & 0 \\
s_2 & q & 0 & p & 0 & 0 \\
s_3 & 0 & q & 0 & 0 & p \\
s_4 & 0 & 0 & q & 0 & p \\
s_5 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
$$

(Eq. 3)

**Example 4**

Assume now that once a boundary state is reached the particle stays at this state with probability $\frac{1}{2}$ and moves to the other boundary state with probability $\frac{1}{2}$. In this case the transition matrix is

$$
P = \begin{pmatrix}
s_1 & s_2 & s_3 & s_4 & s_5 \\
s_1 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\
s_2 & q & 0 & p & 0 & 0 \\
s_3 & 0 & q & 0 & 0 & p \\
s_4 & 0 & 0 & q & 0 & p \\
s_5 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2}
\end{pmatrix}
$$

(Eq. 4)

**Example 5**

As the final choice for the behavior at the boundary, let us assume that when the particle reaches one boundary it moves directly to the other. The transition matrix is

$$
P = \begin{pmatrix}
s_1 & s_2 & s_3 & s_4 & s_5 \\
s_1 & 0 & 0 & 0 & 0 & 0 \\
s_2 & q & 0 & p & 0 & 0 \\
s_3 & 0 & q & 0 & 0 & p \\
s_4 & 0 & 0 & q & 0 & p \\
s_5 & 1 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

(Eq. 5)

**Example 6**

We next consider a modified version of the random walk. If the process is in one of the three interior states, it has equal probability of moving right, moving left, or staying in its present state. If it is
on the boundary, it cannot stay, but has equal probability of moving to any of the four other states. The transition matrix is:

\[
P = \begin{pmatrix}
  s_1 & s_2 & s_3 & s_4 & s_5 \\
  s_1 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
  s_2 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\
  s_3 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\
  s_4 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
  s_5 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\
\end{pmatrix}
\]  

(6)

**Example 7**

A sequence of digits is generated at random. We take as states the following: \( s_1 \) if a 0 occurs, \( s_2 \) if a 1 or 2 occurs, \( s_3 \) if a 3, 4, 5, or 6 occurs, \( s_4 \) if a 7 or 8 occurs, \( s_5 \) if a 9 occurs. This process is an independent trials process, but we shall see that Markov chain theory gives us information even about this special case. The transition matrix is

\[
P = \begin{pmatrix}
  s_1 & s_2 & s_3 & s_4 & s_5 \\
  s_1 & .1 & .2 & .4 & .2 & .1 \\
  s_2 & .1 & .2 & .4 & .2 & .1 \\
  s_3 & .1 & .2 & .4 & .2 & .1 \\
  s_4 & .1 & .2 & .4 & .2 & .1 \\
  s_5 & .1 & .2 & .4 & .2 & .1 \\
\end{pmatrix}
\]  

(7)

**Example 8**

According to *Finite Mathematics* (Chapter V, Section 8), in the Land of Oz they never have two nice days in a row. If they have a nice day they are just as likely to have snow as rain the next day. If they have snow (or rain) they have an even chance of having the same the next day. If there is a change from snow or rain, only half of the time is this a change to a nice day. We form a three-state Markov chain with states R, N, and S for rain, nice, and snow, respectively. The transition matrix is then

\[
P = \begin{pmatrix}
  R & N & S \\
  R & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
  N & \frac{1}{2} & 0 & \frac{1}{2} \\
  S & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\
\end{pmatrix}
\]  

(8)
In our previous examples the Markov property clearly held. In this case it could only be regarded as an approximation since the knowledge of the weather the last two days, for example, might lead us to different predictions than knowing the weather only on the previous day. One way to improve this approximation is to take as states the weather for two successive days. The states would then be NN, NR, NS, RN, RR, RS, SN, SR, SS. New transition probabilities would have to be estimated. A single step would still be one day, so that from NR, for example, we could move only to states RN, RR, RS. In examples of this kind it is possible to improve the approximation, still using the Markov chain theory, but at the expense of increasing the number of states.

**Example 9**

An urn contains two unpainted balls. At a sequence of times a ball is chosen at random, painted either red or black, and put back. If the ball was unpainted, the choice of color is made at random. If it is painted, its color is changed. We form a Markov chain by taking as a state three numbers \((x, y, z)\) where \(x\) is the number of unpainted balls, \(y\) the number of red balls, and \(z\) the number of black balls. The transition matrix is then

\[
\begin{pmatrix}
(0,1,1) & (0,2,0) & (0,0,2) & (2,0,0) & (1,1,0) & (1,0,1) \\
(0,1,1) & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\
(0,2,0) & 1 & 0 & 0 & 0 & 0 & 0 \\
(0,0,2) & 1 & 0 & 0 & 0 & 0 & 0 \\
(2,0,0) & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\
(1,1,0) & 1/4 & 1/4 & 0 & 0 & 0 & 1/2 \\
(1,0,1) & 1/4 & 0 & 1/4 & 0 & 1/2 & 0 \\
\end{pmatrix}
\]

**Example 10**

Assume that a student going to a certain college has each year a probability \(p\) of flunking out, a probability \(q\) of having to repeat the year, and a probability \(r\) of passing on to the next year. We form a Markov chain, taking as states \(s_1\)—has flunked out, \(s_2\)—has graduated, \(s_3\)—is a senior, \(s_4\)—is a junior, \(s_5\)—is a sophomore, \(s_6\)—is a freshman.
The transition matrix is then

$$
P = \begin{pmatrix}
    s_1 & s_2 & s_3 & s_4 & s_5 & s_6 \\
    s_1 & 1 & 0 & 0 & 0 & 0 \\
    s_2 & 0 & 1 & 0 & 0 & 0 \\
    s_3 & p & r & q & 0 & 0 \\
    s_4 & p & 0 & r & q & 0 \\
    s_5 & p & 0 & 0 & r & q \\
    s_6 & p & 0 & 0 & 0 & r & q
\end{pmatrix}.
$$

(10)

**EXAMPLE 11**

A man is playing two slot-machines. The first machine pays off with probability $c$, the second with probability $d$. If he loses, he plays the same machine again; if he wins, he switches to the other machine. Let $s_i$ be the state of playing the $i$-th machine. The transition matrix is

$$
P = \begin{pmatrix}
    s_1 & s_2 \\
    s_1 & 1-c & c \\
    s_2 & d & 1-d
\end{pmatrix}.
$$

(11)

As $c$ and $d$ take on all permissible values ($0 \leq c \leq 1$, $0 \leq d \leq 1$) we get all $2 \times 2$ Markov chains.

**EXAMPLE 12**

Consider the special two-state Markov chain (Example 11) with transition matrix

$$
P = \begin{pmatrix}
    s_1 & s_2 \\
    s_1 & 1/2 & 1/2 \\
    s_2 & 1/4 & 3/4
\end{pmatrix}.
$$

(11a)

(This can be called Example 11a.)

From this Markov chain we form a new Markov chain as follows. A state in the new chain will be a pair of states in the old chain. That is, the states are $s_is_1$, $s_is_2$, $s_2s_1$, $s_2s_2$. The new chain is in state $s_is_j$ on the $n$-th step if the old chain was in state $s_i$ on the $n$-th step and $s_j$ on the $(n+1)$-th step.
The transition matrix for the new chain (Example 12) is

\[
P = \begin{pmatrix}
    s_{1s_1} & s_{1s_2} & s_{2s_1} & s_{2s_2} \\
    \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
    0 & 0 & \frac{1}{4} & \frac{3}{4} \\
    \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
    0 & 0 & \frac{1}{4} & \frac{3}{4}
\end{pmatrix}.
\] (12)

We shall see in § 6.5 that the study of this new chain gives us more detailed information about the original process than could be obtained directly from the two-state chain.

§ 2.3 Connection with matrix theory. In this section we shall show the connection between Markov chain theory and matrix theory. We shall start with the general finite Markov process and then specialize our results to the finite Markov chain.

2.3.1 Theorem. Let \( f_n \) be the outcome function at time \( n \) for a finite Markov process with transition probabilities \( p_{ij}(n) \). then

\[
\Pr[f_n = s_v] = \sum_u \Pr[f_{n-1} = s_u] p_{uv}(n).
\]

Proof. The statement \( f_n = s_v \) is a statement relative to the tree \( T_n \). To find its probability, we add the weights of all paths in its truth set. That is, all possible paths which end in outcome \( s_v \). Thus if \( j, k, \ldots, u \) is a possible sequence of states

\[
\Pr[f_n = s_v]
\]

\[
= \sum_{j, k, \ldots, u} \Pr[f_0 = s_j \land \cdots \land f_{n-1} = s_u \land f_n = s_v].
\]

\[
= \sum_{j, k, \ldots, u} \Pr[f_0 = s_j \land \cdots \land f_{n-1} = s_u] \cdot \Pr[f_n = s_v | f_0 = s_j \land \cdots \land f_{n-1} = s_u].
\]

By the Markov property this is

\[
\sum_{j, k, \ldots, u} \Pr[f_0 = s_j \land \cdots \land f_{n-1} = s_u] p_{uv}(n).
\]

If in this last sum we keep \( u \) fixed and sum over the remaining indices we obtain

\[
\Pr[f_n = s_v] = \sum_u \Pr[f_{n-1} = s_u] p_{uv}(n).
\]

This completes the proof.

We can write the result of this theorem in matrix form. Let \( \pi_n \)
be a row vector which gives the induced measure for the outcome function $f_n$. That is

$$\pi_n = \{p^{(n)}_1, p^{(n)}_2, \ldots, p^{(n)}_r\},$$

where $p^{(n)}_j = \text{Pr}[f_n = s_j]$. Thus, $p^{(n)}_j$ is the probability that the process will after $n$ steps be in state $s_j$. The vector $\pi_0$ is the initial probabilities vector. Let $P(n)$ be the matrix with entries $p_{ij}(n)$. Then the result of Theorem 2.3.1 may be written in the form

$$\pi_n = \pi_{n-1} \cdot P(n)$$

for $n \geq 1$. By successive application of this result we have

$$\pi_n = \pi_0 \cdot P(1) \cdot P(2) \cdot \ldots \cdot P(n).$$

In the case of a Markov chain process, all the $\Gamma(n)$'s are the same and we obtain the following fundamental theorem.

**2.3.2 Theorem.** Let $\pi_n$ be the induced measure for the outcome function $f_n$ for a finite Markov chain with initial probability vector $\pi_0$ and transition matrix $P$. Then

$$\pi_n = \pi_0 \cdot P^n.$$

This theorem shows that the key to the study of the induced measures for the outcome functions of a finite Markov chain is the study of the powers of the transition matrix. The entries of these powers have themselves an interesting probabilistic interpretation. To see this,
take as initial vector $\pi_0$ the vector with 1 in the $i$-th component and 0 otherwise. Then by Theorem 3.2, $\pi_n = \pi_0 P^n$. But $\pi_0 P^n$ is the $i$-th row of the matrix $P^n$. Thus the $i$-th row of the $n$-th power of the transition matrix gives the probability of being in each of the various states under the assumption that the process started in state $s_i$.

In Example 1, let us assume that the process starts in state $s_3$. Then $\pi_0 = \{0, 0, 1, 0, 0\}$. We can find the induced measures (see § 1.7) for the first three outcome functions by constructing a tree and tree measure for the first three experiments. This tree is given in Figure 2-1.

From this tree and tree measure we easily compute the induced measures for the functions $f_1, f_2, f_3$. They are

$$\pi_1 = \{0, \frac{1}{2}, 0, \frac{1}{2}, 0\}$$
$$\pi_2 = \{\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4}\}$$
$$\pi_3 = \{\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4}, \frac{1}{4}\}.$$

By Theorem 2.3.2 these induced measures should also be the third row in the matrices $P$, $P^2$, and $P^3$, since the starting state was $s_3$. These matrices are

$$P = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

$$P^2 = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\
\frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\
0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

$$P^3 = \begin{pmatrix}
1 & 0 & 0 & 0 & \frac{1}{8} \\
\frac{5}{8} & 0 & \frac{1}{4} & 0 & \frac{3}{8} \\
\frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{8} & 0 & \frac{1}{4} & 0 & \frac{5}{8} \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}.$$
We thus see that these matrices furnish us several tree measures simultaneously.

§ 2.4 Classification of states and chains. We wish to classify the states of a Markov chain according to whether it is possible to go from a given state to another given state. This problem is exactly like the one treated in § 1.4. If we interpret \( i \mathbf{T j} \) to mean that the process can go from state \( s_i \) to state \( s_j \) (not necessarily in one step), then all the results of that section are applicable.

In particular, the states are divided into equivalence classes. Two states are in the same equivalence class if they “communicate,” i.e. if one can go from either state to the other one. The resulting partial ordering shows us the possible directions in which the process can proceed.

The minimal elements of the partial ordering are of particular interest.

2.4.1 Definition. The minimal elements of the partial ordering of equivalence classes are called ergodic sets. The remaining elements are called transient sets. The elements of a transient set are called transient states. The elements of an ergodic set are called ergodic (or non-transient) states.

Since every finite partial ordering must have at least one minimal element, there must be at least one ergodic set for every Markov chain. However, there need be no transient set. The latter will occur if the entire chain consists of a single ergodic set, or if there are several ergodic sets which do not communicate with others.

From the results of § 1.4 we see that if a process leaves a transient set it can never return to this set, while if it once enters an ergodic set, it can never leave it. In particular, if an ergodic set contains only one element, then we have a state which once entered cannot be left. Such a state is called absorbing. Since from such a state we cannot go to another state, the following theorem characterizes absorbing states.

2.4.2 Theorem. A state \( s_i \) is absorbing if and only if \( p_{ii} = 1 \).

It is convenient to use our classification to arrive at a canonical form for the transition matrix. We renumber the states as follows: The elements of a given equivalence class will receive consecutive numbers. The minimal sets will come first, then sets that are one level above the minimal sets, then sets two levels above the minimal sets, etc. This will assure us that we can go from a given state to another in the same class, or to a state in an earlier class, but not to a state in a later class. If the equivalence classes arranged as here
described are $u_1, u_2, \ldots, u_k$, then our matrix will appear as follows (where $k$ is taken as 5, for the sake of illustration):

\[
\begin{pmatrix}
  P_1 \\
  R_2 & P_2 & O \\
  R_3 \\
  R_4 & P_4 \\
  R_5 & P_5
\end{pmatrix}
\]

Here the $P_i$ represent transition matrices within a given equivalence class. The region $O$ consists entirely of 0's. The matrix $R_i$ will be entirely 0 if $P_i$ is an ergodic set, but will have non-zero elements otherwise.

In this form it is easy to see what happens as $P$ is raised to powers. Each power will be a matrix of the same form; in $P^n$ we still have zeros in the upper region, and we simply have $P^n_i$ in the diagonal regions. This shows that a given equivalence class can be studied in isolation, by treating the submatrix $P_i$. This will be considered in detail later.

We can also apply the subdivision of an equivalence class considered in the previous chapter. We saw there that each equivalence class can be partitioned into cyclic classes. If there is only one cyclic class, then we say that the equivalence class is \textit{regular}, otherwise we say that it is \textit{cyclic}.

If an equivalence class is regular, then after sufficient time has elapsed the process can be in any state of the class, no matter which of the equivalent states it started in (see § 1.4). This means that all sufficiently high powers of its $P_i$ must be positive (i.e. have only positive entries). If the equivalence class is cyclic, then no power of $P_i$ can be positive.

From this classification of states we can arrive at a classification of Markov chains. We have noted that there must be an ergodic set, but there need be no transient set. This will lead to our primary subdivision. Within this we can subdivide according to the number and type of ergodic sets.

\textbf{I. Chains Without Transient Sets}

If such a chain has more than one ergodic set, then there is absolutely no interaction between these sets. Hence we have two or more unrelated Markov chains lumped together. These chains may be studied separately, and hence without loss of generality we may
assume that the entire chain is a single ergodic set. A chain consisting of a single ergodic set is called an *ergodic chain*.

I-A. The ergodic set is regular. In this case the chain is called a *regular Markov chain*. As we see from previous considerations, all sufficiently high powers of $P$ must be positive in this case. Thus no matter where the process starts, after sufficient lapse of time it could be in any state.

I-B. The ergodic set is cyclic. In this case the chain is called a *cyclic Markov chain*. Such a chain has a period $d$, and its states are subdivided into $d$ cyclic sets ($d > 1$). For a given starting position it will move through the cyclic sets in a definite order, returning to the set of the starting state after $d$ steps. We also know that after sufficient time has elapsed, the process can be in any state of the cyclic set appropriate for the moment.

II. Chains With Transient Sets

In such a chain the process moves towards the ergodic sets. As will be seen in the next chapter, the probability that the process is in an ergodic set tends to 1; and it cannot escape from an ergodic set once it enters it. Hence it is fruitful to classify such chains by their ergodic sets.

II-A. All ergodic sets are unit sets. Such a chain is called an *absorbing chain*. In this case the process is eventually trapped in a single (absorbing) state. This type of process can also be characterized by the fact that all the ergodic states are absorbing states.

II-B. All ergodic sets are regular, but not all are unit sets.

II-C. All ergodic sets are cyclic.

II-D. There are both cyclic and regular ergodic sets.

Naturally, in each of these classes we can further classify chains according to how many ergodic sets there are. Of particular interest is the question whether there are one or more ergodic sets.

We can illustrate all of these types except II-D by the random walk examples.

*For Example 1*: The states $s_1$ and $s_5$ are absorbing states. The states $s_2, s_3, s_4$ are transient states. It is possible to go between any two of these states. Hence they form a single transient set. We have an absorbing Markov chain—that is, case II-A.

*For Example 2*: In this example it is possible to go from any state to any other state. Hence there are no transient states and there is a single ergodic set. Thus we have an ergodic chain. It is possible to return to a state only in an even number of steps. Thus the period
of the states is 2. The two cyclic sets are \( \{s_1, s_3, s_5\} \) and \( \{s_2, s_4\} \). This is type I-B.

*For Example 3:* Again we can go from any state to any other state. Hence we again have an ergodic chain. It is possible to return to state \( s_3 \) from \( s_3 \) in either two or three steps. Hence the greatest common divisor \( d = 1 \), and the period is 1. This is type I-A.

*For Example 4:* In this example \( \{s_1, s_5\} \) is an ergodic set which is clearly regular. The set \( \{s_2, s_3, s_4\} \) is the single transient set. This is type II-B.

*For Example 5:* Here we have a single ergodic set \( \{s_1, s_5\} \) which has period 2. The set \( \{s_2, s_3, s_4\} \) is again a transient set. This is type II-C.

§ 2.5 Problems to be studied. Let us consider our various types of chains, and ask what types of problems we would like to answer in the following chapters.

First of all we may wish to study a regular Markov chain. In such a chain the process keeps moving through all the states, no matter where it starts. Some of the questions of interest are:

1. If a chain starts in \( s_i \), what is the probability after \( n \) steps that it will be in \( s_j \)?
2. Can we predict the average number of times that the process is in \( s_i \)? And if so, how does this depend on where the process starts?
3. We may wish to consider the process as it goes from \( s_i \) to \( s_j \). What is the mean and variance of the number of steps needed? What are the mean and variance of the number of states passed? What is the probability that the process passes through \( s_k \)?
4. We may wish to study a certain subset of states, and observe the process only when it is in these states. How does this modify our previous results? These questions are treated in Chapter IV.

Next we may wish to study a cyclic chain. Here the same kinds of questions are of interest as for a regular chain. Naturally, a regular chain is easier to study; and we will find that, once we have the answers for regular chains, it is not hard to find the corresponding answers for all ergodic chains. This extension of regular chain theory to the theory of ergodic chains is carried out in Chapter V.

Next we may wish to consider a Markov chain with transient states. There are two kinds of questions to be asked here. One will concern the behavior of the chain before it enters an ergodic set, while the other kind will apply after the chain has entered an ergodic set. The latter questions are no different from the ones considered above. Once a chain enters an ergodic set it can never leave it, and hence the existence of states outside the set is irrelevant. Thus questions of the second kind can be answered by considering a chain consisting of a single ergodic set, i.e. an ergodic chain.
The really new questions concern the behavior of the chain up to the moment that it enters an ergodic set. However, for these questions the nature of the ergodic states is irrelevant, and we may make them all into absorbing states if we wish. More generally, if we wish to study the process while it is in a set of transient states, we may make all other states absorbing. This modified process will serve to find all the answers we desire. Hence the only new questions concern the behavior of an absorbing chain.

Some of the questions that are of interest concerning a transient state \( s_i \) are:

1. The probability of entering a given ergodic set, starting from \( s_i \).
2. The mean and variance of the number of times that the process is in \( s_i \) before entering an ergodic set, and how this number depends on the starting position.
3. The mean and variance of the number of steps needed before entering an ergodic set starting at \( s_i \).
4. The mean number of states passed before entering an ergodic set, starting at \( s_i \).

Chapter III will deal with absorbing chains, and all these questions will be answered. Thus we will find the most interesting questions about finite Markov chains answered in Chapters III, IV, and V.

Exercises for Chapter II

For § 2.1

1. Five points are marked on a circle. A process moves from a given point to one of its neighbors, with probability \( \frac{1}{2} \) for each neighbor. Find the transition matrix of the resulting Markov chain.

2. Three tanks fight a duel. Tank A hits its target with probability \( \frac{2}{3} \), tank B with probability \( \frac{1}{2} \), and tank C with probability \( \frac{1}{3} \). Shots are fired simultaneously, and once a tank is hit it is out of action. As a state we choose the set of tanks still in action. If on each step each tank fires at its strongest opponent, verify that the following transition matrix is correct:

\[
\begin{pmatrix}
E & A & B & C & AC & BC & ABC \\
E & 1 & 0 & 0 & 0 & 0 & 0 \\
A & 0 & 1 & 0 & 0 & 0 & 0 \\
B & 0 & 0 & 1 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 1 & 0 & 0 \\
AC & \frac{2}{9} & \frac{4}{9} & 0 & \frac{1}{9} & \frac{2}{9} & 0 \\
BC & \frac{1}{6} & 0 & \frac{1}{3} & \frac{1}{6} & 0 & \frac{1}{3} \\
ABC & 0 & 0 & 0 & \frac{4}{9} & \frac{2}{9} & \frac{2}{9} & \frac{1}{9}
\end{pmatrix}
\]
3. Modify the transition matrix in the previous exercise, assuming that when all tanks are in action, A fires at B, B at C, and C at A.

4. We carry out a sequence of experiments as follows: At first a fair coin is tossed. Then, if experiment \( n - 1 \) comes out heads, we toss a fair coin; if it comes out tails, we toss a coin which has probability \( 1/n \) of coming up heads. What are the transition probabilities? What kind of process is this?

For § 2.2

5. Modify Example 1 by assuming that when the process reaches \( s_1 \) it goes on the next step to state \( s_2 \). Form the new transition matrix.

6. Modify the process described in Example 2 by assuming that when the process reaches \( s_1 \) it stays there for the next two steps and on the third step moves to state \( s_2 \). Show that the resulting process is not a Markov chain (with the five given states).

7. In Exercise 6 show that we can treat the process as a Markov chain, by allowing a larger number of states. Write down the transition matrix.

8. Modify the transition matrix of Example 7, assuming that the digit 0 is twice as likely to be generated than any other digit.

9. Modify Example 7, assuming that the same digit is never generated twice in a row, but otherwise digits are equally likely to occur.

10. In Example 8 allow only two states: Nice and not nice. Show that the process is still a Markov chain, and find its transition matrix.

For § 2.3

11. In Example 11a compute \( P^2, P^4, P^8, \) and \( P^{16} \), and write the entries as decimal fractions. Note the trend, and interpret your results.

12. Show that, no matter how Example 7 is started, the probabilities for being in each of the states after 1 step agree with the common row for the transition matrix. What are the probabilities after \( n \) steps?

13. Assume that Example 8 is started with initial vector \( \pi_0 = (2/5, 1/5, 2/5) \). Find \( \pi_1, \pi_2 \). What is \( \pi_n \)?

14. The weather is nice today in the Land of Oz. What kind of weather is most likely to occur day after tomorrow?

15. In Example 11, assume that \( c = \frac{1}{2} \) and \( d = \frac{1}{4} \). The man randomly chooses the machine to play first. What is the probability that he plays the better machine (a) on the second play, (b) on the third play, and (c) on the fourth play?

16. In Example 2 assume that the process is started in state \( s_3 \). Construct a tree and tree measure for the first three experiments. Use this to find the induced measure for the first three outcome functions. Verify that your results agree with the probabilities found from \( P, P^2, \) and \( P^3 \).

For § 2.4

17. For the following Markov chain, give a complete classification of the states and put the transition matrix in canonical form.
18. A Markov chain has the following transition matrix, with non-zero entries marked by $x$. Give a complete classification of the states and put the transition matrix in canonical form.

$$P = \begin{pmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0
\end{pmatrix}.$$ 

19. Classify the following chains as ergodic or absorbing. Which of the ergodic chains is regular?

(a) $P = \begin{pmatrix}
\frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2}
\end{pmatrix}$

(b) $P = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{pmatrix}$

(c) $P = \begin{pmatrix}
0 & \frac{1}{2} & \frac{1}{2} \\
1 & 0 & 0 \\
1 & 0 & 0
\end{pmatrix}$

(d) $P = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{pmatrix}$

(e) $P = \begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
\frac{1}{2} & \frac{1}{2} & 0
\end{pmatrix}$
20. In Example 9 classify the states. Put the transition matrix in canonical form. What type of chain is this?

21. For an ergodic chain the $i$-th state is made absorbing by replacing the $i$-th row in the transition matrix by a row with a 1 in the $i$-th component. Prove that the resulting chain is absorbing.

22. In Example 11, give conditions on $c$ and $d$ so that the resulting chain is

(a) ergodic  (b) regular  (c) cyclic  (d) absorbing

For the entire chapter

23. In a certain state a voter is allowed to change his party affiliation (for primary elections) only by abstaining from the primary for one year. Let $s_1$ indicate that a man votes Democratic, $s_2$ that he votes Republican, and $s_3$ that he abstains, in the given year. Experience shows that a Democrat will abstain $\frac{1}{2}$ the time in the following primary, a Republican will abstain $\frac{1}{4}$ time, while a voter who abstained for a year is equally likely to vote for either party in the next election. [We will refer to this as Example 13.]

(a) Find the transition matrix.
(b) Find the probability that a man who votes Democratic this year will abstain three years from now.
(c) Classify the states.
(d) In a given year $\frac{1}{4}$ of the population votes Democratic, $\frac{1}{2}$ Republican, the rest abstain. What proportions do you expect in the next primary election?

24. A sequence of experiments is performed, in each of which two fair coins are tossed. Let $s_1$ indicate that two heads come up, $s_2$ that a head and a tail come up, and $s_3$ that two tails turn up. [We will refer to this as Example 14.]

(a) Find the transition matrix.
(b) If two heads turn up on a given toss, what is the probability of two heads turning up three tosses later?
(c) Classify the states.
know exactly the outcome of the last experiment. For example, if we know only that the outcome of the last experiment was either $s_i$ or $s_k$ then knowledge of the truth value of a statement $p$ relating to earlier experiments may affect our future predictions.

2.1.2 Definition. The $n$-th step transition probabilities for a Markov process, denoted by $p_{ij}(n)$ are

$$p_{ij}(n) = \Pr[f_n=s_j|f_{n-1}=s_i].$$

2.1.3 Definition. A finite Markov chain is a finite Markov process such that the transition probabilities $p_{ij}(n)$ do not depend on $n$. In this case they are denoted by $p_{ij}$. The elements of $U$ are called states.

2.1.4 Definition. The transition matrix for a Markov chain is the matrix $P$ with entries $p_{ij}$. The initial probability vector is the vector $\pi_0 = \{p_j^{(0)}\} = \{\Pr[f_0=s_j]\}$.

For a Markov chain we may visualize a process which moves from state to state. It starts in $s_j$ with probability $p^{(0)}_j$. If at any time it is in state $s_i$, then it moves on the next “step” to $s_j$ with probability $p_{ij}$. The initial probabilities are thought of as giving the probabilities for the various possible starting states. The initial probability vector and the transition matrix completely determine the Markov chain process, since they are sufficient to build the entire tree measure. Thus, given any probability vector $\pi_0$ and any probability matrix $P$, there is a unique Markov chain (except possibly for renaming the states) which will have the $\pi_0$ as initial probability vector and $P$ as transition matrix.

In most of our discussions we will consider a fixed transition matrix $P$, but we will wish to vary the initial vector $\pi$. The tree measure assigned will depend on the initial vector $\pi$ that is chosen. Hence if $p$ is any statement relative to the tree, or $f$ is a function with domain the tree, $\Pr[p]$, $M[f]$, and $\Var[f]$ all depend on $\pi$. We indicate this by writing $\Pr_{\pi}[p]$, $M_{\pi}[f]$ and $\Var_{\pi}[f]$. The special case where $\pi$ has a 1 in the $i$-th component (process is started in state $s_i$) is denoted $\Pr_i[p]$, $M_i[f]$, $\Var_i[f]$.

We shall give several examples of Markov chains in the next section. We conclude this section with a few brief remarks about the Markov property.

It can be easily proved that the Markov property is equivalent to the following property more symmetric with respect to time.

\[ (\Pi') \text{ Let } p \text{ be any statement whose truth value depends only on outcomes after the } n \text{-th experiment and } q \text{ be any statement whose truth value depends only on outcomes before the } n \text{-th experiment. Then } \]

$$\Pr[p \land q | f_n = s_j] = \Pr[p | f_n = s_j] \cdot \Pr[q | f_n = s_j].$$