UNIT 2 THE BASICS OF MARKOV CHAIN

Structure **Page No.**

2.1 INTRODUCTION

The **Markov chain** is named after Andrey Markov (1856 – 1922), a Russian mathematician. It is a discrete-time stochastic process with the Markov property. Andrey Markov produced the first results in 1906 for these processes having finite state space. A generalization to countably infinite state spaces was given by Kolmogorov. Further work was done by W. Doeblin, W. Feller, K. L. Chung and others. In most of our study of probability so far, we have dealt with independent trials processes as a sequence of identically and independently distributed random variables. These processes are the basis of classical probability theory, and much of statistics. We have discussed two of the principal theorems for these processes: the Law of Large Numbers, and the Central Limit Theorem. We have seen that when a sequence of repeated chance experiments forms an independent trials process, the possible outcomes for each experiment are the same and occur with the same probability. Further, knowledge of the outcomes of the previous experiments does not influence our predictions for the outcomes of the present or future experiment.

In many cases in real life, we observe a sequence of chance experiments where all of the outcomes in the past experiments may influence our predictions for the next experiment. The sequence of random variables associated with the sequence of such experiments may not be identically and independently distributed. For example, this will happen in predicting a student's grades on a sequence of exams in a course. But to allow too much generality makes the processes mathematically difficult to handle. A. Markov studied this type of chance process where the outcome of current experiment (not previous experiments) can only affect the outcome of the next experiment. This type of process is called a Markov process, a particular case of which, when state space is discrete, is called a Markov chain.

Markovian systems appear extensively in physics. Markov chains can also be used to model various processes in queuing theory. The Page Rank of a webpage as used by Google is defined by a Markov chain. Markov chain methods have also become very important for generating sequences of random numbers to accurately reflect very complicated desired probability distributions – a process called Markov chain Monte Carlo, or MCMC for short. Markov chains also have many applications in biological modeling, particularly population processes, which are useful in modeling processes that are (at least) analogous to biological populations. The Leslie matrix is one such example, though some of its entries are not probabilities (they may be greater than 1).

We will present some discussion about the concept of stochastic processes, definition and understanding of Markov chain in Sec.2.2 and Sec.2.3, respectively. We will also present some examples to illustrate the behavior of Markov chain. Here, we will also learn about the Transition Probability Matrix P, higher order Transition Probabilities, and the famous Chapman-Kolomogorov equation. In Sec. 2.4, we will represent a Markov chain graphically, and in Sec. 2.4, we shall compute higher order transition probability. In Sec. 2.6, we will learn two methods for calculation of $\mathbf{P}^n$, viz., Spectral Decomposition, and Generating Function.

Objectives

After studying this unit you should be able to:

- explain the concept of a stochastic process, and that of a Markov chain as a special case of stochastic process;
- compute the transition probability matrix with some of its applications;
- evaluate higher order transition probabilities, and unconditional probability distribution after a number of steps in a Markov chain;
- explain the two methods of calculating $\mathbf{P}^n$ –Spectral Decomposition and Generating Function.

2.2 STOCHASTIC PROCESS

Let us start this section by discussing the following situations:

(i) Consider a simple experiment like a series of independent throwings of a coin. Suppose that X_n denotes the total number of heads found in the first n throws. Then $\{X_n, n = 1, 2, 3, \ldots\}$ is a family of random variables constituting a stochastic process.

(ii) Consider another simple experiment. Suppose a dice is thrown a number of times, and suppose that X_n is the number of sixes in the first n throws. If we allow n to vary as $n = 1, 2, \ldots$, then we get a sequence random variables $\{X_n, n = 1, 2, 3, \ldots\}$. When n varies, we have a family of random variables constituting a stochastic process.

Stochastic process is also known as chance or random process.

A stochastic process is defined as an indexed collection of random variables $\{\mathbf{X}_n\}$, where the index, n, belongs to an index set, T. In most real life situations, this set represents time, either discrete or continuous. The collection of random variables is defined as some sample space. The set of all possible values taken by these random variables is known as, **state space** of the stochastic process and we will denote it by S. The **state space** is called **discrete** if it contains a finite, or, countably infinite number of points, and it is called **continuous** when it is an interval or union of disjoint intervals.

For example, in situation (i). $\mathbf{X}_n$ denotes the total number of heads found in n independent throws of a coin. Thus the state space, S, will be a finite set of non-negative integers, $0, 1, 2, \ldots, n$. Here, the collection of random variables $\{\mathbf{X}_n\}$, will be a stochastic process having finite state space. In situation (ii), the state space of $\mathbf{X}_n$ is also discrete. We can write $X_n = Y_1 + Y_2 + \cdots + Y_n$, where Y_i is a discrete random variable denoting the outcome of the ith throw and $Y_i = 1$ or 0 accordingly as the ith throw shows a six or not. Representation $X_n = Y_1 + \cdots + Y_n$ is valid in both the situations (i) and (ii). In an another situations, we may consider a collection of random variables $\{\mathbf{X}_n = Y_1 + Y_2 + \cdots + Y_n, n = 1, 2, 3, \ldots\}$ where Y_i is a continuous random variable assuming values in $(0, \infty)$. Here, the set of possible values of X_n belong to the interval $(0, \infty)$, and so the state space S of the stochastic process $\mathbf{X}_n$ is continuous.

From the examples above, it is clear that a stochastic process may be a discrete time stochastic process, when the index set is a discrete set T, often a collection of the non-negative integers $0, 1, 2, 3, \ldots$, or it may be continuous time stochastic process when the index set is continuous (usually space or time interval), resulting in an uncountably infinite number of random variables. We may use alternative notation for a stochastic process such as $X(t)$ or X_t where t indicates space or time in day.

So far, we have discussed the case of a stochastic process in which $X(t)$ are one = dimensional random variable. There may be processes with $X(t)$ that are more than one = dimensional. Consider $X(t) = (X_1(t), X_2(t))$, in which $X_1(t)$ represents the minimum temperature, and $X_2(t)$ represents the maximum temperature in a city in a time interval $[0, t]$, then the stochastic process is two = dimensional. Similarly, we can have a multi-dimensional stochastic process also. In general, stochastic processes can be categorized into the following four types:

(i) discrete state space and discrete time

(ii) discrete state space and continuous time

(iii) continuous state space and discrete time, and

(iv) continuous state space and continuous time.

Thus, we see that the index set, T, and the state space, S, of a stochastic process may be discrete or continuous. Familiar examples of the stochastic processes include prices of shares, varying every moment in a stock market, and exchange rates of our currency fluctuating along with time. Other examples, such as a patient's ECG, blood pressure, or temperature, constitute stochastic processes arising in medical sciences.

In the next section, we shall discuss Markov chains.

2.3 MARKOV CHAIN

A discrete time Markov chain is a stochastic process where both the index set I and the state space S are discrete and the stochastic process satisfies markov property. A sequence of random variables $\{X_n\}$ is said to follow markov property if we are given the present state, that is, the value taken by the random variable is X_n, then the states of the future, that is, values of random variables $X_{n+1}, X_{n+2}, \ldots$ are independent of the states of the past, that is, the value of the random variables $X_{n-1}, X_{n-2}, \ldots$. For example, if the stock price of a stock in the National Stock Exchange follows markov property then the stock price at a future date will depend only on the current price that is known to us, and will not depend on its prices during past dates. The Markov chain and Markov property may be formally defined as follows.

Definition 1: A stochastic process $\{X_i\}$ with the index set $T = \{0, 1, 2, \ldots, i, \ldots\}$ and discrete state space $S = \{1, 2, \ldots, \ell, \ldots s\}$ is called a Markov chain, if for any of the states, $i_0, i_1, i_2, i_3, \ldots, i_{n-1}, i, j \in S$, and any $n \in T$, we have

$$\begin{aligned} P[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_2 = i_2, X_1 = i_1, X_0 = i_0] \\ = P[X_{n+1} = j \mid X_n = i] \end{aligned} \quad (1)$$

and in this situation, the sequence of random variables $\{X_n\}$ is said to possess the **Markov Property.** If X_n has the outcome i (i.e. $X_n = i$), then the Markov chain is said to be in state i at nth trial, or at time n. In the definition above, s may be infinity.

The Markov chain will be called a Finite Markov chain if the state space S is finite.

The probability $P[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_2 = i_2, X_1 = i_1, X_0 = i_0]$, in the above definition denotes the conditional probability that the system will be in state j at time $n+1$, given that the system was in the state i at time n, in the state i_{n-1} at time $n-1, \ldots,$ in the state i_1 at time 1, in the state i_0 initially at time 0. Due to the Markov property this probability depends only on the latest given state, i.e., on the state i, at time n.

Let (i, j) denote a pair of states at the two times, say, at time m and n, $m \leq n$. The transition probability for making the transition from state i, at time m, to state j at time n.

$$P[X_n = j \mid X_m = i] = p_{ij}(m,n) \qquad (2)$$

is called $m-n$ step transition probability.

Here, we have assumed that the transition probabilities depend on both the states i, j, and both the times m, n.

Definition 2: The unconditional probability distribution of the initial random variable X_0 of the Markov chain $\{X_n\}$ is called the **initial distribution** of the chain. The Markov chain starts in a state chosen according to the probability distribution of X_0. Let the vector $\mathbf{u} = (u_1, u_2, u_3, \ldots, u_s)$ be the vector having s elements corresponding to s states namely $1, 2, \ldots, s$ such that $u_i = P(X_0 = i), i = 1,2,\ldots,s$. Thus u_i denotes the probability that the chain starts in state i at time 0.

Definition 3: A Markov chain is called **time homogeneous** or with **stationary transition probabilities,** if its transition probabilities $p_{ij}(m,n)$ do not depend on the specific times, m and n, but depend only on time duration $n-m$, i.e., on the number of steps taken between two times.

In this section, we shall only discuss the time homogeneous chains. In this case, the m-step transition probability for a homogeneous chain may be denoted as

$$P[X_{n+m} = j \mid X_n = i] = p_{ij}^{(m)}, \text{ for any } n \text{ in the index set } I \qquad (3)$$

and one step transition probability as $P[X_{n+1} = j \mid X_n = i] = p_{ij}$ (here, we denote $p_{ij}^{(1)} = p_{ij}$ omitting the superscript (1) for convenience.)

Definition 4 (Transition Matrix): Suppose the state space S of a time homogeneous Markov chain contains s states $1, 2, 3, \ldots, s$, then the $s \times s$ matrix of the one-step transition probabilities (p_{ij}) is called a P-matrix and is denoted by $\mathbf{P}$. This square matrix is also called the **matrix of transition probabilities**, or the **transition matrix.** Since the $(i, j)^{th}$ element of $\mathbf{P}$ represents the one-step transition probability p_{ij}, that is the probability that the chain will move from the state i to the state j in one step.

Therefore, the sum of elements in each row of P is one, i.e. $\sum_{j=1}^{s} p_{ij} = 1$ for all i.

A square matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13}\cdots\cdots \\ p_{21} & p_{22} & p_{23}\cdots\cdots \\ \cdots\cdots\cdots & \cdots\cdots\cdots & \cdots\cdots\cdots \\ \cdots\cdots\cdots & \cdots\cdots\cdots & \cdots\cdots\cdots \end{bmatrix}.$$

with non negative elements, and row-sum unity is called a **stochastic matrix**. When the column-sums are also unity, then this matrix is called a **doubly stochastic**. Thus, a transition matrix is stochastic.

Remark 1: A problem can be modeled as a (homogeneous) **Markov chain** if it has the following properties:

a) For any unit time period, a_n object in the system is in exactly one of the defined states. At the end of the time period, the object either moves to a new state, or stays in that same state for another unit time period.

b) The object moves from one state to the next according to the transition probabilities which depend only on the current state of the object, and not on any previous history of its states. The total probability of movement, out from a state (movement from a state to the same state does count as movement) is equal to one.

c) The transition probabilities do not change over time (the probability of going from state A to state B in the current unit time period is the same as it will be at any other period in the future).

Now, we state below two theorems without proof. **Theorem 1** is known as the **general existence theorem. Theorem 2** states three different conditions identical to the Markov property. For the proof of these theorems, you may refer to **Markov chain with Transition Probabilities** by K. L. Chung (1967).

Theorem 1: The stochastic matrix and the initial distribution completely specify a Markov chain.

Theorem 2: The markov property referred to in Eqn.(1) is equivalent to any one of the following three results. Let the states $i, j, i_1, i_2, i_3, \ldots$ be any states of the Markov chain $\{X_n\}$ then

1. for any $n_1 < n_2 < n_3 < \ldots n_k < n_{k+1}$

$$P[X_{n_{k+1}} = j \mid X_{n_k} = i, X_{n_{k-1}} = i_{k-1}, \ldots, X_{n_2} = i_2, X_{n_1} = i_1] = P[X_{n_{k+1}} = j \mid X_{n_k} = i] \tag{4}$$

2. for any $n_1 < n_2 < n_3 < \ldots . < n_k$

$$\begin{aligned} P[X_{n_k} &= i_k, X_{n_{k-1}} = i_{k-1}, \ldots, X_{n_2} = i_2, X_{n_1} = i_1] \\ &= \{P[X_{n_k} = i_k \mid X_{n_{k-1}} = i_{k-1}] \\ &\ldots\ldots\ P[X_{n_2} = i_2 \mid X_{n_1} = i_1] P[X_{n_1} = i_1]\} \end{aligned} \tag{5}$$

3. $$\begin{aligned} P[X_{n+1} &= j, X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0] \\ &= P[X_{n+1} = j \mid X_n = i] \ \ldots\ldots\ P[X_1 = i_1 \mid X_0 = i_0] P[X_0 = i_0] \end{aligned} \tag{6}$$

From Egn. (6) it follows that the joint probability distribution of $\{X_0, X_1, X_2, X_3, \ldots, X_n\}$ of the Markov chain $\{X_n\}$ is completely determined if the initial distribution, and the transition matrix of the chain are known. Let us prove this result.

Starting with the joint probability of $\{X_0, X_1, X_2, X_3, \ldots, X_n\}$, we have

$$\begin{aligned} P[X_n &= i_n, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0] \\ &= \{P[X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0] \ldots\ldots \\ &\quad P[X_2 = i_2 \mid X_1 = i_1, X_0 = i_0] P[X_1 = i_1 \mid X_0 = i_0] P[X_0 = i_0]\} \end{aligned}$$

using conditional probability and product rule.

$$=\{P[X_n = i_n \mid X_{n-1} = i_{n-1}] \ldots\ldots$$
$$P[X_2 = i_2 \mid X_1 = i_1]P[X_1 = i_1 \mid X_0 = i_0]P[X_0 = i_0]\}$$ using the markov property
$$= p_{i_n i_{n-1}} \cdots p_{i_3 i_2}\, p_{i_2 i_1}\, p_{i_1 i_0}\, u_{i_0},$$

where u_{i_0} is the initial probability and $p_{i_n i_{n-1}}, \ldots, p_{i_3 i_2}, p_{i_2 i_1}$ are transition probabilities.

***

Example 1 (Simple Weather Model): Let us consider three possible conditions of weather at any day say, Sunny (S), Cloudy (C), and Rainy (R). Suppose the probability that a sunny day will follow a sunny day be 0.75, and that the cloudy day will follow a sunny day be 0.15 and the rainy day will follow a sunny day be 0.10. Similarly, the probability that a cloudy day will follow a sunny day be 0.25, it will follow a cloudy day be 0.45, and it will follow rainy day be 0.30. The probability that a rainy day will follow a sunny day be 0.15, it will follow a cloudy day be 0.45 and it will follow that a rainy day be 0.40. We assume that each day's weather condition depends only on the condition of the previous day. Therefore, with this information we may form a Markov chain $\{X_n\}$, where X_n represents weather condition of the n^{th} day. We may take three conditions of weather S, C, and R as the states denoted by, numbers 1, 2, 3 respectively for the Markov chain.

From the information above, we can determine the transition probabilities as follows

$$p_{11} = P(X_n = 1 \mid X_{n-1} = 1) = 0.75, \quad p_{12} = P(X_n = 2 \mid X_{n-1} = 1) = 0.15$$
$$p_{13} = P(X_n = 3 \mid X_{n-1} = 1) = 0.10, \quad p_{21} = P(X_n = 1 \mid X_{n-1} = 2) = 0.25$$
$$p_{22} = P(X_n = 2 \mid X_{n-1} = 2) = 0.45, \quad p_{23} = P(X_n = 1 \mid X_{n-1} = 2) = 0.30$$
$$p_{31} = P(X_n = 1 \mid X_{n-1} = 3) = 0.15, \quad p_{32} = P(X_n = 3 \mid X_{n-1} = 2) = 0.45$$
$$p_{33} = P(X_n = 3 \mid X_{n-1} = 3) = 0.40$$

These are conveniently presented in a 3×3 square transition matrix $\mathbf{P}$ given below.

$$P = \begin{array}{c} \\ S \\ C \\ R \end{array}\!\!\begin{array}{c} \begin{array}{ccc} S & C & R \end{array} \\ \begin{pmatrix} 0.75 & 0.15 & 0.10 \\ 0.25 & 0.45 & 0.30 \\ 0.15 & 0.45 & 0.40 \end{pmatrix} \end{array}$$

***

Example 2: Consider that in a city in the coming week the probability that a healthy person will fall sick is 0.20, and that he will remain healthy is 0.80. Consider another case where, in the coming week, the probabilities that a sick person will become healthy is 0.65, will die be 0.25 and will remain sick is 0.10. We will form a P-matrix on the basis of the above information. In each week, a person will be in any of three conditions – healthy, sick or dead, which are the three states for his health. If each week's health condition depends on the condition on the previous week only, then we have a Markov chain $\{X_n\}$, where X_n represents health condition of a person in the n^{th} week. From the above information we can determine the transition probabilities. Assume the states Healthy, Sick and Dead are denoted as 1, 2, 3 respectively, then

$$p_{11} = P(X_n = 1 \mid X_{n-1} = 1) = 0.80, \quad p_{12} = P(X_n = 2 \mid X_{n-1} = 1) = 0.20,$$
$$p_{13} = P(X_n = 3 \mid X_{n-1} = 1) = 00, \quad p_{21} = P(X_n = 1 \mid X_{n-1} = 2) = 0.65,$$
$$p_{22} = P(X_n = 2 \mid X_{n-1} = 2) = 0.10, \quad p_{23} = P(X_n = 3 \mid X_{n-1} = 2) = 0.25,$$
$$p_{31} = P(X_n = 1 \mid X_{n-1} = 3) = 00, \quad p_{32} = P(X_n = 2 \mid X_{n-1} = 3) = 00,$$
$$p_{33} = P(X_n = 3 \mid X_{n-1} = 3) = 1$$

The transition matrix $\mathbf{P}$ is obtained as shown below.

$$P = \begin{array}{c} \\ H \\ S \\ D \end{array} \begin{array}{c} \begin{array}{ccc} H & S & D \end{array} \\ \begin{pmatrix} 0.80 & 0.20 & 0 \\ 0.65 & 0.10 & 0.25 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

***

So far, we have only defined a Markov chain. Now, let us discuss a graphical representation of a Markov chain.

2.4 GRAPHICAL REPRESENTATION

Markov chains may be depicted by a directed graph or digraph, the graph having directed edges. The states of a Markov chain are represented by the vertices, or the nodes of the graph, and the single step transitions between the states by the directed arcs (edges) joining the vertices. For instance, in Example 1, the probability of the single step transition from the state, rainy (R) to sunny (S) is 0.15. Then vertices labeled R and S are joined by an arc (also a called edge) directed from vertex R to the vertex S. The arc is labeled by the corresponding probability as 0.15, in this case. Likewise, transition from S to S, with probability 0.75, is represented by a self-loop labeled by 0.75 at the vertex S. No edge is drawn corresponding to a transition probability zero. Thus, the number of edges, including self loops, will equal the number of positive entries in the one step transition probability matrix, which is 9 in this case. Let the graph denoted by $G = (V, E)$. Then V, being the set of vertices representing different states of the Markov chain, and E, being the set edges representing all possible non-zero transitional probabilities. This digraph is called a **transition graph**. In a transition graph, the sum of the probabilities of all the edges emanating from each node will be one. Conversely, if in a labeled digraph all the labels of the edges are positive numbers, and the sum of all the labels of the edges emanating from each node is one, then such a graph is called **stochastic graph**, and we can define a Markov chain with this digraph as its transition graph.

Example 3: The directed graph of the Markov chain given in the Example 1 is shown in Fig 1.

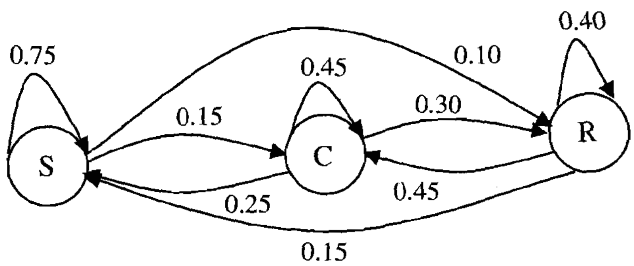

Fig.1

Example 4: The directed graph of the Markov chain given in Example 2 is shown in Fig 2.

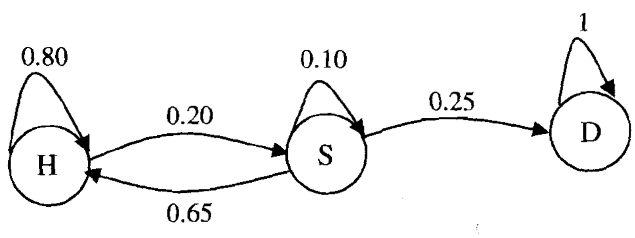

Fig.2

Example 5: The prime minister of a country tells a journalist, X, about his intention to run, or not to run in the next election. The journalist transmits this information to Y and Y transmits it to Z, and so forth. We assume that there is a probability 'a' that a person will change the answer from "yes" to "no" when transmitting it to the next person, and a probability 'b' that a person will change it from "no" to "yes". We choose the messages, either "yes" or "no", to the fellow journalists as states. It may be expressed as a Markov chain $\{X_n\}$ where X_n denotes the transmitted message by the n^{th} person to the next person. X_0 denotes the intention revealed by the prime minister at the start. Here, we have denoted two states "yes" and "no" as 1 and 2 respectively. From the above information, we can determine the transition probabilities.

$$p_{11} = P(X_n = 1 | X_{n-1} = 1) = 1 - a, \qquad p_{12} = P(X_n = 2 | X_{n-1} = 1) = a$$

$$p_{21} = P(X_n = 1 | X_{n-1} = 2) = b, \qquad p_{22} = P(X_n = 2 | X_{n-1} = 2) = 1 - b$$

Therefore, the P-matrix will be

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \end{array}, (0 < a, b < 1)$$

***

Example 6: Each time a certain horse runs in a three-horse race, he has probability $1/2$ of winning (W), $1/4$ of coming in second (S), and 1/4 of coming in third (T), independent of the outcome of any previous race. We have an independent trials process, but it can also be considered as a Markov chain. Here, we choose outcomes of the race, that is, winning, second, and third as three states. It may be modeled as a Markov chain $\{X_n\}$, where X_n denotes the outcome of the n^{th} race. From the above information, we can determine the transition probabilities shown in the transition matrix that follows.

$$P = \begin{array}{c c} & \begin{array}{ccc} W & S & T \end{array} \\ \begin{array}{c} W \\ S \\ T \end{array} & \begin{pmatrix} 0.50 & 0.25 & 0.25 \\ 0.50 & 0.25 & 0.25 \\ 0.50 & 0.25 & 0.25 \end{pmatrix} \end{array}$$

Remark 2: In general, we see that any sequence of discrete i.i.d. (identically and independently distributed) random variables can be considered as a Markov chain. In such a case, the transition matrix has identical rows, each row being the probability distribution of the random variable, X_n.

You may now try the following exercises on the basis of above discussions.

E1) Assume that a man's profession can be classified as business, agriculture, or public servant. It is known from past data that, of the sons of businessmen, 80% are businessmen, 10% are farmers and 10% are public servants. In the case of sons of farmers, 60% are farmers, 20% are businessmen, and 20% are public servants. Finally, in the case of public servants 50% of the sons are public servants and 25% each are in the other two categories. Assume that every man has at least one son. Does the choice of profession by sons in the successive generations in a family form a Markov chain.? If so write down its matrix of transition probabilities.

E2) Draw the transition graph for the problem given in **Example 6.**

E3) The schooling status of a student in any year may be represented by 6 states, namely, nursery, class one, class two, ... class five. Let p_i denotes the probability that a student in state i in any year jumps to a higher class (state $i+1$) and q_i denotes the probability that a student remains to the same class (state i) in the next year. Assume that class 5 is the highest status and it can not be crossed. If X_n denotes the status of a student in the nth year of his schooling, show that $\{X_n\}$ is a Markov chain. Set up the matrix of transition probabilities.

So far, we have learnt about the Markov chain and its graphical representation. Now in this section we shall continue the discussion to the higher step transition probabilities.

2.5 HIGHER ORDER TRANSITION PROBABILITIES

Definition 5 (Higher Steps Transition Probability Matrix): The n –step transition probability for the transition from state i to j in n steps in a homogeneous Markov chain, denoted by $p_{ij}^{(n)}$, was defined in Equation 3. The matrix $P^{(n)} = (p_{ij}^{(n)})$ is called n -step transition probability matrix.
When $n=1$, we have $\mathbf{P}^{(1)} = (p_{ij}^{(1)}) = (p_{ij}) = \mathbf{P}$.

For convenience, we define $\mathbf{P}^{(0)} = \mathbf{I}$, where I is an identity matrix.

The unconditional probability distribution of X_n, the state of Markov chain at the step n, is defined as $u_i^{(n)} = P[X_n = j]$. $j = 1, 2, 3, \ldots, s$.
The unconditional probability distribution of X_n in the vector form may be denoted as
$\mathbf{u}^{(n)} = \left(u_i^{(n)}, u_2^{(n)}, \ldots, u_s^{(n)}\right)$
Now, we will prove some results providing a relation between $P^{(n)}$ and the $\mathbf{P}$ -matrix. These result will be useful in computing higher order transition probabilities.

Theorem 3: The n-step transition probabilities satisfy the recurrence relation
$p_{ij}^{(n)} = \sum_{k=1}^{s} p_{ik}^{(n-1)} p_{kj}$ for $i, j \in S$, the matrix form of which can be written as
$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1)}\mathbf{P}$

Proof: Using the Law of Total Probability and Conditional Probability discussed in Unit 1, we have

$$P_{ij}^{(n)} = P[X_n = j \mid X_0 = i]$$

$$= \sum_{k=1}^{s} P[X_n = j, X_{n-1} = k \mid X_0 = i] \text{ (using the Law of Total Probability)}$$

$$= \sum_{k=1}^{s} \left\{P[X_n = j \mid X_{n-1} = k, X_0 = i]P\left[X_{n-1} = k \mid X_0 = i\right]\right\} \text{ (using conditional Probability)}$$

$$= \sum_{k=1}^{s} \left\{P\left[X_n = j \mid X_{n-1} = k\right]P\left[X_{n-1} = k \mid X_0 = i\right]\right\} \text{ (using the Markov Property)}$$

$$= \sum_{k=1}^{s} p_{ik}^{(n-1)} p_{kj}$$

The last expression is the ij^{th} element in the multiplication of matrices $P^{(n-1)}$ and $P = (p_{ij})$. Thus, we get

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1)}\mathbf{P} \text{ [since } (p_{ij}^{(n)}) = \mathbf{P}^{(n)}\text{]}$$

***

Theorem 4: Let P be the transition matrix of a homogeneous Markov chain. The ij^{th} entry of the matrix $\mathbf{P}^n$ gives the probability that the Markov chain, starting in the state i initially, will be in state j after n steps, i.e.

$$\mathbf{P}^{(n)} = \mathbf{P}^n$$

Proof: Clearly, the probability that the Markov chain, starting in state i, will be in state j after n steps is $p_{ij}^{(n)}$, which is the ij^{th} entry of the matrix $\mathbf{P}^{(n)}$. Therefore, the theorem will be proved if we prove $\mathbf{P}^{(n)} = \mathbf{P}^n$.

Now, let us apply the method of induction to prove this. For $n = 2$,

From **Theorem 4** we have $\mathbf{P}^{(2)} = \mathbf{P}^{(1)}\mathbf{P} = \mathbf{P}\mathbf{P} = \mathbf{P}^2$.

Again, assuming the result for n, we can verify it for $n+1$ as follows.

$\mathbf{P}^{(n+1)} = \mathbf{P}^{(n)}\mathbf{P}$ (using Theorem 4)

$= \mathbf{P}^n\mathbf{P}$ (using the assumption for n)

$= \mathbf{P}^{n+1}$

Hence, by the method of induction, it is proved that the statement is true for every positive integer, n.

***

Theorem 5 (Chapman-Kolomogorov Equation): A time homogeneous Markov chain satisfies the equation

$$p_{ij}^{(m+n)} = \sum_{k=1}^{s} p_{ik}^{(m)} p_{kj}^{(n)},\ i, j = 1, 2, \ldots, s, \text{ for } m, n = 0, 1, 2, \ldots$$

or, in matrix form $\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)}\mathbf{P}^{(n)}$, $\mathbf{P}^{(0)} = I$

Proof: $p_{ij}^{(m+n)} = P[X_{m+n} = j \mid X_0 = i]$

$= \sum_{k=1}^{s} P[X_{m+n} = j, X_n = k \mid X_0 = i]$ (using the Law of Total Probability)

$= \sum_{k=1}^{s} \{P[X_{m+n} = j \mid X_n = k, X_0 = i]$

$P[X_n = k \mid X_0 = i]\}$ (using Conditional Probability)

$= \sum_{k=1}^{s} \{P[X_{m+n} = j \mid X_n = k] P[X_n = k \mid X_0 = i]\}$ (using the Markov Property)

$= \sum_{k=1}^{s} p_{ik}^{(n)} p_{kj}^{(m)},\ i, j = 1, 2, \ldots, s$

Using matrix multiplication representation, we have, for every m, n

$\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)}\ \mathbf{P}^{(n)}$, where we define $\mathbf{P}^{(0)} = I$

***

Theorem 6: Let $\mathbf{P}$ be the transition matrix of a time homogeneous Markov chain, and let $\mathbf{u}$ be the initial probability vector. Then the unconditional probability $P(X_n = j)$ that the chain is in state j after n steps, is the j^{th} entry in the vector $u^{(n)}$ given by

$\mathbf{u}^{(n)} = \mathbf{u}\mathbf{P}^n$

Proof: Since,

$$u_j^{(n)} = P[X_n = j]$$

$$= \sum_{i=1}^{s} P[X_n = j, X_0 = i] \text{ (using the Law of Total Probability)}$$

$$= \sum_{i=1}^{s} P[X_n = j \mid X_0 = i] P[X_0 = i]$$

$$= \sum_{i=1}^{s} p_{ij}^{(n)} u_i$$

The result may be expressed in matrix form as

$$\mathbf{u}^{(n)} = \mathbf{u}\mathbf{P}^{(n)}$$

$$= \mathbf{u}\mathbf{P}^{n} \quad \text{(Using theorem 5)}$$

***

Example 7 (Random Walk): We consider a particle which performs a random walk on a real line on the set of non-negative integers $\{0, 1, 2, \ldots, N\}$ as shown in **Fig. 3** with $N+1$ possible positions. If, at any time, the particle is at the position i (i can be $1, 2, \ldots, N-1$), then in the next unit of time it can move one step forward $(+1)$ to position $i+1$, or one step backward (-1) to the position $i-1$, with probabilities $p(0 < p < 1)$ and $q(q = 1-p)$, respectively.

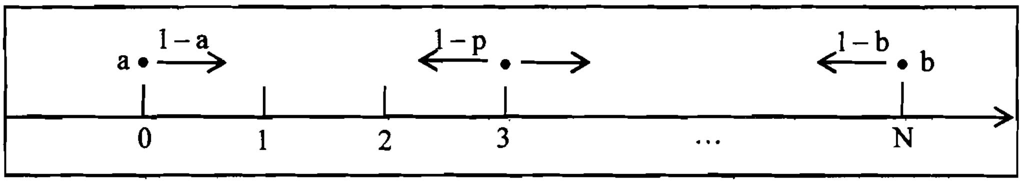

Fig.3: The Random Walk

At the end points, 0 and N, there are two typical behaviours for the particle. If the particle reaches at 0, then it remains at 0 with probability 'a' or moves to 1 with probability '$1-a$'. Similarly, assume the particle remains at N with probability 'b' and moves to $N-1$ with probability '$1-b$' whenever it reaches that position. The position 0 will be an absorbing barrier when $a = 1$, and it will be a reflecting barrier if $a = 0$. The position 0 will be called elastic barrier or partially reflective barrier if $0 < a < 1$. Similarly, the position N will be absorbing when $b = 1$, reflective when $b = 0$, and elastic/partially reflective when $0 < b < 1$.

Suppose the particle starts in a position, $k(0 \le k \le N)$, at time 0. Let X_n denotes the position of the particle at time n. Then, clearly, sequence $\{X_n\}$ follows the Markov property. The $N+1$ possible positions $\{0, 1, 2, \ldots, N\}$ of the particle are the possible states of the chain.

Here, for $0 < r < N$

$$P[X_n = r+1 \mid X_{n-1} = r] = p$$

$$P[X_n = r-1 \mid X_{n-1} = r] = q$$

Also, when $r = 0$

$$P[X_n = 1 \mid X_{n-1} = 0] = 1-a$$

$$P[X_n = 0 \mid X_{n-1} = 0] = a$$

and, when $r = N$

$$P[X_n = N-1 \mid X_{n-1} = N] = 1-b$$

$$P[X_n = N \mid X_{n-1} = N] = b$$

The transition matrix is found as

$$P = \begin{array}{c} \\ 0 \\ 1 \\ \cdots \\ N-1 \\ n \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & \cdots & N-2 & N-1 & N \end{array} \\ \begin{pmatrix} a & 1-a & 0 & \cdots & 0 & 0 & 0 \\ q & 0 & p & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & 0 & \cdots & 0 & 1-b & b \end{pmatrix} \end{array}$$

and, initial probability vector is $u = (0, 0, \ldots, 1, 0, \ldots, 0)$, with 1 at the $k+1^{th}$ place in the vector.

***

Example 8 (Gambler s Ruin Problem): There are two gamblers A and B playing against each other. Let the initial capital of A be x units, and B is $z-x$ units. At each move player A can win one unit from B, with probability p, or can lose one unit to B, with probability $q(p+q=1)$. In due course, after a series of independent moves, if the capital of A reduces to zero, then A is ruined, and the game ends, and if his capital increases to z, then B will be ruined and the game ends. This problem can be modeled as a random walk problem with absorbing barriers at two ends. Here, the Markov chain $\{X_n\}$, represents the capital of A at the nth move of the game. It has $z+1$ states ranging from 0 to z. The transition probability matrix can be obtained directly from Example 7 by putting $a=1, b=1$ and $N=z$. Also, initial state $k=x$ with probability 1.

***

Example 9: Let the initial distribution in Example 1 of the Simple Weather Model be $\mathbf{u} = \{0.7 \quad 0.2 \quad 0.1\}$. Let the three states sunny, cloudy and rainy be represented by integers $1, 2, 3$ respectively.

Then, the probability that the initial day is sunny, the first day is rainy, the second day is cloudy, and the third day is sunny, is given by:

$$\begin{aligned} &P[X_0=1, X_1=3, X_2=2, X_3=1]. \\ &\quad =\{P[X_0=1]\, P[X_1=3 \mid X_0=1] P[X_2=2 \mid X_0=1, X_1=3] \\ &\qquad P[X_3=1 \mid X_0=1, X_1=3, X_2=2]\} \\ &\quad =\{P[X_0=1]\, P[X_1=3 \mid X_0=1] P[X_2=2 \mid X_1=3] \\ &\qquad P[X_3=1 \mid X_2=2]\} \\ &\quad =(0.7)(0.10)(0.45)(0.25) \\ &\quad =0.0079 \end{aligned}$$

Also, the probability that all successive four days starting from the initial day are sunny equals:

$$\begin{aligned} &P[X_0=1, X_1=1, X_2=1, X_3=1] \\ &\quad =\{P[X_0=1]\, P[X_1=1 \mid X_0=1] P[X_2=1 \mid X_0=1, X_1=1] \\ &\qquad P[X_3=1 \mid X_0=1, X_1=1, X_2=1]\} \\ &\quad =\{P[X_0=1]\, P[X_1=1 \mid X_0=1] P[X_2=1 \mid X_1=1] \\ &\qquad P[X_3=1 \mid X_2=1]\} \\ &\quad =(0.7)(0.75)^3 = 0.2953 \end{aligned}$$

***

Example 10: Suppose that in Example 9 the P-Matrix is modified as given below

$$\mathbf{P}=\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0.500 & 0.250 & 0.250 \\ 0.450 & 0.100 & 0.450 \\ 0.250 & 0.250 & 0.500 \end{pmatrix} \end{array}$$

Let us now find the probability distribution of weather for the first day, second day, and the third day, and also the probability distribution of the sixth day. The probability distribution of weather for the first day is the probability distribution of X_1. Now

$$\begin{aligned} u_1^{(1)} &= P[X_1=1] \\ &= \sum_{i=1}^{3} P[X_1=1, X_0=i] \\ &= \sum_{i=1}^{3} P[X_0=i]P[X_1=1 | X_0=i] \\ &= \sum_{i=1}^{3} u_i\, p_{i1} = (0.7)(0.5)+(0.2)(0.45)+(0.1)(0.25) \\ &= 0.465 \end{aligned}$$

Similarly,

$$\begin{aligned} u_2^{(1)} &= P[X_1=2] \\ &= \sum_{i=1}^{3} u_i\, p_{i2} = (0.7)(0.25)+(0.2)(0.1)+(0.1)(0.25) \\ &= 0.22 \end{aligned}$$

and

$$\begin{aligned} u_3^{(1)} &= P[X_1=3] \\ &= \sum_{i=1}^{3} u_i\, p_{i3} = (0.7)(0.25)+(0.2)(0.45)+(0.1)(0.5) \\ &= 0.315 \end{aligned}$$

The probabilities may be written in vector form as

$\mathbf{u}^{(1)} = \{0.465 \;\; 0.22 \;\; 0.315\}$

The distribution of X_1 may also be obtained by using formula $\mathbf{u}^{(n)} = \mathbf{u}\mathbf{P}^n$.

$\mathbf{u}^{(1)} = \mathbf{u}\mathbf{P}$

$$= (0.7 \;\; 0.2 \;\; 0.1)\begin{pmatrix} 0.500 & 0.250 & 0.250 \\ 0.450 & 0.100 & 0.450 \\ 0.250 & 0.250 & 0.500 \end{pmatrix} = (0.465 \;\; 0.22 \;\; 0.315)$$

We may get distribution of X_2, the probability distribution of weather for the second day, as

$\mathbf{u}^{(2)} = \mathbf{u}\mathbf{P}^2$

$$= (0.7 \;\; 0.2 \;\; 0.1)\begin{pmatrix} 0.500 & 0.250 & 0.250 \\ 0.450 & 0.100 & 0.450 \\ 0.250 & 0.250 & 0.500 \end{pmatrix}\begin{pmatrix} 0.500 & 0.250 & 0.250 \\ 0.450 & 0.100 & 0.450 \\ 0.250 & 0.250 & 0.500 \end{pmatrix}$$

$$= (0.7 \;\; 0.2 \;\; 0.1)\begin{pmatrix} 0.425 & 0.213 & 0.363 \\ 0.383 & 0.235 & 0.383 \\ 0.363 & 0.213 & 0.425 \end{pmatrix}$$

$$= (0.410 \quad 0.217 \quad 0.373)$$

Likewise the distribution of X_3, the probability distribution of weather for the third day, as

$\mathbf{u}^{(3)} = \mathbf{uP}^3$

$$= (0.7 \quad 0.2 \quad 0.1)\begin{pmatrix} 0.399 & 0.218 & 0.383 \\ 0.393 & 0.215 & 0.393 \\ 0.383 & 0.218 & 0.399 \end{pmatrix}$$

$$= [0.396 \quad 0.217 \quad 0.387]$$

and the distribution of X_6, the probability distribution of weather for the sixth day, will be

$\mathbf{u}^{(6)} = \mathbf{uP}^6$

$$= (0.7 \quad 0.2 \quad 0.1)\begin{pmatrix} 0.391 & 0.217 & 0.391 \\ 0.391 & 0.217 & 0.391 \\ 0.391 & 0.217 & 0.391 \end{pmatrix}$$

$$= (0.391 \quad 0.217 \quad 0.391)$$

***

Remark 3: Here we see that sixth day probability distribution of weather has become independent of the initial distribution. You may verify that the same distribution for the sixth day will be found if any other initial distribution is used. This happens since all rows of the $\mathbf{P}^6$ are identical and define a probability distribution on a set of states. You may also find more powers of P higher than 6. Are they identical to $\mathbf{P}^6$? If, for a large n, all rows of the $\mathbf{P}^n$ become identical and define probability distribution, then the Markov chain is called **Regular Markov chain**. We shall discuss these chains in Unit 3.

Example 11 (Partial Sum): Let $\{X_n\}$ be i.i.d. (identically and independently distributed) random variables, taking only non-negative integral values. Let $S_n = \sum_{i=1}^{n} X_i$ and $S_0 = 0$. Then $S_n = S_{n-1} + X_n$. Since, distribution of S_n depends only on S_{n-1} and not on any of the $S_{n-2}, S_{n-3}, \ldots, S_0$, the sequence $\{S_n\}$ is a Markov chain with state space $S = \{0, 1, 2, \ldots, j, \ldots\}$.

Again,

$$p_{ij} = P[S_n = j \mid S_{n-1} = i]$$
$$= P[X_n = j - i] = p_{j-i} \quad \text{(say)}$$

Therefore, Markov chain $\{S_n\}$ is time homogeneous, or has stationary transition probabilities p_{ij}, given above. Here, p_{ij} depends only on the $j-i$, in such a case a Markov chain is said to have stationary independent increments, and the Markov chain is called **additive process**. If the sequence $\{X_n\}$ is a sequence of i.i.d Bernoulli random variables with $P[X_n = 1] = p$ and $P[X_n = 0] = q$, then the P-Matrix of $\{S_n\}$ will be

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ - \\ - \\ - \end{array}\begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & - & - & - \end{array} \\ \begin{pmatrix} q & p & 0 & 0 & - & - \\ 0 & q & p & 0 & - & - \\ 0 & 0 & q & p & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \end{pmatrix} \end{array}$$

***

Example 12 (Ehrenfest Model): This example is a special case of a model, called the Ehrenfest model, given by P. and T. Ehrenfest in 1907. It has been used to explain the diffusion of gases. Suppose we have two urns that contain between them four balls. At each step, one of the four balls is chosen at random and moved from its present urn to the other urn. We choose, as states, the number of balls in the first urn. Thus, the set of states are $\{0, 1, 2, 3, 4\}$. The sequence of random variables $\{X_n\}$, the denoting number of balls in the first urn at successive steps is a Markov chain.

$p_{0j} = P(X_n = j \mid X_{n-1} = 0) = 0$, when $j \neq 1$, $p_{0j} = P(X_n = 1 \mid X_{n-1} = 0) = 1$. Since, when the first urn is empty then the chosen ball is certainly from the second urn and it will be transferred to the first urn.

$p_{10} = P(X_n = 0 \mid X_{n-1} = 1) = 1/4$, $p_{12} = P(X_n = 2 \mid X_{n-1} = 1) = 3/4$, $p_{11} = p_{13} = p_{14} = 0$.

Since, when the first urn has one ball, then the chosen ball will be in the first urn with probability, $1/4$, and in the second urn with probability, $3/4$. Similarly, the other transition probabilities can be obtained.

The transition matrix is then

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

***

Example 13: A Markov chain has the following initial distribution $\mathbf{u}$, and P-matrix.

$\mathbf{u} = \{1/3, 1/3, 1/3\}$ and

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

We have

$$\mathbf{P}^2 = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$$

We find the following results

$\mathbf{uP} = (1/3\ \ 1/3\ \ 1/3) = \mathbf{u}$, and

$\mathbf{uP}^2 = (1/3\ \ 1/3\ \ 1/3) = \mathbf{u}$.

We can find the same relation for all the higher powers of $\mathbf{P}$. Therefore, we get, in general, $\mathbf{uP}^n = (1/3\ \ 1/3\ \ 1/3) = \mathbf{u}$ for every n, and thus using **Theorem 6**, we have $\mathbf{u}^{(n)} = \mathbf{u}$ for all n.

With such initial distribution $\mathbf{u}$, the Markov chain will be called a **Stationary Markov chain**. The probability distribution, $\mathbf{u}$, is then called the **Stationary Distribution** of the Markov chain. This type of Markov chain will be discussed in detail in Unit 3.

***

You may now try the following exercises.

E4) In Example 5, let $a = 0$ and $b = 1/2$. Compute $\mathbf{P}$, $\mathbf{P}^2$, and $\mathbf{P}^3$. What will $\mathbf{P}^n$ be? What happens to $\mathbf{P}^n$, as n tends to infinity? Interpret this result.

E5) In Example 6, compute $\mathbf{P}$, $\mathbf{P}^2$ and $\mathbf{P}^3$. What will be $\mathbf{P}^n$?

E6) Compute the matrices $\mathbf{P}^2, \mathbf{P}^3, \mathbf{P}^4$ for the Markov chain defined by the transition matrix $\mathbf{P}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Do the same for the transition matrix $\mathbf{P}=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Interpret the results in each of these processes.

E7) Assume in Exercise 1 that every man has at least one son. Find the probability that a randomly chosen grandson of a businessman is former.

So far, we have been discussing Markov chains, the related probability matrices including initial distributions, and their interpretations. Now, in the next section we shall discuss two important methods of calculating P^n.

2.6 METHODS OF CALCULATING P^n

In this section, we shall discuss the following two methods of evaluating P^n for a given P.

(1) Method of Spectral Decomposition
(2) Method of Generating function.

First, let us understand the method of Spectral Decomposition.

2.6.1 Method of Spectral Decomposition

Let **P** be the transition matrix of a finite order, $s \times s$. Suppose P has distinct **eigen values** (or latent roots, or characteristic roots, or spectral values) $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_s$. They are the roots of the **characteristic equation** $|\mathbf{P} - \lambda \mathbf{I}| = 0$, where $\mathbf{I}$ is $s \times s$ identity matrix.

A non-zero column vector $\mathbf{x}$ is called a **right eigen vector** (or latent, or characteristic vector) of $\mathbf{P}$, corresponding to **eigen value** λ_i if it satisfies the vector equation $(\mathbf{P} - \lambda_i \mathbf{I})\mathbf{x} = \mathbf{0}$. A non-zero row vector $\mathbf{y}'$ is called a **left eigen vector** (or latent, or characteristic vector) of $\mathbf{P}$ corresponding to **eigen value** λ_i, if it satisfies the vector equation $\mathbf{y}'(\mathbf{P} - \lambda_i \mathbf{I}) = \mathbf{0}$. The right and left **eigen vectors** are not unique. For examples, if $\mathbf{x}$ is a **right eigen vector** the $\mathbf{kx}$ is also a **right eigen vector,** $k \neq 0$ is scalar. A similar rule holds for the **left eigen vector.**

Let $\mathbf{x}_i, \mathbf{y}'_i$ be right and left **eigen vectors** corresponding to λ_i, $(i = 1, 2, \ldots, s)$.

Let $c_i = 1/\mathbf{y}'_i\mathbf{x}_i$ and $\mathbf{B}_i = c_i\, \mathbf{x}_i\mathbf{y}'_i$. The product $\mathbf{B}_i$ is a matrix of order $s \times s$, and is called a constituent matrix corresponding to λ_i, $(i = 1, 2, \ldots, s)$.

We have the following properties in the context of constituent matrices:

(i) $\mathbf{B}_i\mathbf{B}_j = \mathbf{0}\, (i \neq j)$, (orthogonal)

(ii) $\mathbf{B}_i^2 = \mathbf{B}_i$ (Idempotent)

(iii) $\sum_{i=1}^{s} \mathbf{B}_i = \mathbf{I}$, where $\mathbf{I}$ is an $s \times s$ identity matrix.

(iv) $\mathbf{P} = \sum_{i=1}^{s} \lambda_i \mathbf{B}_i$, the Spectral Decomposition

In general, we have, the following result, using the above properties

$$\mathbf{P}^n = \left(\sum_{i=1}^{s} \lambda_i \mathbf{B}_i\right)^n = \sum_{i=1}^{s} \lambda_i^n \mathbf{B}_i$$

From this, we can get $p_{ij}^{(n)}$, as (i, j) th element of $\mathbf{P}^n$.

***

Remark 4: (i) Since the row sum equals unity for all the rows of $\mathbf{P}$, therefore one is always an eigen value of $\mathbf{P}$, and the corresponding **right eigen vector** $\mathbf{x}$ has all the elements unity. Therefore, the constituent matrix $\mathbf{B}_1$ corresponding to eigen value one will have all rows identical. It is illustrated below.

$$\mathbf{B}_1 = c_1 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} y_1 & y_2 & \cdots & y_s \end{pmatrix} = c_1 \begin{pmatrix} y_1 & y_2 & \cdots & y_s \\ y_1 & y_2 & \cdots & y_s \\ \vdots & \vdots & & \vdots \\ y_1 & y_2 & \cdots & y_s \end{pmatrix}$$

(ii) All the eigen values of $\mathbf{P}$ are less than or equal to unity in absolute value.

(iii) If the matrix $\mathbf{P}$ is positive and irreducible, then it has only one eigen value equal to unity, while if $\mathbf{P}$ is non-negative, irreducible and cyclic of order h, then it may have $h(\geq 1)$ repeated eigen values equal to unity. .

(iv) If unity is the non-repeated eigen value of $\mathbf{P}$, then $\lim_{n\to\infty} \mathbf{P}^n \to \mathbf{B}_1$, the constituent matrix for eigenvalue 1.

***

Example 14: We will find $\mathbf{P}^n$ for the transition matrix $\mathbf{P}$, given in Example 5.

$$P = \begin{array}{c} \\ 1 \\ 2 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 1 & \quad 2 \end{array} \\ \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \end{array}, \quad (0<a,\, b<1)$$

For eigen values λ of $\mathbf{P}$, the characteristic equation is

$$|\mathbf{P} - \lambda \mathbf{I}| = 0, \text{ or}$$

$$\begin{vmatrix} 1-a-\lambda & a \\ b & 1-b-\lambda \end{vmatrix} = 0 \text{ and solving it, we get}$$

$$\lambda_1 = 1, \quad \lambda_2 = 1-a-b$$

The **right eigen vector** $\mathbf{x}_1$ corresponding to $\lambda_1 = 1$, will satisfy following

$$(\mathbf{P} - \lambda_1 \mathbf{I})\mathbf{x}_1 = \mathbf{0}, \text{ i.e. } \mathbf{P}\mathbf{x}_1 = \mathbf{x}_1$$

Therefore, we have to solve the system of equations

$$\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} \text{ i.e.}$$

$$(1-a)x_{11} + a x_{12} = x_{11}$$
$$b x_{11} + (1-b) x_{12} = x_{12}$$

Which gives $x_{11} = x_{12}$, and therefore $x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Similarly, the **right eigen vector** $\mathbf{x}_2$ corresponding to $\lambda_2 = 1-a-b$ can be obtained by solving following

$$\mathbf{P}\mathbf{x}_2 = \lambda_2 \mathbf{x}_2, \text{ i.e.}$$

$$\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} = (1-a-b) \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} \text{ i.e.}$$

$$(1-a)x_{21} + a\,x_{22} = (1-a-b)\,x_{21} \text{ i.e.}$$
$$bx_{21} + (1-b)\,x_{22} = (1-a-b)\,x_{22}\,.$$

Solving this, we get

$$bx_{21} = -a\,x_{22} \text{ and thus, } x_2 = \begin{pmatrix} a \\ -b \end{pmatrix}$$

The left **eigen vector** $\mathbf{y}_1'$, corresponding to $\lambda_1 = 1$, will be obtained by solving

$$\mathbf{y}_1'(\mathbf{P} - \lambda_1\mathbf{I}) = \mathbf{0}\text{, which reduces to}$$
$$(1-a)y_{11} + by_{12} = y_{11}$$
$$ay_{11} + (1-b)y_{12} = y_{12}$$

This gives

$$ay_{11} = by_{12}$$

and so, we get

$$\mathbf{y}_1' = (b \quad a)$$

Similarly, we get left **eigen vector** $\mathbf{y}_2'$, corresponding to $\lambda_2 = 1-a-b$, as

$$\mathbf{y}_2' = (1 \quad -1)$$

and, therefore,

$$1/c_1 = \mathbf{y}_1'\mathbf{x}_1 = (b\ a)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = a+b,\ 1/c_2 = \mathbf{y}_2'\mathbf{x}_2 = (1\ -1)\begin{pmatrix} a \\ -b \end{pmatrix} = a+b$$

Next, we compute constituent matrices

$$\mathbf{B}_1 = c_1\mathbf{x}_1\mathbf{y}_1' = \frac{1}{(a+b)}\begin{pmatrix} 1 \\ 1 \end{pmatrix}(b \quad a) = \frac{1}{(a+b)}\begin{pmatrix} b & a \\ b & a \end{pmatrix}$$

$$\mathbf{B}_2 = c_2\mathbf{x}_2\mathbf{y}_2' = \frac{1}{(a+b)}\begin{pmatrix} a \\ -b \end{pmatrix}(1 \quad -1) = \frac{1}{(a+b)}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix}$$

And thus,

$$\mathbf{P}^{(n)} = \mathbf{P}^n = \sum_{i=1}^{2}\lambda_i^n\,\mathbf{B}_i = \frac{1}{(a+b)}\begin{pmatrix} b & a \\ b & a \end{pmatrix} + \frac{(1-a-b)^n}{(a+b)}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix}.$$

Since, $0 < |1-a-b| < 1 \Rightarrow (1-a-b)^n \to 0$ as $n \to \infty$ and thus, $\mathbf{P}^n \to \frac{1}{(a+b)}\begin{pmatrix} b & a \\ b & a \end{pmatrix}$

We also get $p_{ij}^{(n)}$, the probability of transition from state i to j in n-step, as the $(i, j)^{th}$ element of $\mathbf{P}^n$. Therefore,

$$p_{11}^{(n)} = \frac{b + a(1-a-b)^n}{(a+b)},\quad p_{12}^{(n)} = \frac{a - a(1-a-b)^n}{(a+b)},$$

$$p_{21}^{(n)} = \frac{b - b(1-a-b)^n}{(a+b)},\quad p_{22}^{(n)} = \frac{a + b(1-a-b)^n}{(a+b)}$$

As $n \to \infty$,

$$p_{11}^{(n)} \to b/(a+b),\qquad p_{12}^{(n)} \to a/(a+b),$$
$$p_{21}^{(n)} \to b/(a+b),\qquad p_{22}^{(n)} \to a/(a+b)$$

Let the initial distribution be $\mathbf{u} = (p \quad 1-p)$. Then, the unconditional probability distribution of X_n is

$$\mathbf{u}^{(n)} = \mathbf{u}\mathbf{P}^n$$

$$= (p \quad 1-p)\left(\frac{1}{(a+b)}\begin{pmatrix} b & a \\ b & a \end{pmatrix} + \frac{(1-a-b)^n}{(a+b)}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix}\right)$$

$$= \frac{1}{(a+b)}(b \quad a) + \frac{(1-a-b)^n}{(a+b)}(ap - bq \quad -ap + bq)$$

and hence,

$$u_1^{(n)} = \frac{b}{(a+b)} + \frac{(ap-bq)(1-a-b)^n}{(a+b)}$$

$$u_2^{(n)} = \frac{a}{(a+b)} + \frac{(-ap+bq)(1-a-b)^n}{(a+b)}, \text{ where, } q = 1-p$$

***

Example 15: Three girls, A, B, and C stand in a circle to play a ball throwing game. Each one can throw the ball to one of her two neighbours, each with probability 0.5. The sequence of random variables $\{X_n\}$, where X_n denotes the player with whom the ball will lie at n^{th} throw will form a Markov chain. The Markov chain will have following P-matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$

It is doubly stochastic as all the row sums and column sums are unity. Therefore, corresponding to eigen value $\lambda_1 = 1$, the left eigen vector $\mathbf{y}_1'$ and right eigen vector $\mathbf{x}_1$, both will have all the elements as one only. Thus, the constituent matrix $\mathbf{B}_1$ will have all rows identical and all columns identical (Bhat, 2000, 109p.). Therefore, all the elements of $\mathbf{B}_1$ will be identical, and each will be equal to $s^{-1} = 1/3$. Thus,

$$\mathbf{P}^n \to \mathbf{B}_1 = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \text{ as } n \to \infty.$$

***

Let us discuss the second method of calculating P^n which is known as the method of generating function.

2.6.2 Method of Generating Function

As the name suggests in this method, a function is determined which generates the P^n for different values of n.

Define the generating function

$$P(s) = \mathbf{I} + s\mathbf{P} + s^2\mathbf{P}^2 + s^3\mathbf{P}^3 + \cdots + s^n\mathbf{P}^n + \cdots \text{ where } |s| < 1$$

(Here, s is a variable of the function $P(s)$, and not the size of state space as before.)

Since, as $n \to \infty$, $s^n\mathbf{P}^n \to 0$, therefore, $P(s) = (\mathbf{I} - s\mathbf{P})^{-1}$, the inverse of matrix $(\mathbf{I} - s\mathbf{P})$.

Thus, we may obtain $\mathbf{P}^n$ by extracting the coefficient of s^n in the expansion $(\mathbf{I} - s\mathbf{P})^{-1}$, and $P_{ij}^{(n)}$ as the $(i, j)^{th}$ component of $\mathbf{P}^n$.

Example 16: Let us find $\mathbf{P}^n$, where $\mathbf{P}$ the transition matrix is given below:

$$\mathbf{P} = \begin{pmatrix} q & p & 0 & 0 \\ 0 & q & p & 0 \\ 0 & 0 & q & p \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ where } q = 1-p, \text{ and } 0 < p < 1$$

Here,

$$\mathbf{I} - s\mathbf{P} = \begin{pmatrix} 1-sq & -sp & 0 & 0 \\ 0 & 1-sq & -sp & 0 \\ 0 & 0 & 1-sq & -sp \\ 0 & 0 & 0 & 1-s \end{pmatrix} \text{ and thus,}$$

Now, as $|\mathbf{I}-s\mathbf{P}|=(1-s)\,(1-sq)^3$, we have

$$(\mathbf{I}-s\mathbf{P})^{-1}=\frac{\mathrm{adj}(\mathbf{I}-s\mathbf{P})}{|\mathbf{I}-s\mathbf{P})}$$

$$=\frac{1}{(1-s)(1-sq)^3}\begin{pmatrix}(1-s)(1-sq)^2 & sp(1-s)\,(1-sq) & s^2P^2(1-s) & s^3p^3\\ 0 & (1-s)(1-sq)^2 & sp(1-s)(1-sq) & s^2p^2(1-sq)\\ 0 & 0 & (1-s)(1-sq)^2 & sp(1-sq)^2\\ 0 & 0 & 0 & (1-sq)^3\end{pmatrix}$$

$$=\begin{pmatrix}1-sq)^{-1} & sp(q-sq)^{-2} & s^2p^2(1-sq)^{-3} & s^3p^3(1-s)^{-1}(1-sq)^{-3}\\ 0 & (1-sq)^{-1} & sp(1-sq)^2 & s^2p^2(1-s)^{-1}(1-sq)^{-2}\\ 0 & 0 & (1-sq)^{-1} & sp(1-s)^{-1}(1-sq)^{-1}\\ 0 & 0 & 0 & (1-s)^{-1}\end{pmatrix}$$

To obtain $\mathbf{P}^n$, we collect coefficients of s^n by expanding each element of $(\mathbf{I}-s\mathbf{P})^{-1}$ in powers of s, as below.

$$\mathbf{P}^n=\begin{pmatrix}q^n & npq^{n-1} & \frac{n(n-1)}{2}p^2q^{n-2} & p^3(1+3q+6q^2+\ldots+\frac{(n-1)(n-2)}{2}q^{n-3})\\ 0 & q^n & npq^{n-1} & p^2(1+2q+3q^2+\ldots(n-1)q^{n-2})\\ 0 & 0 & q^n & p(1+q+q^2+\ldots.+q^{n-1})\\ 0 & 0 & 0 & 1\end{pmatrix}$$

Simplifying,

$$\mathbf{P}^n=\begin{pmatrix}q^n & npq^{n-1} & \frac{n(n-1)}{2}p^2q^{n-2} & (1-\frac{n(n-1)}{2}p^2q^{n-2}-npq^{n-1}-q^n)\\ 0 & q^n & npq^{n-1} & (1-npq^{n-1}-q^n)\\ 0 & 0 & q^n & (1-q^n)\\ 0 & 0 & 0 & 1\end{pmatrix}$$

***

You may now try the following exercises.

E8) Let state space of a Markov chain be $S=(0,1,2)$, and its P-Matrix is

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

Obtain $\mathbf{P}^n$.

E9) Find $\mathbf{P}^n$ for large n for the matrix given below.

$$\mathbf{P}=\begin{pmatrix}0.5 & 0.3 & 0.2\\ 0.2 & 0.4 & 0.4\\ 0.1 & 0.5 & 0.4\end{pmatrix}$$

E10) Find $\mathbf{P}^n$ and its limiting value for large n for the matrix given below.

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p_1 & p_2 & p_3 \end{pmatrix}, \text{ where } p_1, p_2, p_3 \geq 0 \text{ and } \sum p_i = 1$$

E11) (**Gene Model**) The simplest type of inheritance of a trait in animals is governed by a pair of genes, each of which may be of two types, say G and g. An individual may have either a combination GG, or Gg (which is genetically the same as gG), or gg. Very often, the GG and Gg types are indistinguishable in appearance, and then we say that the G gene dominates the g gene. An individual is called *dominant* if he or she has GG genes, *recessive* if he or she has gg, and *hybrid* if a Gg mixture is present. Consider a process of continued matings. We start with an individual of known genetic character and mate it with a hybrid. We assume that there is at least one offspring. An offpring is chosen at random and is mated with a hybrid, and this process repeated through a number of generations. The genetic type of the chosen offpring in successive generations forms a Markov chain with states, dominant GG, hybrid Gg, and recessive gg, represented by 1, 2, and 3 respectively. The transition probability matrix is

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{pmatrix}. \text{ Find } \mathbf{P}^n \text{ and its limit for large n.}$$

Now we bring this unit to a close. But before that let's briefly recall the important concepts that we studied in it.

2.7 SUMMARY

In this unit, we have tried to acquaint you with the basic features of a stochastic process, and Markov chains. We are summarizing these below:

1. We introduced the idea of Stochastic process and presented their classification according to the nature of time and state space. The Markov chain was explained as a particular case of the Stochastic process.
2. We defined the Markov property and the Markov chain, and presented some examples suitable to a Markov model.
3. We studied properties of transition probabilities, and the transition matrix.
4. We described how the one-step transition in a Markov chain can be represented as a digraph.
5. We have acquainted you with the concept of higher order transition probabilities.
6. We have defined initial distribution, and illustrated the method of computing unconditional probability distribution of states of Markov chain at n^{th} step in terms of transition matrix, and the initial distribution.
7. We have described the method of spectral decomposition and generating functions to compute $\mathbf{P}^n$.

2.8 SOLUTIONS/ANSWERS

E1) Let the three states, business, agriculture and public servant be denoted by 1, 2, 3, respectively. Let the random variable X_n denote the choice of profession of the sons in n^{th} generation. Let P_{ij} denote the probability that given a person is in i^{th} state (profession) his son will choose j^{th} state (profession). Therefore, we get the following P-Matrix

$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.25 & 0.25 & 0.50 \end{pmatrix}$$

E2)

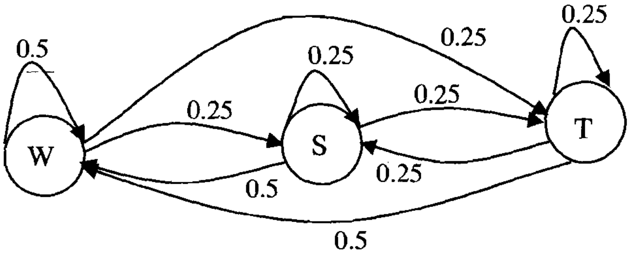

Fig.4

E3) Assume that the states nursery, class one, class two, …, class 5, are denoted by numbers $0, 1, 2, \ldots, 5$. The P-Matrix will be

$$\mathbf{P} = \begin{pmatrix} q_0 & p_0 & 0 & 0 & 0 & 0 \\ 0 & q_1 & p_1 & 0 & 0 & 0 \\ 0 & 0 & q_2 & p_2 & 0 & 0 \\ 0 & 0 & 0 & q_3 & p_3 & 0 \\ 0 & 0 & 0 & 0 & q_4 & p_4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{where } q_i + p_i = 1$$

E4) $\mathbf{P} = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1 & 0 \\ 0.5 & 0.5 \end{pmatrix} \end{array}$, $\mathbf{P}^2 = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1 & 0 \\ 0.75 & 0.25 \end{pmatrix} \end{array}$, $\mathbf{P}^3 = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1 & 0 \\ 0.875 & 0.125 \end{pmatrix} \end{array}$

Using, **Example 13**, we may get following result by putting $a = 0, b = 0.5$ in the expression of $\mathbf{P}^n$

$$\mathbf{P}^n = \frac{1}{(0.5)}\begin{pmatrix} 0.5 & 0 \\ 0.5 & 0 \end{pmatrix} + \frac{(0.5)^n}{(0.5)}\begin{pmatrix} 0 & 0 \\ -0.5 & 0.5 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + (0.5)^n \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix}$$

and as n is large $\mathbf{P}^n \to \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$

E5) $\mathbf{P}^2 = \mathbf{P}^3 = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$, $\quad \mathbf{P}^n = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$

E6) When $\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ then $\mathbf{P}^2 = \mathbf{P}^3 = \mathbf{P}^4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and in this case $\mathbf{P}^n = \mathbf{P}$

and when $\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ then $\mathbf{P}^2 = \mathbf{P}^4 = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}$ and $\mathbf{P}^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

In this case $\mathbf{P}^n = \begin{cases} \mathbf{P} & \text{when n is odd} \\ \mathbf{I} & \text{when n is even} \end{cases}$

E7) We want the probability $P_{12}^{(2)}$. Since,

$$\mathbf{P}^2 = \begin{pmatrix} 0.685 & 0.190 & 0.125 \\ 0.330 & 0.480 & 0.190 \\ 0.363 & 0.450 & 0.188 \end{pmatrix}$$

Therefore $P_{12}^{(2)} = 0.19$.

E8) Using the method of generating function

$$\mathbf{I} - s\mathbf{P} = \begin{pmatrix} 1 & -s & 0 \\ 0 & 1 & -s \\ -s & 0 & 1 \end{pmatrix} \text{ and } |\mathbf{I} - s\mathbf{P}| = 1 - s^3 \neq 0, \text{ for } |s| < 1$$

$$(\mathbf{I} - s\mathbf{P})^{-1} = \frac{\text{adj}(\mathbf{I} - s\mathbf{P})}{1 - s^3} = \frac{1}{1 - s^3} \begin{pmatrix} 1 & s & s^2 \\ s^2 & 1 & s \\ s & s^2 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} (1-s^3)^{-1} & s(1-s^3)^{-1} & s^2(1-s^3)^{-1} \\ s^2(1-s^3)^{-1} & (1-s^3)^{-1} & s(1-s^3)^{-1} \\ s(1-s^3)^{-1} & s^2(1-s^3)^{-1} & (1-s^3)^{-1} \end{pmatrix}$$

We can get the following coefficients easily since $(1-s^3)^{-1}$ has only powers of s^3,

$$\mathbf{P}^{3n} = \text{Coefficient of } s^{3n} \text{ in } (\mathbf{I} - s\mathbf{P})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{P}^{3n+1} = \text{Coefficient of } s^{1+3n} \text{ in } (\mathbf{I} - s\mathbf{P})^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{P}^{3n+2} = \text{Coefficient of } s^{2+3n} \text{ in } (\mathbf{I} - s\mathbf{P})^{-1} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

where n is a non-negative integer.

E9) The P-matrix is

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}$$

the eigen values of the P are 1, 0.1 and 0.2.

For the eigen value $\lambda_1 = 1$, the right eigen vector is $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, and the left eigen vector $\mathbf{y}_1'$, is the solution of $\mathbf{y}_1'(\mathbf{P} - \lambda_1 \mathbf{I}) = \mathbf{0}$.

Equivalently, the solution of

$$0.5y_{11} + 0.2y_{12} + 0.1y_{13} = y_{11}$$
$$0.3y_{11} + 0.4y_{12} + 0.5y_{13} = y_{12}$$
$$0.2y_{11} + 0.4y_{12} + 0.4y_{13} = y_{13}$$

This gives

$\mathbf{y}_1' = (0.16 \quad 0.28 \quad 0.24)$ and thus

$1/c_1 = \mathbf{y}_1' \mathbf{x}_1 = 0.68$

$$\mathbf{B}_1 = c_1 \mathbf{x}_1 \mathbf{y}_1'$$

$$= 0.68 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (0.16 \ 0.28 \ 0.24) = \begin{pmatrix} 0.235 & 0.412 & 0.353 \\ 0.235 & 0.412 & 0.353 \\ 0.235 & 0.412 & 0.353 \end{pmatrix}$$

Since for large n, $\mathbf{P}^n \to \mathbf{B}_1$, therefore we get the result.

E10) Using the method of generating function

$$\mathbf{I} - s\mathbf{P} = \begin{pmatrix} 1-s & 0 & 0 \\ 0 & 1-s & 0 \\ -p_1 s & -p_2 s & 1-p_3 s \end{pmatrix},$$

$|\mathbf{I} - s\mathbf{P}| = (1-s)^2 (1-p_3 s) \neq 0$, for $|s| < 1$,

$$(\mathbf{I} - s\mathbf{P})^{-1} = \begin{pmatrix} (1-s)^{-1} & 0 & 0 \\ 0 & (1-s)^{-1} & 0 \\ p_1 s(1-s)^{-1}(1-p_3 s)^{-1} & p_2 s(1-s)^{-1}(1-p_3 s)^{-1} & (1-p_3 s)^{-1} \end{pmatrix}$$

$\mathbf{P}^n$ = coefficient of s^n in $(\mathbf{I} - s\mathbf{P})^{-1}$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p_1 \sum_{r=0}^{n-1} p_3^{\,r} & p_2 \sum_{r=0}^{n-1} p_3^{\,r} & p_3^{\,n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \dfrac{p_1(1-p_3^{\,n})}{1-p_3} & \dfrac{p_2(1-p_3^{\,n})}{1-p_3} & p_3^{\,n} \end{pmatrix}$$

As n becomes large,

$$\mathbf{P}^n \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p_1/(1-p_3) & p_2/(1-p_3) & 0 \end{pmatrix}$$

E11) Solving $|\mathbf{P} - \lambda \mathbf{I}| = 0$, we may get eigen values for $\mathbf{P}$ as $\lambda_1 = 1, \lambda_2 = 0.5, \lambda_3 = 0$

For the eigen value $\lambda_1 = 1$, the right eigen vector $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, while the left eigen vector $\mathbf{y}_1'$, will be obtained by solving $\mathbf{y}_1'(\mathbf{P} - \lambda_1 \mathbf{I}) = \mathbf{0}$. This gives

$\mathbf{y}_1' = (2 \quad 4 \quad 2)$ and thus,

$1/c_1 = \mathbf{y}_1' \mathbf{x}_1 = 8$

$$\mathbf{B}_1 = c_1 \mathbf{x}_1 \mathbf{y}_1' = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.5 & 0.25 \end{pmatrix}$$

For the eigen value 0.5, the right eigen vector $\mathbf{x}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$, and the left eigen vector, will be

$\mathbf{y}'_2 = (-1 \;\; 0 \;\; 1)$ and thus

$1/c_2 = \mathbf{y}'_2\mathbf{x}_2 = 2$. Therefore,

$$\mathbf{B}_2 = c_2\mathbf{x}_2\mathbf{y}'_2 = (0.5)\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

Thus, we have

$$\mathbf{P}^n = \sum_{i=1}^{3} \lambda_i{}^n \mathbf{B}_i = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.5 & 0.25 \end{pmatrix} + (0.5)^{n+1}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

As $n \to \infty$, $\mathbf{P}^n \to \mathbf{B}_1$ therefore, the matrix $\begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.5 & 0.25 \end{pmatrix}$ will be the limiting value of $\mathbf{P}^n$.

—x—