UNIT 1 CONDITIONAL PROBABILITY

Structure

Page No.

1.1 INTRODUCTION

This unit introduces you to the pre-requisites of probability and statistics, which you studied as undergraduates. We recall the concepts of conditional probability, compound probability, Bayes' theorem, conditional distribution, and conditional expectations here. These are fundamental to the study of probability and statistics.

The history of probability can be traced back to the beginning of the mankind in the games of chance. Archaeologists have found evidence of games of chance in prehistoric digs, showing that gaming and gambling have been a major pastime for the peoples in Greece, Egypt, China, and India since the dawn of civilization. However, it wasn't until the 17^{th} century that a rigorous mathematics of probability was developed by French mathematicians Pierre de Fermat and Blaise Pascal. The basic concept of conditional probability and the famous Bayes' theorem was the pioneer work of Thomas Bayes (1707–1761). However, it was Laplace, who generalized, completed and consummated the ideas provided by his predecessors in his book *Théorie analytique des probabilités* in 1812. It gave a comprehensive system of probability theory (The elements of probability calculus – addition, multiplication, division – were by that time firmly established.).

We shall start our discussion with conditional probability in Sec.1.2. Here, we present its concept and definition along with some examples. In Sec.1.3, we learn the Compound Probability Law. In Sec.1.4, we recall the Law of Total Probability along with the very widely used Bayes' theorem. In Sec.1.5, we discuss the conditional distribution. Finally we conclude with defining the conditional expectation of the random variables, which is very important, and give some examples of it.

Objectives

After studying this unit, you should be able to:

- define and compute the conditional probability of an event;
- distinguish between the conditional and unconditional probability of an event;
- evaluate the change in the probability of an event after the occurrence of another event;
- apply the Bayes' theorem in different situations;
- apply the concept of conditional distribution and conditional expectation and their important properties in various problems.

1.2 CONDITIONAL PROBABILITY

Let us start with an example to understand the concept of conditional probability. A teacher gave two tests in succession to the students of a class. 75% Students of the class passed the first test, 35% of the class passed the second test and 15% of the class passed both tests. We may want to find, "What percent of the students, who passed the first test passed the second, who passed the second test, not the first test?" Through this example we shall illustrate the concept of conditional probability. Suppose we want to find the probability of the event that the second test will be passed by a student given that he/she had passed first test. Here, it is given that 35% of the students passed the second test and therefore, the probability that a student of the class will pass the second test will be 0.35. It is an unconditional probability of the event that a student passes second test. When we are given prior information that a student has passed the first test and we want to know the probability of the same event – the student passes the second test – under this condition, then the probability will not remain the same. Since out of 75%, of the students who passed the first test only 15% could pass the second test, therefore, $\frac{15\text{x}100}{75} = 20\%$ students of those students who passed the first test passed the second test. It gives the conditional probability of the event that a student passes the second test given that he/she has passed the first test, and the probability will be equal to 0.20.

This example can also be shown clearly by a Venn diagram as depicted in Fig.1.

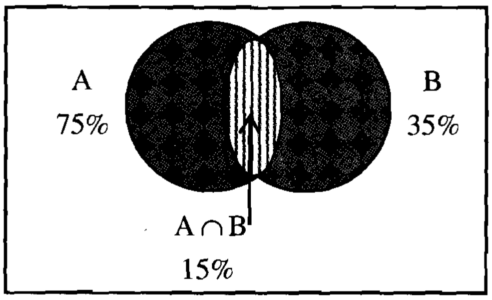

Fig.1

Let the event $A =$ a student who passed the first test, and the event $B =$ a student who passed the second test; clearly, the event $A \cap B =$ a student who passed both the tests.

It is given that $P(B) = 0.35$, which is the probability of the event B without any additional condition, i.e. inconditional probability which means that the probability is evaluated considering a full class of students as the sample space. If we want to find the probability of the event that, has a student passed the second test, who has already passed the first test, then our sample space reduces to the set of those students who have passed the first test, i.e. the event A. This probability is evaluated as the ratio of the probability of B included in A (which is $P(A \cap B)$) to the probability of A. This ratio comes out to be 0.20 or 20% as evaluated above. In this case, this probability will be termed as the conditional probability of event B (passed second test), given that the event A (passed the first test) has happened.

This discussion enables us to introduce the following formal definitions. In what follows we assume that we are given a random experiment with discrete sample space S, and all relevant events are subsets of S.

Definition 1: Let S be a sample space of an experiment. Let A and B be any two events defined on this sample space with $P(B) > 0$ (not allowing an event of probability zero). The conditional probability of an event A, given that the other event B has happened, is denoted by the symbol $P(A \mid B)$, and is read as "the probability of A, given B" and this conditional probability, $P(A \mid B)$, is defined as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}; \qquad P(B) > 0 \qquad (1)$$

Similarly, we may define conditional probability of B given A as:

$$P(B|A)=\frac{P(A\cap B)}{P(A)}; \qquad P(A)>0 \qquad (2)$$

Conditional probability $P(A\,|\,B)$ is a set function defined on the subsets of event B. It can easily be verified that $P(A\,|\,B)$ satisfies all the axioms of Probability.

The probability, which is not conditional, is said to be unconditional probability or probability. Also, there does not exist any ordinal relationship between conditional and unconditional probabilities. Depending on the size of the new sample space (here, we denote it as B) under the condition, and the size of $A\cap B$ the conditional probability $P(A\,|\,B)$ may be smaller or larger than its unconditional probability $P(A)$. Now, let us look at some important properties of conditional probability.

Properties of Conditional Probability

From the above definition we may easily verify the following intuitive results for any three events A, B, and C of a sample space, S. Let us discuss a few properties of conditional probability.

1. $P(A\,|\,A)=1$, which is the conditional probability of the reduced sample space itself. $P(A\,|\,A)$, the probability of event A when A has happened, is clearly 1.
   Also $P(A\,|\,A)=\frac{P(A\cap A)}{P(A)}$ (from (1)) i.e. $P(A\,|\,A)=1$. This is the axiom of normedness of probability measure.
2. $P(A|B)\geq 0$. Since numerator and denominator both in Eqn.(1) of conditional probability are non-negative, it is non-negativity axiom of probability.
3. $P(A|B)\leq 1$. Since in Eqn.(1) the event involved in the numerator, $A\cap B$ is always a subset to the event involved in the denominator B. Hence, from the monotone property of probability the result follows.
4. $P(A\,|\,B)=0$; if the events A and B are mutually exclusive. If the events A and B are mutually exclusive, then $A\cap B$ will be empty. Then the numerator in the Eqn.(1) will be zero. Therefore, the result follows.
5. $P(\phi\,|\,B)=0$
6. $P(S\,|\,B)=1$
7. $P(A^C\,|\,B)=1-P(A\,|\,B)$

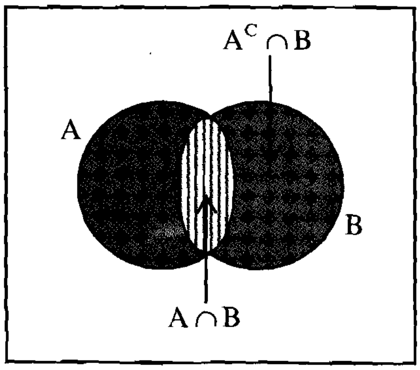

Fig.2

Proof: From the set theory, we have

$$B=(A\cap B)U(A^C\cap B).$$

which is also shown in the Venn diagram in Fig.2.

Since $(A \cap B)$ and $(A^C \cap B)$ are two disjoint events. Therefore using the addition law of probability for disjoint events, we get:

$$P(B) = P(A \cap B) + P(A^C \cap B)$$

Dividing both sides by $P(B)$, we get

$$1=\frac{P(A \cap B)}{P(B)}+\frac{P(A^C \cap B)}{P(B)} \text{ and}$$

$1=P(A \mid B)+P(A^C \mid B)$ and we get the result.

8. $P(A \cup B|C)=P(A|C)+P(B|C)-P(A \cap B|C)$ where $P(C)>0$. It is parallel to the addition law of probability.

Proof: From the distributive law of set theory for the three sets A, B, and C, we know that:

$$(A \cup B) \cap C=(A \cap C) \cup (B \cap C).$$

Using the addition law of the probability of events, we get:

$$P((A \cup B) \cap C)=P(A \cap C) + P(B \cap C)-P(A \cap B \cap C).$$

Dividing both sides by $P(C)$, we have:

$$\frac{P((A \cup B) \cap C)}{P(C)}=\frac{P(A \cap C)}{P(C)}+\frac{P(B \cap C)}{P(C)}-\frac{P(A \cap B \cap C)}{P(C)}$$

and, hence, we get the result.

In terms of mathematical or classical interpretation of the probability, in case all outcomes in S are considered equally likely, the conditional probability of an event A, given the event B, $P(A \mid B)$, can also be defined as:

$$P(A \mid B)=\frac{n(A \cap B)}{n(B)};\ n(B) \neq 0 \tag{3}$$

where $n(B)$ denotes the number of outcomes favorable to event B, and similarly $n(A \cap B)$ denotes the number of outcomes favorable to the event $A \cap B$.

Let us consider some examples to illustrate the concept of conditional probability.

Example 1: Suppose that A and B are two events in an experiment with $P(A)=1/3$, $P(B)=1/4$, and $P(A \cap B)=1/10$. Find each of the following:

a) $P(A \mid B)$
b) $P(B \mid A)$
c) $P(A^c \mid B)$
d) $P(A \mid B^c)$
e) $P(A^c \mid B^c)$

Solution: a) By definition, we have $P(A|B)=\dfrac{P(A \cap B)}{P(B)}=\dfrac{1/10}{1/4}=4/10=0.4$

b) Similarly, we have $P(B|A)=\dfrac{P(A \cap B)}{P(A)}=\dfrac{1/10}{1/3}=3/10=0.3$

c) From Property 7 in this section, we have:
$P(A^c \mid B)=1-P(A \mid B)=1-0.4=0.6$

d) For any two events A and B, we know $P(A)=P(A \cap B)+P(A \cap B^C)$

thus $P(A \cap B^c) = 1/3 - 1/10 = 7/30$, and, therefore:

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{7/30}{1-P(B)} = \frac{7/30}{1-1/4} = 0.31$$

e) $P(A^c \mid B^c) = 1 - P(A \mid B^c) = 1 - 0.31 = 0.69$

***

Example 2: In a survey, the question, "Do you smoke?" was asked to 100 people. Results are shown in the following table:

	Yes (A)	No (A^C)	Total
Male (B)	19	41	**60**
Female (B^C)	12	28	**40**
Total	**31**	**69**	**100**

An individual is chosen from them at random. Find:

a) What is the probability that the selected individual smokes?

b) What is the probability that the selected individual is a male and smokes?

c) What is the probability that the selected individual is a male?

d) What is the probability of selected individual being a smoker if he was found to be a male?

Solution: Define event $A =$ an individual who smokes, and $B =$ an individual is male

a) We want to find the probability that the selected individual smokes, $P(A) = 31/100 = 0.31$.

b) Here, we want to obtain the probability that the selected individual a male and he smokes, which is $P(A \cap B) = 19/100 = 0.19$

c) The probability that the selected individual is a male is $P(B)$ which is $\frac{60}{100} = 0.60$.

d) Here, we want to find the conditional probability of a selected individual smoking given that he is a male: $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.19}{0.60} = \frac{19}{60}$.

***

In the example given above, it may be noted that, the difference in part b and d is that we evaluate $P(A \cap B)$ when the simultaneous occurrence of both events A and B is required, whereas, we evaluate $P(A \mid B)$ when the chances of occurrence of event A from event B is required as it is the conditional probability.

Example 3: In a card game, suppose a player wants to draw two cards of the same suit in order to win. Out of a total of 52 cards, there are 13 cards in each suit. Suppose at first draw, the player draws a diamond. Now, the player wishes to draw a second diamond to win. What is the probability of his winning?

Solution: Let the event A denotes getting a diamond at the first draw, and event B denotes getting a diamond at the second draw. Clearly, we have to find the conditional probability of B, given A, $P(B|A) = \frac{P(A \cap B)}{P(A)}$

Here, $P(A) = 13/52 = 1/4$ and $P(A \cap B) = {}^{13}C_2 / {}^{52}C_2 = \frac{13 \times 12}{52 \times 51} = \frac{1}{17}$

Thus, $P(B|A)=\frac{1/17}{14}=4/17$.

We may arrive at this result by reducing the sample space under the condition and by getting the outcomes favorable to picking a diamond in the reduced space. At the time of the second draw, one diamond has already been chosen, and there are only 12 diamonds remaining in a deck of remaining 51 cards. Thus, the total number of possible outcomes will be 51, and the outcomes favorable to picking a diamond will be 12. Thus, $P(B|A)=12/51$.

***

You may now try the following exercises.

E1) Suppose that A and B are events in a random experiment with $P(B)>0$. Prove each of the following:
  a) If $B \subseteq A$, then $P(A|B)=1$.
  b) If $B \subseteq A$, then $P(A|B)=P(A)/P(B)$.
  c) If A and B are disjoint, then $P(A|B)=0$.

E2) Suppose that A and B are events in a random experiment, each having positive probability. Show that:
  a) $P(A|B)>P(A) \Leftrightarrow P(B|A)>P(B) \Leftrightarrow P(A \cap B)>P(A)P(B)$.
  b) $P(A|B)<P(A) \Leftrightarrow P(B|A)<P(B) \Leftrightarrow P(A \cap B)<P(A)P(B)$.
  c) $P(A|B)=P(A) \Leftrightarrow P(B|A)=P(B) \Leftrightarrow P(A \cap B)=P(A)P(B)$.

E3) The probability that it is Friday and that a student is absent is 0.03. There are 6 school days in a week. What is the probability that a student is absent given that today is Friday?

E4) Suppose that a bag contains 12 coins of which 5 are fair, 4 are biased, each with the probability of heads being $1/3$; and 3 are two-headed. A coin is chosen at random from the bag and tossed.
  a) Find the probability that the coin is biased.
  b) Find the probability that the biased coin was selected and the coin lands showing a head.
  c) Given that the coin is biased, find the conditional probability of getting a head.

In the next section, we shall discuss the concept of compound probability.

1.3 COMPOUND PROBABILITY

From the definition of conditional probability given in Eqn. (1), we can easily derive the following multiplication rule by cross multiplication

$$P(A \cap B)=P(A)P(B|A) \tag{4}$$

and, similarly from Eqn. (2), we get

$$P(A \cap B)=P(B)P(A|B) \tag{5}$$

Now we want to extend it for three events. Let A, B and C be three events, then we write

$$P(A \cap B \cap C)=P(A)\frac{P(A \cap B)}{P(A)}\frac{P(A \cap B \cap C)}{P(A \cap B)}$$

$$=P(A)P(B|A)P(C|A \cap B). \tag{6}$$

The above multiplication rule for three events can easily be extended using induction for n events $A_1, A_2, \ldots, A_n$ belonging to a sample space as follows:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2)\ldots$$
$$\ldots P(A_n | A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \qquad (7)$$

These relations are called compound probability law or multiplication law. This rule is applied to find the probability of the concurrent occurrence of two or more events using conditional probability as illustrated in the following example.

Example 4: A bag contains 5 white balls, and 4 black balls. Two balls are drawn from the bag randomly, one by one, without replacement. Find the probability that the first ball is black, and second is white.

Solution: Let events $A =$ first ball is black, and $B =$ second ball is white. Clearly, we have to find out $P(A \cap B)$.

Since $P(A) = 4/9$ and $P(B | A) = 5/8$ (Under the condition that A has happened, the reduced sample space has a total of 8 outcomes, out of which 5 are favorable to B.)

Thus, using the multiplication law, we get:

$$P(A \cap B) = P(A)P(B | A) = \frac{4}{9} \cdot \frac{5}{8} = \frac{5}{18}$$

***

Example 5: In a production process, three units are selected randomly without replacement from lots of 100 units for inspection for quality control. If all three selected units are found defective then the lot is rejected, otherwise it is accepted. If a lot contains 15 defective items, then find the probability that this lot will be:

a) rejected

b) accepted.

Solution: Let the event A be that the first selected unit is defective. Event B be equals that the second selected unit is defective, and event C is that the third selected unit is defective.

a) The lot is rejected if all three units are found defective. Thus, we need to obtain $P(A \cap B \cap C)$.
At the first draw, $P(A) = 15/100$, getting a defective unit from 100 units containing 15 defectives units.
In the second draw, assuming event A, the lot now contains 99 units with 14 defectives units. Thus $P(B | A) = 14/99$.
Similarly, $P(C | A \cap B)$ is the probability of getting a defective unit in the third draw, given that in both the earlier draws there were defective units. Thus, the lot now contains 98 units with 13 of which are defectives. Therefore, $P(C | A \cap B) = 13/98$.
Therefore, using the law of compound probability, we get:
$$P(A \cap B \cap C) = P(A)P(B|A)P(C | A \cap B)$$
$$= \frac{15}{100} \cdot \frac{14}{99} \cdot \frac{13}{98} = \frac{13}{4620}.$$

b) The lot will be accepted if it is not rejected. Clearly, the probability that the lot will be accepted, $1 - P(A \cap B \cap C)$, which is $1 - \frac{13}{4620} = \frac{4604}{4620}$ or $\frac{1151}{1144}$.

***

You may now try some exercises.

E5) A box contains 8 balls. Three of them are red and the remaining 5 are blue. Two balls are drawn successively, at random and without replacement. Find the probability that, the first draw results in red, and the second draw results in blue.

E6) In a certain population, 30% of the persons smoke, and 8% have a certain type of heart disease. Moreover, 12% of the persons who smoke have the heart disease.

a. What percentage of the population smoke and have the heart disease?

b. What percentage of the population with the heart disease smoke?

E7) Consider the experiment that consists of rolling two fair dice. Let X denotes, the score of the first die, and Y denotes the sum of the scores on both the dice.

a. Find the probability that $X = 3$, and $Y = 8$.

b. Find the probability that $X = 3$, given that $Y = 8$.

c. Find the probability that $Y = 8$, given that $X = 3$.

In the next section, we shall talk about an important law, which is the law of Total Probability. It also includes the celebrated Bayes' Theorem.

1.4 BAYES THEOREM

To understand Baye's theorem, we need another result in probability which is also of independent interest. Let us first prove that.

The Law of Total Probability

Let us suppose that $B_1, B_2, B_3, \ldots, B_n$ are the events of a set which form a partition of the sample space S. It means that all these events are mutually exclusive and their union is the sample space.

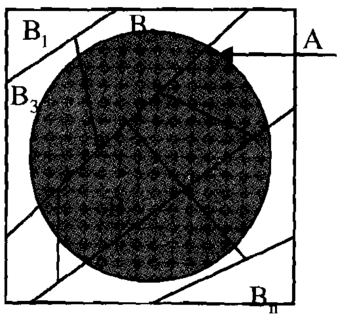

Fig.3

Symbolically, $B_i \cap B_j = \phi$ for $i \neq j$ and i,j=1,2,3,...,n and and

$P(B_i) > 0, i = 1, 2, \ldots, n$. Let A be another event over the sample space. Then we can write $\bigcup_{i=1}^{n} B_i = S$

$$A = A \cap S = A \cap \left\{\bigcup_{i=1}^{n} B_i\right\} = \bigcup_{i=1}^{n} \{A \cap B_i\} \qquad \text{[using the law of distribution]} \qquad (8)$$

Here, $A \cap B_i$ and $A \cap B_j$ are mutually exclusive for all $i \neq j$, $i, j = 1, 2, 3, \ldots, n$

Therefore, using the law of addition, we get

$$P(A) = P\left(\bigcup_{i=1}^{n} \{A \cap B_i\}\right) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i) \qquad (9)$$

Relation (9) is called the Law of Total Probability.
Now, We are ready to state Bayes' Theorem.

Bayes' Theorem was given by Thomas Bayes a British Mathematician in 1763.

Theorem 1 (Bayes Theorem): Let $B_1, B_2, B_3, \ldots, B_n$ be a set of events which form a partition of the sample space, S. Let A be any event with $P(A) > 0$. Then,

$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{\sum_{i=1}^{n} P(B_i)P(A \mid B_i)}, \quad i = 1, 2, 3, \ldots, n \qquad (10)$$

Proof: From the definition of conditional probability, for two events A and B_i, we have

$$P(B_i \mid A) = \frac{P(A \cap B_i)}{P(A)}$$

$$= \frac{P(B_i)P(A \mid B_i)}{P(A)} \qquad \text{[using Eqn.(4)]}$$

$$= \frac{P(B_i)P(A \mid B_i)}{\sum_{i=1}^{n} P(A \cap B_i)} \qquad \text{[using Eqn.(9)]}$$

$$= \frac{P(B_i)P(A \mid B_i)}{\sum_{i=1}^{n} P(B_i)P(A \mid B_i)} \qquad \text{[using Eqn.(4)]}$$

***

In the context of Bayes theorem, the probability $P(B_i)$ is called **a priori probability** of B_i, because it exists prior to the happening of event A in the experiment.
The probability $P(B_i \mid A)$ is termed '**a posteriori probability** , because it is determined after the happening of the event A, i.e. posterior to the event A.
Since the probability $P(B_i \mid A)$ represents the likelihood of the event B_i after event A happens, the probability $P(B_i \mid A)$ is called a 'likelihood' of event B_i after the happening of event A.

Let us apply the above results in the following examples to understand this concept.

Example 6: Suppose in a group of individuals 31% were smokers. It was also observed that $19/31$ of smokers and $41/69$ of non-smokers were male. An individual was chosen at random from the group. What is the probability of the selected individual being a smoker, if he was found to be a male? (Compare with the problem given in Example 2(d).)

Solution: Let the events B_1 be individual is smoker, B_2 be individual is non-smoker, and A be individual is a male.

Clearly $P(B_1) = 0.31$ and $P(B_2) = 1 - 0.31 = 0.69$ and

$$P(A \mid B_1) = \frac{19}{31}, \quad P(A \mid B_2) = \frac{41}{69}$$

Substituting $n = 2$ and $i = 1$ in Eqn.(9) of Bayes' rule, we get

$$P(B_1 \mid A) = \frac{P(B_1)P(A \mid B_1)}{\sum_{i=1}^{2} P(B_i)P(A \mid B_i)} = \frac{P(B_1)P(A \mid B_1)}{P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2)}$$

$$= \frac{0.31 \times \frac{19}{31}}{0.31 \times \frac{19}{31} + 0.69 \times \frac{41}{69}} = \frac{19}{60}, \text{ which is the result in Example 2(d).}$$

***

Here you can see that by applying Bayes' theorem, how easily the problem can be solved.

Example 7: There are three bags. The first bag contains 6 red balls, and 4 blue balls. The second bag contains 2 red balls, and 8 blue balls. The third bag contains 5 red balls, and 5 blue balls. A bag was selected at random from the three bags, and a ball was drawn randomly from it. The ball was found to be blue. What is the probability that the ball came from second bag?

Solution: Let the events B_1 be selecting the first bag, B_2 be selecting the second bag, B_3 be selecting the third bag, and A be the ball drawn is blue.

Thus $P(B_1) = P(B_2) = P(B_3) = 1/3$

$P(A|B_1) =$ the probability of getting a blue ball from the first bag $= 4/10$. Similarly, $P(A|B_2) = 8/10$, $P(A|B_3) = 5/10$.

We want to find the probability that the selected bag was second, given that a blue ball came in the draw, i.e. $P(B_2|A)$. Using the Bayes' theorem, we have

$$P(B_2|A) = \frac{P(B_2)P(A|B_2)}{\sum_{i=1}^{3} P(B_i)P(A|B_i)} = \frac{\frac{1}{3}\cdot\frac{8}{10}}{\frac{1}{3}\cdot\frac{4}{10}+\frac{1}{3}\cdot\frac{8}{10}+\frac{1}{3}\cdot\frac{5}{10}} = \frac{8}{17}$$

***

You may now try the following exercises.

E8) In a die-coin experiment, a fair die is rolled and then a fair coin is tossed a number of times, equal to the score on the die.

a) Find the probability that the coin shows head in every toss.

b) Given that the coin shows heads in all tosses, find the probability that the die score was i, $i = 1, 2, 3, 4, 5, 6$.

E9) A plant that produces memory chips has 3 assembly lines. Line 1 produces 40% of the chips with a defective rate of 5%, line 2 produces 25% of the chips with a defective rate of 6% and line 3 produces 35% of the chips with a defective rate of 3%. A chip is chosen at random from the plant.

a) Find the probability that the chip is defective.

b) Given that the chip is defective, find the probability that the chip was produced by the Line 3.

So far, we have discussed the conditional probability, compound probability, and Bayes' theorem. Now, let us discuss conditional distribution.

1.5 CONDITIONAL DISTRIBUTION

We have learnt about Random Variable in the undergraduate course in detail. We may recall that a random variable is a mathematical function over the sample space of an experiment that maps its outcomes to real numbers. Unlike other mathematical variables, a random variable cannot be assigned a value independently. It only describes the possible outcomes of an experiment in terms of real numbers. Due to this, some people consider the name **random variable** a misnomer.

A **probability distribution**, more properly called a **probability distribution function**, assigns a probability to every interval of the real numbers, so that the probability axioms are satisfied. The probability distribution of the variable X can be uniquely described by its cumulative distribution function, $F(x)$, which is defined by

$$F(x) = P[X \le x] \qquad \text{for every } x \text{ in } \mathbf{R}.$$

Every random variable gives rise to a probability distribution, and this distribution contains most of the important information about the variable. If X is a random variable, the corresponding probability distribution assigns to the interval $(a, b]$, the probability $Pr[a < X \le b]$, i.e. the probability that the variable X will take a value in the interval $(a, b]$. This probability can be expressed in terms of the cumulative distribution function $F(x)$.

A probability distribution is called **discrete** if its cumulative distribution function is a step function consisting of a sequence of a countable number of jumps, which means that it corresponds to a discrete random variable: a variable which can only attain values from a certain finite, or countable, set. Here, we use **probability mass function** (abbreviated **p.m.f.**) to represent the probability distribution. It gives the probability that a discrete random variable is exactly equal to some value, that is **p.m.f.** is $f(x) = P(X = x)$.

A probability distribution is called **continuous** if its cumulative distribution function is continuous, which means that it corresponds to a random variable X for which $P[X = x] = 0$ for all x in $\mathbf{R}$. In this case, we use a **probability density function:** a non-negative function f defined on the real numbers, such that

$$P[a<X \le b]= \int_a^b f(x)\,dx$$

for all a and b, to assign the probability that the variable X will take a value in the interval $(a, b]$.

You must have studied conditional distribution in your earlier course. Just to recaptulate, let us discuss it here again. Let us first discuss the formal definition of conditional distribution.

Given two jointly distributed random variables X and Y, that is, a two-dimensional random variable or vector (X, Y), the **conditional probability distribution** of Y given X (written "$Y|X$") is the probability distribution of Y when X is known to have taken a particular value.

Definition 2: Let X and Y be two discrete random variables (r.v.s.) associated with the same random experiment, taking values in countable sets, T_x and T_y respectively. The function $f(x, y)$ defined for all ordered pairs (x, y), $x \in T_X$ and $y \in T_Y$ by the relation

$$f(x, y) = P[X = x, Y = y]$$

is called the **joint probability mass function** of X and Y.

Note: By definition,

$$f(x, y) \ge 0$$

and

$$\sum_{x \in T_X} \sum_{y \in T_Y} f(x_j, y_k) = 1$$

Moreover, we should clarify that $[X = x, Y = y]$ really stands for the event $[X = x] \cap [Y = y]$, and that $[X = x, Y = y]$ is a simplified and accepted way of expressing the intersection of the two events, $[X = x]$ and $[Y = y]$.

Let us consider the following example.

Example 8: A committee of two persons, is formed by selecting them at random and without replacement from a group of 10 persons, of whom 2 are mathematicians, 4 are statisticians and 4 are engineers. Let X and Y denote the number of mathematicians and statisticians, respectively, in the committee. The possible values of X are 0, 1, 2, which are also the possible values of Y. Thus, all the ordered pairs (x, y) of the values of X and Y are

$(0, 0), (0,1), (0, 2), (1, 0), (1, 1), (2, 0), (1, 2), (2, 1)$ and $(2, 2)$.

The total number of ways of selecting two persons from a group of 10 persons is $^{10}C_2 = 45$. Since the persons are selected at random, each of these 45 ways has the same probability $\frac{1}{45}$. Consider the event $[X = 1, Y = 1]$ that a committee has one mathematician and one statistician. One mathematician can be selected from two in

${}^2C_1 = 2$ ways, and one statistician can be selected from 4 statisticians in ${}^4C_1 = 4$ ways. Hence, the total number of committees with 1 mathematician and 1 statistician is $2 \times 4 = 8$. Thus, $P[x=1, Y=1] = \frac{8}{45}$.

To obtain the probability of the event $[X=0, Y=1]$, observe that if $X=0, Y=1$, this means that 1 statistician is on the committee, and that no mathematician is on it. Then, the other person on the committee has to be one of the 4 engineers. This engineer can be selected in ${}^4C_1 = 4$ ways. Hence,

$$P[X=0, Y=1] = \frac{{}^4C_1 \times {}^4C_1}{45} = \frac{16}{45}.$$

Similarly, we can obtain

$$P[X=0, Y=0] = \frac{{}^4C_2}{45} = \frac{6}{45}$$

$$P[x=0, Y=2] = \frac{{}^4C_2}{45} = \frac{6}{45}$$

$$P[X=1, Y=0] = \frac{{}^2C_1 \times {}^4C_1}{45} = \frac{8}{45}$$

$$P[X=2, Y=0] = \frac{{}^2C_2}{45} = \frac{1}{45}$$

Since the committee has only two members, it is obvious that there are no sample points corresponding to the events $[X=1, Y=2]$, $[X=2, Y=1]$ and $[X=2, Y=2]$. Hence, the probabilities $P[X=1, Y=2] = P[X=2, Y=1] = P[X=2, Y=2] = 0$.
We now summarise these calculations in the following table.

Table 1: $P[X=x, Y=y]$ for $x, y = 0, 1, 2$.

x \ y	0	1	2
0	6/45	16/45	6/45
1	8/45	8/45	0
2	1/45	0	0

Note: If we denote probability $P[X=x, Y=y]$ by $f(x, y)$, then the function $f(x, y)$ is defined for all pairs (x, y) for values x and y of X and Y, respectively.
Moreover,

$$f(x, y) \geq 0$$

and

$$\sum_{x=0}^{2} \sum_{y=0}^{2} f(x, y) = 1$$

We say that the function $f(x, y)$ is the joint probability mass function of the r.v.s. X, Y, or random vector (X, Y).

We now define the p.m.f. of the marginal distribution.

Let X and Y be r.v.s. taking values $x \in T_X$ and $y \in T_Y$, respectively and joint p.m.f. $f(x, y) = P[X=x, Y=y]$.

We define new functions, g and h , as follows:

$$g(x)=\sum_{y\in T_Y} f(x, y) \tag{11}$$

and $$h(y)=\sum_{x\in T_X} f(x, y) \tag{12}$$

In Eqn.(11), we keep the value x of X fixed and sum $f(x, y)$ over all values y of Y. On the other hand, in Eqn.(12), y is kept fixed and $f(x, y)$ is summed over all values of X. We wish to interpret the function $g(x)$ defined for all value, x of X and the function $h(y)$ defined for all values y of Y. Notice that both g and h, being sums of non-negative numbers, are themselves non-negative. Further,

$$\sum_{x\in T_X} g(x)=\sum_{x\in T_X}\sum_{y\in T_Y} f(x, y)=1$$

Thus, $g(x)$ has all the properties of a p.m.f. of one=dimensional r.v. Similarly, you can verify that $h(y)$ also has all the properties of a p.m.f. We call these the p.m.f. of the marginal distribution of X and Y respectively, as you can see from the following definition.

Definition 3: The function $g(x)$ defined for all values $x\in T_x$ of the r.v. X by the relation

$$g(x)=\sum_{y\in T_Y} f(x, y)$$

is called the p.m.f. of the marginal distribution of X. Similarly, $h(y)$ defined for all the values $y\in T_y$ of the r.v. Y by the relation

$$h(y)=\sum_{x\in T_X} f(x, y)$$

is called the p.m.f. of the marginal distribution of Y.

Definition 4: As usual, assume a random experiment that has a sample space S and a probability function P on S. Suppose that X and Y be two discrete random variables for the experiment, taking values in the sets T_X, T_Y respectively. For discrete random variables (X, Y), the conditional probability mass function Y given $X=x$, $x\in T_X$ can be written as $P(Y=y\,|\,X=x)$, $y\in T_Y$. From the definition of conditional probability,

$$P[Y=y\,|\,X=x]=\frac{P[Y=y, X=x]}{P[X=x]}, \text{ provided } P[X=x]>0 \tag{13}$$

and if $f_{X,Y}(x,y)$ be the joint probability mass function of X and Y, and $f_X(x)(>0)$ be the marginal probability mass function of X, then the conditional probability mass function $P(Y=y\,|\,X=x)$, for a given $x\in T_x$ can be expressed as

$$P[Y=y\,|\,X=x]=\frac{f_{X,Y}(x,y)}{f_X(x)};\quad y\in T_y \tag{14}$$

Similarly, if X and Y be continuous random variables for an experiment, then letting $f_{X,Y}(x,y)$ to be the joint probability density function of X and Y, and $f_X(x)$, $f_Y(y)$ be the marginal probability density function of X and Y respectively, then the conditional probability density function of Y given $X=x$ can be written as $f_{Y|X}(y\,|\,x)$, and is defined by

$$f_{Y|X}(y\,|\,x)=\frac{f_{X,Y}(x,y)}{f_X(x)},\qquad -\infty<y<\infty \tag{15}$$

provided $f_X(x)>0$.

Likewise, the conditional probability density function of X given $Y=y$ can be denoted as $f_{X|Y}(x\,|\,y)$ and can be defined by

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad -\infty < x < \infty \tag{16}$$

provided $f_Y(y) > 0$.

(We will use the notation $f_{X,Y}(x,y)$ for joint probability density function (p.d.f.) and for joint probability mass function (p.m.f.), both. In the examples and exercises, for simplicity, $f(x, y)$ has been used to represent $f_{X,Y}(x,y)$.

Definition 5: For two-dimensional discrete random variable (X, Y), the conditional cumulative distribution function of a discrete random variable Y, given $X = x$, is defined as

$$F_{Y|X}(y \mid x) = \sum_{u \le y} P[Y = u \mid X = x] \qquad -\infty < y < \infty \tag{17}$$

and the conditional cumulative distribution function of random variable X, given $Y = y$, is defined as

$$F_{X|Y}(x \mid y) = \sum_{u \le x} P[X = u \mid Y = y] \qquad -\infty < x < \infty \tag{18}$$

Similarly, if (X, Y) is two-dimensional continuous r.v., the conditional cumulative distribution function of random variable Y, given $X = x$, is defined as

$$F_{Y|X}(y \mid x) = \int_{-\infty}^{y} f_{Y|X}(u \mid x)\,du \qquad -\infty < y < \infty \tag{19}$$

and conditional cumulative distribution function of random variable X, given $Y = y$, is defined as

$$F_{X|Y}(x \mid y) = \int_{-\infty}^{x} f_{X|Y}(u \mid y)\,du \qquad -\infty < x < \infty \tag{20}$$

The conditional probability mass functions and conditional probability density functions of random variables also satisfy properties of unconditional probability distributions. They are non=negative on the basis of being the ratio of non=negative and positive functions. It can easily be shown that they also sum up, or integrate to 1.

Theorem 2: For any pair of discrete random variable, X, Y

$$\sum_{y \in T_y} P[Y = y \mid X = x] = 1$$

Proof: $$\sum_{y \in T_y} P[Y = y \mid X = x] = \sum_{y \in T_y} \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$= \frac{\sum_{y \in T_y} f_{X,Y}(x,y)}{f_X(x)}$$

$$= \frac{f_X(x)}{f_X(x)} = 1$$

***

Theorem 3: For a pair of continuous random variables, X, Y,

$$\int_{-\infty}^{\infty} f_{Y|X}(y \mid x)\,dy = 1$$

Proof: $$\int_{-\infty}^{\infty} f_{Y|X}(y \mid x)\,dy = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy$$

$$= \frac{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy}{f_X(x)}$$

$$= \frac{f_X(x)}{f_X(x)} = 1$$

***

Definition 6: Two discrete random variables, X and Y, are called independent if, and only if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \text{ for all } x \in T_x, \text{ and all } y \in T_y \tag{21}$$

where $f_{X,Y}(x,y)$ is the joint probability mass function of X and Y, and $f_X(x)$, $f_Y(x)$ are the marginal probability mass functions of X and Y respectively. Likewise, two continuous random variables, X and Y are called independent, if and only if

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y), \quad -\infty < x < \infty, \ -\infty < y < \infty \tag{22}$$

where $f_{X,Y}(x,y)$ is the joint probability density function of X and Y, and $f_X(x)$, $f_Y(y)$ are the marginal probability density functions of random variables X and Y, respectively.

It may be easily verified that the following conditions with usual notations are equivalent for the independent random variables, X and Y, both for discrete and continuous case.

a. $f_{X|Y}(x \mid y) = f_X(x)$

b. $f_{Y|X}(y \mid x) = f_Y(y)$

c. $f_{X,Y}(x, y) = f_X(x) f_Y(y)$

for all $-\infty < x < \infty, -\infty < y < \infty$.

Now, look at some examples.

Example 9: The random variables, X and Y, have the following joint probability mass function $f_{X,Y}(x,y)$ in the cells of a bivariate probability table.

(x, y)	0	1	2
0	2/9	1/3	1/15
1	2/9	2/15	
2	1/45		

Find

(i) The conditional probability distribution of Y given X

(ii) The conditional probability distribution of X given Y

(iii) Are X and Y independent?

Solution: (i) Here, $T_x = \{0, 1, 2\}$, $T_y = \{0, 1, 2\}$

Clearly, marginal probability mass function of X, may be evaluated as follows:

$$f_X(x) = \sum_{y \in T_y} f_{X,Y}(x,y)$$

$$= \sum_{y=0}^{2} f_{X,Y}(x,y)$$

$$= f_{X,Y}(x,0) + f_{X,Y}(x,1) + f_{X,Y}(x,2)$$

$$f_X(0) = f_{X,Y}(0,0) + f_{X,y}(0,1) + f_{X,y}(x,2)$$
$$= 2/9 + 1/3 + 1/15 = 28/45$$
$$f_X(1) = f_{X,Y}(1,0) + f_{X,Y}(1,1) + f_{X,Y}(1,2)$$
$$= 2/9 + 2/15 + 0 = 16/45$$
$$f_X(2) = f_{X,Y}(2,0) + f_{X,Y}(2,1) + f_{X,Y}(2,2)$$
$$= 1/45 + 0 + 0 = 1/45$$

These are basically row totals in the bivariate table. Similarly, column totals give $f_Y(y)$.

(x, y)	0	1	2	$f_X(x)$
0	2/9	1/3	1/15	28/45
1	2/9	2/15		16/45
2	1/45			1/45
$f_Y(y)$	21/45	7/15	1/15	1

The conditional probability distribution of Y, given $X = x,\ x \in T_x$, may be calculated as follows:

$$P[Y = y \mid X = x] = \frac{f_{X,Y}(x,y)}{f_X(x)}; \qquad y \in T_y$$

$$P[Y = y \mid X = 0] = \frac{f_{X,Y}(0,y)}{f_X(0)} = \frac{45}{28} f_{X,Y}(0,y); \qquad y = 0,1,2$$

$$P[Y = y \mid X = 1] = \frac{f_{X,Y}(1,y)}{f_X(1)} = \frac{45}{16} f_{X,Y}(1,y); \qquad y = 0,1,2$$

$$P[Y = y \mid X = 2] = \frac{f_{X,Y}(2,y)}{f_X(2)} = \frac{45}{1} f_{X,Y}(2,y); \qquad y = 0,1,2$$

Thus, conditional probability distribution of Y, given X, are found as shown in the rows of the following table.

(x, y)	$P[Y = y \mid X = x]$			Total
	0	1	2	
0	(45/28)(2/9) = 5/14	(45/28)(1/3) = 15/28	(45/28)(1/15) = 3/28	1
1	(45/16)(2/9) = 5/8	(45/16)(2/15) = 3/8	(45/16)(0) = 0	1
2	45 (1/45) = 1	45 (0) = 0	45 (0) = 0	1

(ii) In a similar way, we may obtain conditional probability distribution of X given Y.

(iii) Since $f_{X,Y}(0,0) \neq f_X(0) f_Y(0)$, therefore, X and Y are not independent.

***

Example 10: Suppose that (X, Y) has joint density function

$$f(x,y) = \begin{cases} x + y, & \text{if } 0 < x < 1,\ 0 < y < 1 \\ 0 \quad , & \text{otherwise} \end{cases}$$

a) Find the conditional density of X given $Y = y$.

b) Find the conditional density of Y given $X = x$.

c) Are X and Y independent?

d) Compute $P[0 < X < 1/2 \mid Y = 1/3]$.

Solution: The marginal probability density function of X

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

$$= \int_0^1 (x + y)dy$$

$$= \left[xy + \frac{y^2}{2.} \right]_{y=0}^{y=1}$$

or, $$f_X(x) = \begin{cases} x + 1/2, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

The marginal probability density function of Y

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$$

$$= \int_0^1 (x + y)dx$$

$$= \left[\frac{x^2}{2} + yx \right]_{x=0}^{x=1} = \frac{1}{2} + y \text{ and } f_Y(y) = 0 \text{ otherwise}$$

$$f_Y(y) = \begin{cases} 1/2 + y, & \text{if } 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

a) The conditional probability density function of X given $Y = y, 0 < y < 1$ will be

$$f_{X|Y}(x \mid y) = \frac{f_{x,y}(x, y)}{f_y(y)} \qquad -\infty < x < \infty$$

or $$f_{X|Y}(x \mid y) = \begin{cases} \dfrac{x + y}{y + 1/2}, & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

b) The conditional probability density function of Y, given $X = x, 0 < x < 1$, will be

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \qquad -\infty < y < \infty$$

or $$f_{Y|X}(y \mid x) = \begin{cases} \dfrac{x + y}{x + 1/2}, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

c) Since $f_{X,Y}(x, y) \neq f_X(x)f_Y(y)$, the random variables are not independent.

d) Now, the $P[0 < X < 1/2 \mid Y = 1/3]$ will be obtained by integrating the conditional probability density function of X given $Y = 1/3$ over $[0, 1/2]$.

$$P[0 < X < 1/2 \mid Y = 1/3] = \int_0^{1/2} f_{X|Y}(x \mid 1/3)dx$$

Since, $$f_{X|Y}(x \mid 1/3) = \left. \frac{x + y}{y + 1/2} \right|_{y=1/3}$$

$$= \frac{x + 1/3}{1/3 + 1/2} = \frac{6}{5}(x + 1/3)$$

Thus, $P[0<X<1/2 \mid Y=1/3]=\int_0^{1/2}\frac{6}{5}(x+1/3)\,dx$

$$=\frac{6}{5}\left[\frac{x^2}{2}+\frac{x}{3}\right]_0^{1/2}$$

$$=\frac{6}{5}\cdot\frac{7}{24}=\frac{7}{20}$$

***

Example 11: Suppose that random variable (X, Y) has joint probability density function, f as given below

$$f(x,y)=\begin{cases}2(x+y), & \text{if } 0<x<y<1\\ 0, & \text{otherwise}\end{cases}$$

a) Find the conditional density of X, given $Y=y$.

b) Find the conditional density of Y, given $X=x$.

c) Are X and Y independent?

Solution: The marginal probability density function of X

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

$$=\int_x^1 2(x+y)dy$$

$$=2\left[xy+\frac{y^2}{2}\right]_{y=x}^{y=1}$$

$$=2(x+\frac{1}{2}-\frac{3}{2}x^2)=1+2x-3x^2,\ \text{if } 0<x<1$$

and $f_X(x)=0$, otherwise

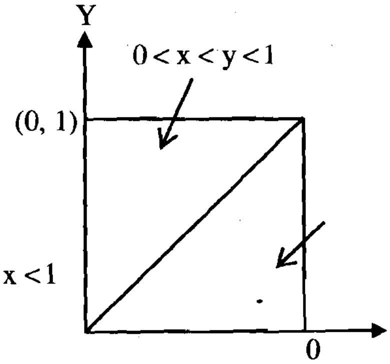

Fig.4

The marginal probability density function of Y

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

$$=\int_0^y 2(x+y)dx$$

$$=2\left[\frac{x^2}{2}+yx\right]_{x=0}^{x=y}$$

$$=2(\frac{y^2}{2}+y^2)=3y^2,\quad \text{if } 0<y<1$$

and $f_Y(y) = 0,$ otherwise

a) The conditional probability density function of X, given $Y=y, 0<y<1$ will be

$$f_{X|Y}(x\mid y)=\frac{f_{X,Y}(x,y)}{f_Y(y)},\qquad -\infty<x<\infty$$

$$=\frac{2(x+y)}{3y^2},\qquad 0<x<y$$

and $f_{X|Y}(x\mid y)=0,$ otherwise

b) The conditional probability density function of Y, given $X = x, 0 < x < 1$ will be

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \qquad -\infty < y < \infty$$

$$= \frac{2(x+y)}{1+2x-3x^2}, \qquad x < y < 1$$

and $\quad f_{Y|X}(y \mid x) = 0, \qquad$ otherwise

c) Since $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$, $0 < x < y < 1$, the random variables X and Y are not independent.

***

You may now try the exercises that follow.

E10) Two dice are thrown. Let X denotes the sum of the scores on two dice and Y denotes the absolute value of their difference.

a) Find the joint probability mass function of X and Y.
b) Find the marginal probability mass function of X.
c) Find the marginal probability mass function of Y.
d) Find the conditional probability mass function of Y given $X = x$.
e) Find the conditional probability mass function of X given $Y = y$.
f) Are X and Y independent?

E11) Suppose that the random variables X and Y have a joint probability density function f as given below

$$f(x,y) = \begin{cases} (x+y)/4, & \text{if } 0 < x < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

a) Find the conditional density of X, given $Y = y$.
b) Find the conditional density of Y, given $X = x$.
c) Are X and Y independent?

E12) Suppose that the random variables X and Y have a joint probability density function f as given below

$$f(x,y) = \begin{cases} e^{-(x+y)}, & \text{if } 0 < x < \infty,\ 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

a) Find the conditional density of X, given $Y = y$.
b) Find the conditional density of Y, given $X = x$.
c) Find $P[X < 1]$, $P[X < Y]$, and $P[X + Y < 1]$

E13) Suppose that the random variables X and Y have a joint probability density function f as given below

$$f(x,y) = \begin{cases} 2, & \text{if } 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

a) Check whether or not independence of X and Y holds.
b) Find $P[X < 0.2 \mid Y > 0.1]$, $P[0.1 < Y < 0.4]$.

The exercises in this section would have given you enough practice to compute the density functions and distribution functions of bivariate random variables. Next, we shall discuss measures of central tendency of the probability distribution of bivariate random vectors.

1.6 CONDITIONAL EXPECTATIONS

We shall begin this section with the definition of the conditional expectation of a function of one random variable, given that the other variable has taken a given value.

Definition 7: Assume a random experiment has a sample space S, and a probability function, P on S. Let that X and Y be two discrete random variables for the experiment, taking values in the countable sets, T_X and T_Y (being subsets of $\mathbf{R}$), respectively. Let $f_{X,Y}(x,y)$ be the joint probability mass function of random variables X and Y, and $f_X(x), f_Y(y)$, be the marginal probability mass functions of X and Y, respectively. Then, the conditional expectation of Y, given $X=x$, denoted by $E(Y|X=x)$, or $E(Y|x)$ is defined as

$$E(Y|X=x)=\sum_{y\in T_Y} y\frac{f_{X,Y}(x,y)}{f_X(x)}, \tag{23}$$

provided the series on right hand side of Eqn. (23) is absolutely convergent. Here, $E(Y|X=x)$ is a function of x since x can take any value in T_X, and the conditional expectation of X, given $Y=y$, denoted by $E(X|Y=y)$ or $E(X|y)$ is defined as

$$E(X|Y=y)=\sum_{x\in T_X} x\frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{24}$$

provided right hand side of Eqn. (24) converges absolutely. Here, $E(X|Y=y)$ is a function of y, since y can take any value in T_Y. Similarly, if X and Y be continuous random variables for the experiment, having $f_{X,Y}(x,y)$ as the joint probability density function and $f_X(x), f_Y(y)$ be the marginal probability density functions of X and Y, respectively, then the conditional expectation of Y, given $X=x$, is defined as

$$E(Y|X=x)=\int_{-\infty}^{\infty} y\frac{f_{X,Y}(x,y)}{f_X(x)}dy \tag{25}$$

provided the integral on right hand side of Eqn. (25) converges absolutely, and the conditional expectation of X, given $Y=y$, is defined as

$$E(X|Y=y)=\int_{-\infty}^{\infty} x\frac{f_{X,Y}(x,y)}{f_Y(y)}dx \tag{26}$$

provided the right hand side of Eqn. (24) is absolutely convergent. Here, again we note that $E(Y|X=x)$ is a function of x, and $E(X|Y=y)$ is a function of y, as both x and y can vary in $\mathbf{R}$. The conditional expectation of a function of a random variable can also be defined in similar way. For the discrete random variables, X and Y, as specified above, the conditional expectation of $\phi(Y)$, a function of random variable Y, given $X=x$, will be

$$E(\phi(Y)|X=x)=\sum_{y\in T_Y} \phi(y)\frac{f_{X,Y}(x,y)}{f_X(x)} \tag{27}$$

and for the continuous random variables, X and Y, as specified above, this conditional expectation will be

$$E(\phi(Y)|X=x)=\int_{-\infty}^{\infty} \phi(y)\frac{f_{X,Y}(x,y)}{f_X(x)}dy \tag{28}$$

The conditional variance of X, given $Y=y$, can now be defined as

$$V(X|Y=y)=E(\{X-E(X|Y=y)\}^2|Y=y) \tag{29}$$

It can be easily shown that the above expression has following equivalent form

$$V(X|Y=y)=E(X^2|Y=y)-\{E(X|Y=y)\}^2 \tag{30}$$

Example 12: For the problem discussed in **Example 9,** obtain

(i) Expectation of Y, given $X=1$.

(ii) Expectation of Y, given $X=x$, i.e., $E(Y|X=x)$.

(iii) Expectation of the expectation of Y, given $X=x$ i.e., $E(E(Y|X))$

(iv) Find $V(X|Y=0)$

Solution: (i) Since,

$$E(Y|X=1)=\sum_{y\in T_Y} y\frac{f_{X,Y}(1,y)}{f_X(1)}$$

$$=\frac{1}{f_X(1)}\sum_{y=0}^{2} y\,.\,f_{X,Y}(1,y)$$

$$=\frac{45}{16}(0\,.\,f_{X,Y}(1,0)+1\,.\,f_{X,Y}(1,1)+2\,.\,f_{X,Y}(1,2))$$

$$=\frac{45}{16}\left(0\,.\,\frac{2}{9}+1\,.\,\frac{2}{15}+2\,.\,0\right)=\frac{3}{8}$$

(ii) $$E(Y|X=x)=\sum_{y\in T_Y} y\frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$=\frac{1}{f_X(x)}\sum_{y=0}^{2} y\,.\,f_{X,Y}(x,y)$$

$$=\frac{1}{f_X(x)}(0\,.\,f_{X,Y}(x,0)+1\,.\,f_{X,Y}(x,1)+2\,.\,f_{X,Y}(x,2))$$

(iii) $$E(E(Y|X))=E(\sum_{y\in T_Y} y\frac{f_{X,Y}(X,y)}{f_X(X)})$$

$$=E(\frac{1}{f_X(X)}\sum_{y=0}^{2} y\,.\,f_{X,Y}(X,y))$$

$$=\sum_{x=0}^{2}\{\frac{1}{f_X(x)}(0\,.\,f_{X,Y}(x,0)+1\,.\,f_{X,Y}(x,1)+2\,.\,f_{X,Y}(x,2))\}f_X(x)$$

$$=0+\sum_{x=0}^{2} f_{X,Y}(x,1)+2\sum_{x=0}^{2} f_{X,Y}(x,2)=\frac{7}{15}+2\,.\,\frac{1}{15}=\frac{3}{5}$$

which is the same as $E(Y)$. Since,

$$E(Y)=\sum_{y=0}^{2} y\,.\,P(Y=y)$$

$$=0\,.\,P(Y=0)+1\,.\,P(Y=1)+2\,.\,P(Y=2)$$

$$=0+1.\frac{7}{15}+2\,.\,\frac{1}{15}=\frac{3}{5}$$

(iv) $$V(X|Y=0)=E(X^2|Y=0)-\{E(X|Y=0)\}^2$$

$$=\sum_{x\in T_X} x^2\frac{f_{X,Y}(x,0)}{f_Y(0)}-\left\{\sum_{x\in T_X} x\frac{f_{X,Y}(x,0)}{f_Y(0)}\right\}^2$$

$$=\frac{1}{f_Y(0)}\{0\,.\,f_{X,Y}(0,0)+1\,.\,f_{X,Y}(1,0)+4\,.\,f_{X,Y}(2,0)\}-$$

$$\left\{\frac{1}{f_Y(0)}\{0\,.\,f_{X,Y}(0,0)+1\,.\,f_{X,Y}(1,0)+2\,.\,f_{X,Y}(2,0)\}\right\}^2$$

$$= \frac{45}{21}\left\{0+\frac{2}{9}+4\cdot\frac{1}{45}\right\}-\left\{\frac{45}{21}\left\{0+\frac{2}{9}+2\cdot\frac{1}{45}\right\}\right\}^2$$

$$= \frac{2}{3}-\left\{\frac{12}{21}\right\}^2 = \frac{294-144}{441} = \frac{50}{147}$$

***

Example 13: Let the continuous random variables, X and Y, have the following joint probability density function

$$f(x,y) = \begin{cases} 8xy, & \text{if } 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

(i) Find the expectation of Y, given $X = x$, i.e., $E(Y \mid X = x)$

(ii) Find $V(Y \mid X = x)$

Solution: (i) Since

$$f_x(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

$$= \int_x^1 8xy\,dy$$

$$= 4x(1-x^2) \text{ for } 0 < x < 1, \text{ and } f_X(x) = 0, \text{ otherwise}$$

Therefore,

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy$$

$$= \int_x^1 y \frac{8xy}{4x(1-x^2)}\,dy = \frac{2}{1-x^2}\left[\frac{y^3}{3}\right]_x^1$$

$$= \frac{2}{3}\left(\frac{1+x+x^2}{1+x}\right)$$

(ii) Since

$$E(Y^2 \mid X = x) = \int_{-\infty}^{\infty} y^2 \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy$$

$$= \int_x^1 y^2 \frac{8xy}{4x(1-x^2)}\,dy$$

$$= \frac{1}{2}\left(\frac{1-x^4}{1-x^2}\right) = \frac{1+x^2}{2}$$

Therefore,

$$V(Y \mid X = x) = E(Y^2 \mid X = x) - \{E(Y \mid X = x)\}^2$$

$$= \frac{1+x^2}{2} - \frac{4}{9}\left(\frac{1+x+x^2}{1+x}\right)^2$$

***

You must have seen in (iii) part of Example 12, that $E(E(y \mid x))$ and $E(y)$ both attain the same value, $3/5$. Now let us try to prove this in the theorem that follows.

Theorem 4: The expectation of the conditional expectation of Y, given X, is equal to the expectation of Y i.e. $E(E(Y|X)) = E(Y)$.

Proof: Suppose X and Y are discrete random variables. Therefore, from Eqn.(23), we have

$$E(Y \mid X = x) = \sum_{y \in T_Y} y \frac{f_{X,Y}(x, y)}{f_X(x)}, \text{ which is a function of } x.$$

Thus,

$$E(E(Y \mid X)) = E\left\{ \sum_{y \in T_Y} y \frac{f_{X,Y}(X, y)}{f_X(X)} \right\}$$

$$= \sum_{x \in T_X} \left\{ \sum_{y \in T_Y} y \frac{f_{X,Y}(x, y)}{f_X(x)} \right\} f_X(x)$$

$$= \sum_{x \in T_X} \frac{1}{f_X(x)} \left\{ \sum_{y \in T_Y} y f_{X,Y}(x, y) \right\} f_X(x)$$

$$= \sum_{y \in T_Y} \left\{ \sum_{x \in T_X} y f_{X,Y}(x, y) \right\} \qquad \text{(changing the order of summation)}$$

$$= \sum_{y \in T_Y} y \left\{ \sum_{x \in T_X} f_{X,Y}(x, y) \right\} = \sum_{y \in T_Y} y f_Y(y) = E(Y)$$

This result holds for the continuous random variables also. In similar way, the result can be proved for the continuous random variables also.

***

Now let us prove another theorem.

Theorem 5: Prove that $V(X) = E(V(X \mid Y)) + V(E(X \mid Y))$

Proof: We have proved that $E(E(Y|X)) = E(Y)$. We can prove, similarly, that $E(E(X|Y)) = E(X)$ and $E(E(X^2|Y)) = E(X^2)$

We start from right hand side of the statement,

$$E(V(X \mid Y)) + V(E(X \mid Y))$$
$$= E\{E(X^2 \mid Y) - (E(X \mid Y))^2\} + E\{E(X \mid Y)\}^2 - \{E(E(X \mid Y))\}^2$$
$$= E\{E(X^2 \mid Y)\} - E(E(X \mid Y))^2 + E\{E(X \mid Y)\}^2 - \{E(X)\}^2$$
$$= E(X^2) - \{E(X)\}^2 = V(X), \text{ the left hand side.}$$

In the proof, we used the fact that

$$E(E(X^2|Y)) = E(X^2)$$

***

You may now try the following exercises based on these discussions.

E14) Suppose that (X, Y) is uniformly distributed on the square $R = \{(x, y): -6 < x < 6, -6 < y < 6\}$. That is, the joint probability density of (X, Y) is

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{144}, & -6 < x < 6, -6 < y < 6 \\ 0, & \text{otherwise} \end{cases}.$$

Then find

(i) Expectation of Y given $X = x$, i.e., $E(Y \mid X = x)$

(ii) Expectation of X given $Y = y$, i.e., $E(X \mid Y = y)$

(iii) $V(Y \mid X = x)$

(iv) $V(X \mid Y = y)$

E15) Suppose that a random vector (X, Y) has a joint probability density function, f as given below

$$f(x,y)=\begin{cases} kxy, & \text{if } 0<y<x<1 \\ 0, & \text{otherwise} \end{cases}$$

(i) Find k.

(ii) Find $E(Y|X=x)$.

(iii) Find $V(Y|X=x)$.

Here, we close the discussion on conditional probability. We hope that you have gained considerable knowledge about conditional probability, and conditional distribution. Now let us summarise what we discussed in this unit.

1.7 SUMMARY

In this unit, we have covered the following points.

1. We illustrated the idea of conditional probability, and presented some examples to elaborate its concept. The conditional probability of an event is obtained on the basis of prior knowledge of the happening of another event. For the evaluation of the conditional probability of an event, the sample space gets reduced to the event whose occurrence has taken place.
2. We attempted to describe conditional probability. Conditional probability is influenced by the happening of another event if the events are dependent, otherwise, it is not influenced and the events are called independent.
3. We studied some basic properties of conditional probability. They were similar to the general properties of a probability function on a sample space.
4. We stated and proved the famous Bayes' theorem. We presented simple examples to illustrate it.
5. We have acquainted you with the concept of conditional distribution with some examples.
6. We have defined conditional expectation of a random vector, and some important properties. Finally, we defined conditional variance.

1.8 SOLUTIONS/ANSWERS

E1) a) If $B \subseteq A$, then $A \cap B = B$, therefore, we have

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(B)}{P(B)}=1$$

b) If $A \subseteq B$ then $A \cap B = A$, therefore, we get

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A)}{P(B)}$$

c) If A and B are disjoint then $A \cap B = \phi$, therefore, we get

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(\phi)}{P(B)}=0$$

E2) a) Since, $P(A)>0$, and $P(B)>0$

$$P(A|B)>P(A)$$

$$\Leftrightarrow \frac{P(A\cap B)}{P(B)}>P(A)$$

$$\Leftrightarrow \frac{P(A \cap B)}{P(A)} > P(B) \Leftrightarrow P(A \cap B) > P(A)P(B) \Leftrightarrow P(B \mid A) > P(B)$$

b) Proof is left for you.

c) Since, $P(A) > 0, \text{and} P(B) > 0$, thus

$$P(A|B) = P(A)$$

$$\Leftrightarrow \frac{P(A \cap B)}{P(B)} = P(A)$$

$$\Leftrightarrow \frac{P(A \cap B)}{P(A)} = P(B) \Leftrightarrow P(B|A) = P(B)$$

and $\frac{P(A \cap B)}{P(B)} = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$

E3) Let event A = the student is absent, and the event B = today is Friday. Since there are six school days, thus, $P(B) = 1/6$, and $P(A \cap B) = 0.03$. Therefore, the required probability is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.03}{1/6} = 0.18.$$

E4) Let event A = the coin is biased, and the event B = the coin lands heads up. The bag contains 12 coins: 5 fair, 4 biased, each with probability of heads $1/3$; and, 3 two-headed.

a) P (coin is biased) $= P(A) = 4/12 = 1/3$.

b) P (coin is biased and it lands head) $= \frac{1}{9}$

c) P (the coin lands heads, given that it is biased)

$$= P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/9}{1/3} = 1/3.$$

E5) A box contains 8 balls: 3 of them are red, and the remaining 5 are blue. Two balls are drawn successively, at random, and without replacement. Let the event A be that a red ball is drawn in the first draw and event B be that the blue ball is drawn in the second draw. The required probability is $P(A \cap B)$, and

$$P(A \cap B) = P(A)\ P(B \mid A) = \frac{3}{8}\frac{5}{7} = \frac{15}{56}$$

E6) Event A = a person smokes, and event B = a person has heart disease. We are given

$$P(A) = 0.3,\ P(B) = 0.08$$

$$P(B \mid A) = 0.12$$

a) We require $P(A \cap B)$

$$\begin{aligned} P(A \cap B) &= P(A)\ P(B \mid A) \\ &= 0.3 \times 0.12 \\ &= 0.036 \end{aligned}$$

Thus, the percentage of population that smoke is 3.6%.

b) We require $P(A \mid B)$ here

$$\begin{aligned} P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{0.03}{0.08} = 0.45 \end{aligned}$$

Therefore the percentage of smokers with heart disease is 45%.

E7) The sample space is $S=\{(a, b): a, b=1, 2,\ldots, 6\}$, where a and b are scores on the first and second die, respectively. $X=3$ and $Y=8$ means that the score at first die 3, and the score at second die is 5, i.e. the outcome is $(3, 5)$. Event $Y=8$ may occur if the outcomes be either of, the following $(2, 6), (3, 5), (4,4), (5, 3), (6, 2)$. Thus, $P(Y=8)=5/36$, assuming 36 outcomes in S are equally likely.

a) $P(X=3 \text{ and } Y=8)=P\{(3, 5)\}=1/36$

b) $P(X=3|Y=8)=\dfrac{P(X=3 \text{ and } Y=8)}{P(Y=8)}=\dfrac{1/36}{5/36}=1/5$

c) $P(Y=8|X=3)=\dfrac{P(X=3 \text{ and } Y=8)}{P(X=3)}=\dfrac{1/36}{6/36}=1/6$

E8) Let the events B_i = the score on the die is i, where $i=1, 2, 3, 4, 5, 6$, and event A = all tosses of coin show heads

Clearly, $P(B_i)=1/6, i=1, 2, 3, 4, 5, 6$ and
$P(A|B_1)=1/2, P(A|B_2)=(1/2)(1/2)=1/4, P(A|B_3)=(1/2)(1/2)(1/2)=1/8$
$P(A|B_4)=(1/2)(1/2)(1/2)(1/2)=1/16$ similarly
$P(A|B_5)=1/32, P(A|B_6)=1/64$.

a) $$P(A)=\sum_{i=1}^{6} P(B_i)P(A|B_i)$$
$$=\frac{1}{6}\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+\frac{1}{64}\right)$$
$$=\frac{63}{6.64}=\frac{21}{128}$$

b) Since
$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{P(A)};\ i=1,2,3,4,5,6 \text{ thus}$$

$$P(B_1|A)=\frac{P(B_1)P(A|B_1)}{P(A)}=\frac{\frac{1}{6}\cdot\frac{1}{2}}{\frac{21}{128}}=\frac{32}{63}$$

$$P(B_2|A)=\frac{P(B_2)P(A|B_2)}{P(A)}=\frac{\frac{1}{6}\cdot\frac{1}{4}}{\frac{21}{128}}=\frac{16}{63}$$

$$P(B_3|A)=\frac{P(B_3)P(A|B_3)}{P(A)}=\frac{\frac{1}{6}\cdot\frac{1}{8}}{\frac{21}{128}}=\frac{8}{63}$$

$$P(B_4|A)=\frac{P(B_4)P(A|B_4)}{P(A)}=\frac{\frac{1}{6}\cdot\frac{1}{16}}{\frac{21}{128}}=\frac{4}{63}$$

$$P(B_5|A)=\frac{P(B_5)P(A|B_5)}{P(A)}=\frac{\frac{1}{6}\cdot\frac{1}{32}}{\frac{21}{128}}=\frac{2}{63}$$

$$P(B_6|A) = \frac{P(B_6)P(A|B_6)}{P(A)} = \frac{\frac{1}{6}\cdot\frac{1}{64}}{\frac{21}{128}} = \frac{1}{63}$$

E9) Let the events B_i = the chip is produced by line i, where $i = 1, 2, 3$ and event A = the chip is defective.

$P(B_1) = 0.40,\ P(B_2) = 0.25,\ P(B_3) = 0.35$

$P(A \mid B_1) = 0.05,\ P(A \mid B_2) = 0.06,\ P(A \mid B_3) = 0.03$

a) $P(A) = \sum_{i=1}^{3} P(B_i)P(A \mid B_i)$

$= (0.40)(0.05) + (0.25)(0.06) + (0.35)(0.03)$

$= 0.0455$

b) The chip was produced from the Line 3, given that the chip is defective

$$= P(B_3|A) = \frac{P(B_3)P(A|B_3)}{P(A)} = \frac{(0.35)\,(0.03)}{0.0455} = 0.2308.$$

E10) The sample space, as given below, consists of 36 equally likely outcomes.

$$S = \begin{Bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{Bmatrix}$$

a., b., c. Since random variable X = sum of scores on the two faces, Y = absolute difference of scores on the two faces. Clearly, X takes values 2, 3,..., 12 and Y takes values 0, 1, 2,..., 5. Joint p.m.f. $f_{X,Y}(x, y)$ of X and Y and their marginal p.m.f. may be very easily obtained from the bivariate probability table as shown in the following table:

(x, y)	0	1	2	3	4	5	$f_X(x)$
2	1/36						1/36
3		2/36					2/36
4	1/36		2/36				3/36
5		2/36		2/36			4/36
6	1/36		2/36		2/36		5/36
7		2/36		2/36		2/36	6/36
8	1/36		2/36		2/36		5/36
9		2/36		2/36			4/36
10	1/36		2/36				3/36
11		2/36					2/36
12	1/36						1/36
$f_Y(y)$	6/36	10/36	8/36	6/36	4/36	1/36	1

d) Conditional p.m.fs of Y, given X, X, $P[Y = y \mid X = x]$, are given in the table below:

	$P[Y=y\mid X=x]=\dfrac{f_{X,Y}(x,y)}{f_X(x)}$						Total
(x, y)	0	1	2	3	4	5	
2	1						1
3		1					1
4	1/3		2/3				1
5		1/2		1/2			1
6	1/5		2/5		2/5		1
7		1/3		1/3		1/3	1
8	1/5		2/5		2/5		1
9		1/2		1/2			1
10	1/3		2/3				1
11		1					1
12	1						1

e) Similarly, conditional pmf of X given Y, can be obtained.

f) Since $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$. say for $x=2,\ y=0$, therefore X and Y are not independent.

E11) The marginal probability density function of X

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

$$=\int_{x}^{2}(x+y)/4\,dy$$

$$=\frac{1}{4}\left[xy+\frac{y^2}{2}\right]_{y=x}^{y=2}=\frac{1}{8}(4+4x-3x^2)\ ,\quad \text{if } 0<x<2$$

and $f_X(x)=0$, otherwise

The marginal probability density function of Y

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

$$=\int_{0}^{y}(x+y)/4\,dx$$

$$=\frac{1}{4}\left[xy+\frac{x^2}{2}\right]_{x=0}^{x=y}=\frac{3}{8}y^2\ ,\quad \text{if } 0<y<2$$

and $f_Y(y)=0$, otherwise

a) The conditional probability density function of X, given $Y=y,\ 0<y<2$ will be

$$f_{X|Y}(x\mid y)=\frac{f_{X,Y}(x,y)}{f_Y(y)},\quad -\infty<x<\infty=\frac{(x+y)/4}{(3/8)y^2},\quad 0<x<y$$

$f_{X|Y}(x/y)=0$, otherwise

b) The conditional probability density function of Y, given $X=x,\ 0<x<2$ will be

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \qquad -\infty < y < \infty$$

$$= \frac{(x+y)/4}{(4+4x-3x^2)/8}, \qquad x < y < 2$$

$$= 0 \qquad \text{otherwise}$$

c) Since $f_{X,Y}(x,y) \neq f_X(x)f_Y(y),\ 0 < x < y < 2$, the random variables X and Y are not independent.

E12) The marginal probability density function of X

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

$$= \int_0^{\infty} e^{-(x+y)}dy$$

$$= \left[-e^{-(x+y)}\right]_{y=0}^{y=\infty} = e^{-x} \qquad \text{if } 0 < x < \infty$$

and $\quad f_X(x) = 0 \qquad$ otherwise

The marginal probability density function of Y

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

$$= \int_0^{\infty} e^{-(x+y)}dx$$

$$= \left[-e^{-(x+y)}\right]_{x=0}^{x=\infty} = e^{-y} \qquad \text{if } 0 < y < \infty$$

and $\quad f_Y(y) = 0 \qquad$ otherwise

a) The conditional probability density function of X, given $Y = y$, will be

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad -\infty < x < \infty$$

$$= \frac{e^{-(x+y)}}{e^{-y}} = e^{-x}, \qquad 0 < x < \infty$$

$$= 0 \qquad \text{otherwise}$$

b) The conditional probability density function of Y, given $X = x$, will be

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \qquad -\infty < y < \infty$$

$$= \frac{e^{-(x+y)}}{e^{-x}} = e^{-y}, \qquad 0 < y < \infty$$

$$= 0 \qquad \text{otherwise}$$

c) $P(X < 1) = \int_0^1 e^{-x}\,dx$

$$= [-e^{-x}]_0^1 = 1 - e^{-1}$$

$$P(X<Y)=\int_{-\infty}^{\infty}\int_{x}^{\infty} f_{X,Y}(x,y)\,dx\,dy$$

$$=\int_{0}^{\infty}\int_{x}^{\infty} e^{-(x+y)}\,dx\,dy$$

$$=\int_{0}^{\infty}[-e^{-(x+y)}]_{y=x}^{y=\infty}\,dx$$

$$=\int_{0}^{\infty}[-e^{-(x+y)}]_{y=x}^{y=\infty}\,dx$$

$$=\int_{0}^{\infty} e^{-2x}\,dx=\left[\frac{e^{-2x}}{-2}\right]_{x=0}^{x=\infty}=\frac{1}{2}.$$

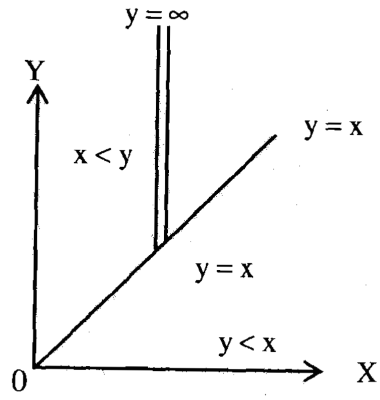

Fig.5

$$P(X+Y<1)=\int_{0}^{1}\int_{0}^{1-x} f_{X,Y}(x,y)\,dx\,dy$$

$$=\int_{0}^{1}\int_{0}^{1-x} e^{-(x+y)}\,dx\,dy$$

$$=\int_{0}^{1}[-e^{-(x+y)}]_{y=0}^{y=1-x}\,dx$$

$$=\int_{0}^{1}(e^{-x}-e^{-1})\,dx$$

$$=[-e^{-x}-xe^{-1}]_{x=0}^{x=1}=1-2e^{-1}.$$

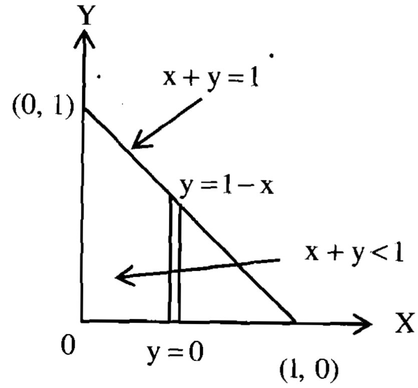

Fig.6

E13) $f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$

$$=\int_{x}^{1} 2\,dy$$

$$=2[x]_{y=x}^{y=1}=2(1-x) \qquad \text{if } 0<x<1$$

and $f_X(x)=0$ otherwise

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

$$=\int_{0}^{y} 2\,dx$$

$$=2[x]_{x=0}^{x=y}=2y \qquad \text{if } 0<y<1$$

and $f_Y(y)=0$ otherwise

a) Since $f_{X,Y}(x,y)\neq f_X(x)f_Y(y)$, $0<x<y<1$, the random variables, X and Y, are not independent.

b) $P[X<0.2|Y>0.1]=\dfrac{P[X<0.2,Y>0.1]}{P[Y>0.1]}$

$$P[Y > 0.1] = \int_{0.1}^{\infty} f_Y(y)\,dy$$

$$= \int_{0.1}^{1} 2y\,dy = \left[y^2\right]_{0.1}^{1} = 0.99$$

$$P[X<0.2, Y>0.1] = \int_{0}^{0.1}\int_{0.1}^{1} 2\;dxdy + \int_{0.1}^{0.2}\int_{x}^{1} 2\;dxdy$$

$$= \int_{0}^{0.1} [2y]_{0.1}^{1}\,dx + \int_{0.1}^{0.2} [2y]_{x}^{1}\,dx$$

$$= 1.8\int_{0}^{0.1} dx + 2\int_{0.1}^{0.2} (1-x)\,dx$$

$$= 0.18 + 2\left[x - \frac{x^2}{2}\right]_{0.1}^{0.2} = 0.35$$

$$P[X<0.2 \mid Y>0.1] = \frac{P[X<0.2, Y>0.1]}{P[Y>0.1]} = \frac{0.35}{0.99} = 0.3535$$

$$P[0.1<Y<0.4] = \int_{0.1}^{0.4} f_Y(y)\,dy$$

$$= \int_{0.1}^{0.4} 2y\,dy = \left[y^2\right]_{0.1}^{0.4} = 0.15$$

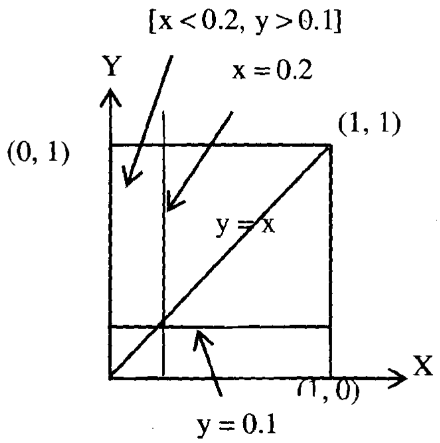

Fig.7

E14) Let the joint probability density function of X and Y be as follows

$$f_{X,Y}(x,y) = \begin{cases} k; & -6 < x, y < 6 \\ 0 & \text{elsewhere} \end{cases}$$

Since,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dxdy = \int_{-6}^{6}\int_{-6}^{6} k\,dxdy = 1$$

therefore, $k = 1/144$

Again,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

$$= \int_{-6}^{6} (1/144)\,dy$$

$$= 12(1/144) = 1/12 \qquad \text{for} -6 < x < 6$$

(i) $$E(Y \mid X = x) = \int_{-\infty}^{\infty} y \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy$$

$$= \int_{-6}^{6} y\frac{1/144}{1/12}\,dy = \frac{1}{12}\left[\frac{y^2}{2}\right]_{-6}^{6}$$

$$= 0$$

(ii) We may get $E(X \mid Y = y) = 0$ in similar way.

(iii) $V(Y \mid X = x) = E(Y^2 \mid X = x) - \{E(Y \mid X = x)\}^2$

$$= \int_{-\infty}^{\infty} y^2 \frac{f_{X,Y}(x,y)}{f_X(x)} dy - 0$$

$$= \int_{-6}^{6} y^2 \frac{1/144}{1/12} dy = \frac{1}{12}\left[\frac{y^3}{3}\right]_{-6}^{6}$$

$$= \frac{1}{36}(432) = 12$$

(iv) We may get $V(X \mid Y = y) = 12$ in similar way.

E15) Since

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = \int_0^1\int_0^x kxydxdy$$

$$= \int_0^1 kx\left[\frac{y^2}{2}\right]_0^x dx = k\int_0^1 \frac{x^3}{2}dx$$

$$= \frac{k}{2}\left[\frac{x^4}{4}\right]_0^1 = \frac{k}{8} = 1$$

therefore

(i) $k = 8$

(ii) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$

$$= \int_0^x 8xydy = 4x^3, \qquad 0 < x < 1$$

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y\frac{f_{X,Y}(x,y)}{f_X(x)}dy$$

$$= \int_0^x y\frac{8xy}{4x^3}dy = \frac{2}{x^2}\left[\frac{y^3}{3}\right]_0^x = \frac{2}{3}x$$

(iii) $E(Y^2 \mid X = x) = \int_{-\infty}^{\infty} y^2 \frac{f_{X,Y}(x,y)}{f_X(x)}dy$

$$= \int_0^x y^2\frac{8xy}{4x^3}dy$$

$$= \frac{2}{x^2}\left[\frac{y^4}{4}\right]_0^x = \frac{1}{2}x^2$$

$$V(Y \mid X = x) = E(Y^2 \mid X = x) - \{E(Y \mid X = x)\}^2 = \frac{1}{2}x^2 - \frac{4}{9}(x)^2 = \frac{1}{18}x^2.$$

—x—

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,..., 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

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

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.

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

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

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

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

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

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

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

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

$$P[X_{n_k} = i_k, X_{n_{k-1}} = i_{k-1}, \ldots, X_{n_2} = i_2, X_{n_1} = i_1] = \{P[X_{n_k} = i_k \mid X_{n_{k-1}} = i_{k-1}] \ldots\ldots P[X_{n_2} = i_2 \mid X_{n_1} = i_1] P[X_{n_1} = i_1]\} \quad (5)$$

3. $$P[X_{n+1} = j, X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0] = P[X_{n+1} = j \mid X_n = i] \ldots\ldots P[X_1 = i_1 \mid X_0 = i_0] P[X_0 = i_0] \quad (6)$$

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

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

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.

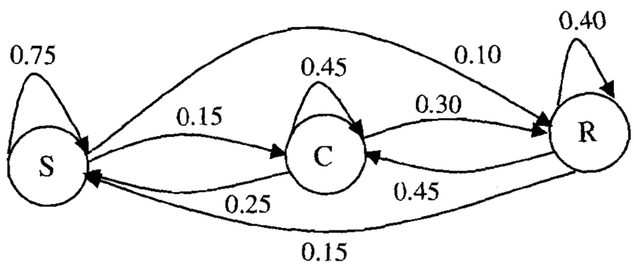

Fig.1

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

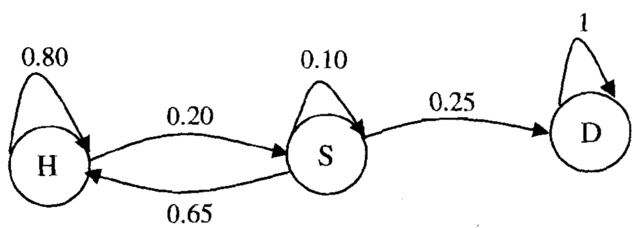

Fig.2

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

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

***

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.

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

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

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 $\mathbf{P}^{(0)} = \mathbf{I}$

***

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.

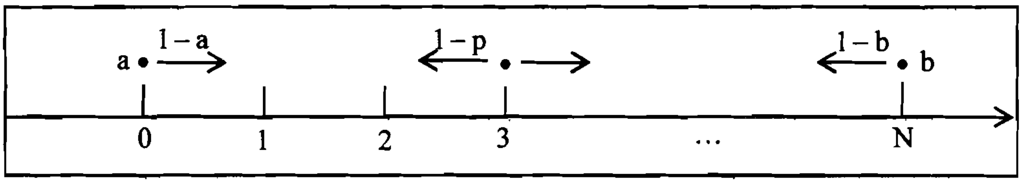

Fig.3: The Random Walk

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:

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

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

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

***

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

***

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.

$p_{0j} = P(X_n = j | X_{n-1} = 0) = 0$, when $j \neq 1$, $p_{0j} = P(X_n = 1 | X_{n-1} = 0) = 1$. 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.

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

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 $\mathbf{P}$ 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.

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

(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.

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

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

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

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

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

The **right eigen vector** $\mathbf{x_1}$ corresponding to $\lambda_1 = 1$, will satisfy following

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

Therefore, we have to solve the system of equations

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

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

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

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

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

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

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

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

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

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

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

and hence,

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

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

***

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.

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

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)

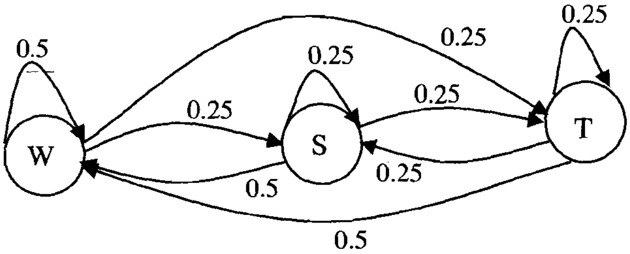

Fig.4

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

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—

UNIT 3 STATIONARY MARKOV CHAINS

Structure

3.1 INTRODUCTION

In Unit 2, we defined the Markov chain and its basic properties. In that unit, we limited our discussions only to the finite state Markov chains. Therefore, transition matrices were only of finite order. Here, in Unit 3, we will deal mostly with the Markov chain with countable states. Therefore, the transition matrices will, generally, be of infinite order. We will study the classification of states under various conditions. Mainly, we will gain knowledge of the limiting behaviour of the chain. Some chains stabilize after a long time. Their distributions become independent of the initial distribution of the chain. Due to this property, the limiting distribution is called the stationary distribution. We will learn the criterion under which the chains achieve the limiting distribution. We shall start the discussion in Sec. 3.2 with the classification of states of the Markov chains. Here, we will present the concepts of communication of states, closed set, and irreducibility. We will study about first passage time to the states and their expectations. In Sec. 3.3, we will present the concepts of recurrence and transience of states. We will develop some mechanism to identify states of the Markov chain. We will present some examples to illustrate these concepts. In Sec. 3.4, we will study the limiting behaviour of the chains. We will define stationary distributions and will study various conditions under which the chains will approach to the stationary distribution. In this unit, we will present various theorems without proofs.

Objectives

After studying this, unit you should be able to:

- classify and categorize the states of the Markov chain into communicating classes and closed sets;
- learn about first passage time to a state and time of first return (recurrence time) to a state;
- find the mean first passage time to the states, and mean time of first return (mean recurrence time) to the states;
- recognize the recurrent and transient states;
- understand the concept of Stationary Distribution and conditions for the existence of the limiting distribution of Markov chains.

3.2 CLASSIFICATION OF STATES

In this section, we will classify the states of the discrete time Markov chain on the basis of some of its transition properties. As in the previous unit, we will denote a Markov chain by a sequence $\{X_n\}$, satisfying "Markov Property" whose state space, S, is assumed to be discrete. The index set will also be denoted by T which is a discrete set. The transition probability matrix, $\mathbf{P}$, whose $(i,j)^{th}$ entry, p_{ij}, $(i, j \in S)$ denotes the probability of transition from the state i, to the state j in a step, or unit time. As in **Unit 2,** we assume that the transition probability in zero step is defined by $\mathbf{P}^{(0)} = (p_{ij}^{(0)}) = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix.

3.2.1 Irreducible Chains

Let us first discuss few definitions.

Definition 1: A state j in S is said to be **accessible** from the state i, if and only if, there exists a non-negative integer m, such that $p_{ij}^{(m)} > 0$. The symbol, $i \to j$, denotes this relation between states i and j.
Thus, if for all non-negative integers m, $p_{ij}^{(m)} > 0$, then state j is **not accessible** from i, and we will denote this by $i \not\to j$. When two states, i, and j are **accessible** to each other, then we say that the states i and j communicate with each other. In other words two states, i, and j are called **communicative,** if and only if, there exists integers m, n (≥ 0), such that $p_{ij}^{(m)} > 0, p_{ij}^{(n)} > 0$. The symbol $i \leftrightarrow j$ denotes the relation that i and j **communicate** with each other.

Definition 2: Let j be a state in the state space S of the Markov chain. Then a sub set $C(j)$ of S is called the **communicating class of** j if all the states in $C(j)$ communicate with j. Symbolically, given, $k, j \in S$, then $k \in C(j)$ if and only if $j \leftrightarrow k$.

We present below this theorem, without proof, stating a property of the communication among the states of a Markov chain. The theorem follows from the fact that the communication relation on the state space S is reflexive, symmetric, and transitive.

Theorem 1: The communication relation on the state space S is an equivalence relation.

Remark

(i) The relation of accessibility is neither reflexive, nor symmetric. However it is transitive. Since, $i \to j \Rightarrow \exists$ an integer m such that $p_{ij}^{(m)} > 0$, and $j \to k \Rightarrow \exists$ an integer n such that $p_{jk}^{(n)} > 0$,

From the Chapman-Kolomogorov equation, we have

$$p_{ik}^{(m+n)} = \sum_{l \in S} p_{il}^{(m)} p_{lk}^{(n)} > p_{ij}^{(m)} p_{jk}^{(n)} > 0 \tag{1}$$

Therefore, we find that state k is accessible from i. Thus, accessibility is transitive.

(ii) Let i, j are states in a Markov chain, and they communicate with each other, then they belong to same communicating class, that is $C(i) = C(j)$, if $i \leftrightarrow j$.

(iii) If C_1 and C_2 are two communicating classes in S, then either they will be equal, or they will be disjoint. The state space S can be partitioned into the equivalence classes by the communication relation as it is an equivalence relation. These equivalence classes are communicating classes.

Definition 3: A subset C of the state space S is called **closed** if it is a communicating class and no state outside C can be reached from any state within C. Symbolically, a subset C of the state space S will be called closed, if and only if, for any states j, k in S, $j \leftrightarrow k$, and for $j \in C$ and $k \notin C$, $p_{jk} = 0$.

Remark

(i) Let C be a closed set. Then, for any states j, k, in S, such that $j \in C$ and $k \notin C$, we have $p_{jk}^{(m)} = 0$ for all positive integers, m.

(ii) All the states within closed set C communicate with each other, but the reverse is not true, that is, if a set contains states that communicate with each other it does not imply that the set is closed. However, if all the states of the state space are communicable, then the state space will be closed.

(iii) A subset of states C is closed if for every state $i \in C$, $\sum_{j \in C} p_{ij} = 1$. In this case, the matrix $\mathbf{P}$ can be rearranged in the following canonical form, $\mathbf{P} = \begin{pmatrix} P_1 & 0 \\ R & Q \end{pmatrix}$ where 0 is a zero matrix. Here, the sub-matrix $\mathbf{P}_1 = (p_{ij}), i, j \in C$ is also stochastic, and states in C forms a sub-Markov chain.

(iv) If a closed set is a singleton, then the state comprising this set is called an **absorbing state,** i.e., a state j will be absorbing if, and only if, $p_{jj} = 1$, and $p_{jk} = 0$ for $k \neq j$.

(v) No proper subset of a closed set will be closed.

Definition 4 (Irreducible Chain): A Markov chain is called **irreducible** if there does not exist any closed set other than state space S, itself. If a chain is not irreducible then it is called reducible. Markov chains which are not irreducible are called **reducible**.

We should note that whether or not a Markov chain is irreducible is determined by the state space S, and the transition matrix (p_{ij}), the initial distribution, is irrelevant in this matter. If all the elements of the transition matrix (p_{ij}) are non-zero, then the Markov chain will necessarily be irreducible. All the off-diagonal elements of the transition matrix (p_{ij}) of an irreducible Markov chain can not be zero. In fact, no row can have all the off-diagonal elements zero.

Let us discuss the following example to understand various Markov chains.

Example 1: Let a Markov chain with the state space $S = \{0, 1, 2, 3, 4, 5\}$ with the following transition matrix:

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1/5 & 1/5 & 1/10 & 1/10 & 1/5 & 1/5 \\ 0 & 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

For set of states $C_1 = \{0, 1\}$, the states are communicating with each other.

Since $p_{01} = 1$, $p_{00}^{(2)} = p_{01}p_{10} = 1$

$p_{10} = 1$, $p_{11}^{(2)} = p_{10}p_{01} = 1$, and for state 0, we have $p_{0j} = 0$ for all $j \notin C_1$

for state 1, we have $p_{1j} = 0$ for all $j \notin C_1$

therefore, $C_1 = \{0, 1\}$ is a **closed set** of the given Markov chain.

Similarly, we may show that sets $C_2 = \{3, 4\}$ is a **communicable** set, and the sets outside it are not accessible from C_2 thus, C_2 is **closed set.**

Here, the state 5 is an absorbing state since the set $\{5\}$ is closed and it is a singleton. It may be verified that the sub matrices formed by the closed sets are stochastic as follows.

We can verify for $C_1 = \{0, 1\}$

for state $0 \in C_1$, $\sum_{j \in C_1} p_{0j} = p_{00} + p_{01} = 1$

for state $1 \in C_1$, $\sum_{j \in C_1} p_{1j} = p_{10} + p_{11} = 1$.

We can also verify it for other closed sets $C_2 = \{3, 4\}$ and $C_3 = \{5\}$.

The transition matrix can also be rearranged in following canonical form.

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 3 \\ 4 \\ 5 \\ 2 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 3 & 4 & 5 & 2 \end{array} \\ \left(\begin{array}{cc|cc|c|c} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 1/5 & 1/5 & 1/10 & 1/10 & 1/5 & 1/5 \end{array}\right) \end{array} = \left(\begin{array}{ccc|c} \mathbf{P_1} & \mathbf{0} & \mathbf{0} & \\ \mathbf{0} & \mathbf{P_2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P_3} & \\ \hline \mathbf{R} & & & \mathbf{Q} \end{array}\right)$$

where $\mathbf{P_1}, \mathbf{P_2}, \mathbf{P_3}$ are sub matrices of $\mathbf{P}$ corresponding to the three closed sets, $\mathbf{0}$ are zero matrices, $\mathbf{Q}$ is sub matrix corresponding to the transient state and $\mathbf{R}$ is remaining sub matrix.

The Markov chain is reducible since it has three closed sets and a transient set.

***

3.2.2 First Return and First Passage Probabilities

Let i be any state of a time homogeneous Markov chain $\{X_n\}$. Define

$$f_{ii}^{(n)} = P[X_n = i, X_{n-1} \neq i, \ldots, X_1 \neq i \mid X_0 = i], \qquad (2)$$

Thus, $f_{ii}^{(n)}$ is the probability that the chain starting in state i returns to state i for the first time after n steps. Clearly, $f_{ii}^{(1)}$ - p_{ii}, and we define $f_{ii}^{(0)} = 0$, for all states i in state space S. We call $f_{ii}^{(n)}$, the **probability of first return** (also called time of first recurrence) to state i, in time n.

Similarly, we may define, the **probability of first passage** from state i to state $j, i \neq j$ in time n denoted by $f_{ij}^{(n)}$ as

$$f_{ij}^{(n)} = P[X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = i], \qquad (3)$$

Thus, $f_{ij}^{(n)}$ is the probability that the chain starting in state i and visits the j for the first time after n steps. Clearly, $f_{ij}^{(1)} = p_{ij}$, and we now define $f_{ij}^{(0)} = 0$ for all i, j in S. As defined in **Unit** 1, $\mathbf{P}^{(0)} = \mathbf{I}$, i.e.

$$p_{jj}^{(0)} = 1 \text{ and } p_{jk}^{(0)} = 0 \text{ for } k \neq j \text{ for all } j, k \text{ in } S.$$

We present below a theorem without proof, which provides two equations: the first, a relationship between $f_{ii}^{(n)}$, the **probability of first return** to state i in time n and $p_{ii}^{(n)}$, the n-step transition probability from state i to itself, and the second relates the **probability of first passage** from state i to state j in time n given by $f_{ij}^{(n)}$ and the n-step transition probability from state i to state j given by $p_{ij}^{(n)}$. These relations may help in computation of n-step transition probabilities and in proving results on limiting behaviors of states of Markov chain.

Theorem 2: For any state i in S, we have

$$p_{ii}^{(n)} = \sum_{r=0}^{n} f_{ii}^{(r)} p_{ii}^{(n-r)}$$

and for any two states i and j in S, we have

$$p_{ij}^{(n)} = \sum_{r=0}^{n} f_{ij}^{(r)} p_{jj}^{(n-r)} \tag{4}$$

Definition 5: Assume that a time homogeneous Markov chain starts in state i, and define

$$f_{ii} = \sum_{n=0}^{\infty} f_{ii}^{(n)} \tag{5}$$

Then f_{ii} is the probability of ultimate or eventual return to the state i, having started in this state, i.e., the probability that the chain ever returns to the state i. A state i is called a **recurrent** state or **persistent** state if $f_{ii} = 1$, i.e., when the return to the state i is certain. We will use both the terms **recurrent** and **persistent** for this purpose in this unit. A state i is called **transient** when the ultimate, or eventual, return to the state i is not certain, i.e., $f_{ii} < 1$.

Definition 6 (Mean Recurrence Time): Let i be a recurrent (persistent) state, then $f_{ii}^{(n)}$, $n = 0, 1, 2, 3, \ldots$ is a probability distribution of time to return to state i and the mean of this distribution is defined as

$$\mu_{ii} = \sum_{n=0}^{\infty} n\, f_{ii}^{(n)}. \tag{6}$$

m_i is called the mean recurrence time of the state i. A recurrent state i is called **non = null recurrent** (also called positive recurrent, or positive persistent) if $\mu_{ii} < \infty$, i.e., if its mean recurrence time is finite, whereas it is called **null recurrent** if $\mu_{ii} = \infty$, i.e., if its mean recurrence time is infinite.

Definition 7 (Recurrent Periodic State): A recurrent state i is called **periodic** if the return to the state i can occur only in the t^{th}, $2t^{th}$, …, step where t is an integer greater than 1. In such a case, the integer t is called the **period** of the **periodic** state. Symbolically,

$$t = \text{G.C.D.}\{m : p_{ii}^{(m)} > 0\} = \text{G.C.D.}\{m : f_{ii}^{(m)} > 0\} \tag{7}$$

(Here, $f_{ii}^{(m)}$ is the probability of first return to the state i in m steps, defined in the Equation (2).)

If there does not exist such a $t(>1)$ then the recurrent state i is not periodic, and this state is called **aperiodic,** i.e., if $t=1$, then the state is said to be **aperiodic.** Every state in a communicating class must have the same period. Thus, in an irreducible Markov chain all states are either aperiodic or periodic with same period. An irreducible Markov chain is said to be **aperiodic** if its all states are aperiodic, and the irreducible Markov chain is said to be **periodic,** having period $t(>1)$, if its all states are periodic with period $t(>1)$. In an irreducible chain, all the states are of the same type. In fact, we have the following important results.

Theorem 3 (Recurrence a Class Property): Let two states, i and j, in state space S, $i \leftrightarrow j$, (that is, both states are in the same communicating class), then both the states are either transient, both are persistent null, or, both are persistent non-null together. Both are aperiodic or periodic with same period. Thus, all the states in a communicating class have the same classification. Either all are transient, or non-null persistent, or null persistent. All are aperiodic, or periodic with the same period.

***

Corollary 1: In an irreducible chain, all the states are either transient, all are persistent null, or all are persistent non-null together. If all are periodic then all will have the same period.

Definition 8 (Passage Time): Parallel to the recurrence time, now we define the passage time. Firstly, define

$$f_{ij} = \sum_{n=0}^{\infty} f_{ij}^{(n)}, \tag{8}$$

the probability that the chain starting in state i will ever reach the state j, i.e., the probability of ultimate passage from state i to j. If $f_{ij}=1$, then the ultimate passage to state j is certain given that the chain starts in the state i. In such a case, $f_{ij}^{(n)}$, $n=0, 1, 2, 3, \ldots$ is the probability distribution of first passage time to the state j given that the chain starts from i. Then, we may define the mean of the first passage time from the state i to state j as,

$$\mu_{ij} = \sum_{n=0}^{\infty} n\, f_{ij}^{(n)} \tag{9}$$

Definition 9 (Recurrent Chain): A Markov chain is called recurrent, or persistent, if all its states are recurrent

.

Transient Chain: A Markov chain is called transient if all its states are transient.
Ergodic State and Ergodic Chain: A persistent, non-null, aperiodic state of a Markov chain is called **ergodic** state. If all states in a Markov chain are ergodic, then the chain is said to be ergodic.

Let us discuss the following example.

Example 2: Let a Markov chain with state space $S=\{1, 2, 3, 4, 5\}$ have the following transition matrix. We will determine the nature of the states of the chain.

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/2 & 0 \\ 0 & 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

On the basis of the probability of first return to the states, we will classify the states as follows. Since,

$$f_{11} = f_{11}^{(1)} + f_{11}^{(2)} + f_{11}^{(3)} + \cdots$$
$$= 0 + 1.1 + 0 + 1\ldots = 1$$

therefore, state 1 is **persistent.** Again

$$f_{22} = f_{22}^{(1)} + f_{22}^{(2)} + f_{22}^{(3)} + \cdots$$
$$= 0 + 1.1 + 0 + \ldots = 1$$

therefore, state 2 is **persistent.** Similarly,

$$f_{33} = f_{33}^{(1)} + f_{33}^{(2)} + f_{33}^{(3)} + \cdots$$
$$= 1/4 + 0 + 0 + \ldots = 1/4$$

therefore, state 3 is **transient** and

$$f_{44} = f_{44}^{(1)} + f_{44}^{(2)} + f_{44}^{(3)} + \cdots$$
$$= \frac{1}{4} + \frac{3}{4}.1 + 0 + \ldots = 1$$

therefore, state 4 is **persistent**. Finally,

$$f_{55} = f_{55}^{(1)} + f_{55}^{(2)} + f_{55}^{(3)} + \cdots$$
$$= 0 + 1.\frac{3}{4} + 1.\frac{1}{4}\frac{3}{4} + 1.\frac{1}{4}\frac{1}{4}\frac{3}{4} + 1.\frac{1}{4}\frac{1}{4}\frac{1}{4}\frac{3}{4}. + \ldots = 1$$

therefore, state 5 is **persistent.**

The states 1 and 2 are **periodic** with period 2 since, for state 1

$$t = \text{G.C.D.}\{m : f_{11}^{(m)} > 0\} = \text{G.C.D.}(2) = 2$$

and for state 2

$$t = \text{G.C.D.}\{m : f_{22}{}^{(m)} > 0\} = \text{G.C.D.}(2) = 2$$

The **Mean Recurrence Time** of the persistent (recurrent) states are obtained as follows:

$$\mu_{11} = 1.f_{11}^{(1)} + 2.f_{11}^{(2)} + 3.f_{11}^{(3)} + \ldots$$
$$= 1.0 + 2.1 + 0.0 = 2$$
$$\mu_{22} = 1.f_{22}^{(1)} + 2.f_{22}^{(2)} + 3.f_{22}^{(3)} + \cdots$$
$$= 1.0 + 2.1 + 0.0 = 2$$
$$\mu_{44} = 1.f_{44}^{(1)} + 2.f_{44}^{(2)} + 3.f_{44}^{(3)} + \cdots$$
$$= 1.\frac{1}{4} + 2.\frac{3}{4} + 3.0 + \cdots = \frac{7}{4}$$
$$\mu_{55} = 1.f_{55}^{(1)} + 2.f_{55}^{(2)} + 3.f_{55}^{(3)} + \cdots$$
$$= 1.0 + 2.\frac{3}{4} + 3.\frac{3}{16} + \cdots = \frac{7}{3}$$

The states {4, 5} are persistent, non-null, and aperiodic. Therefore, they are ergodic. The states {1, 2} are persistent and periodic with period 2. The state 3 is transient. It may be easily verified that the given Markov chain is reducible. Its state space can be decomposed into three communicating classes $C_1 = \{1, 2\}$, $C_2\{4, 5\}$ and $C_3 = \{3\}$. Further, C_1, C_2 are closed sets. At states in C_1 are aperiodic and positive recurrent. Whereas all states in C_2 are positive recurrent and periodic, each with period 2. This verifies the results of Theorem 3, and the fact that periodicity is a class property.

***

Example 3: Let a Markov chain have following transition matrix.

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

All the states are communicable. Therefore, it has only one closed set, the state space $S = \{0, 1, 2\}$. The chain is irreducible.

The probability of ultimate return to state 0 will be

$$f_{00} = \sum_{n=0}^{\infty} f_{00}^{(n)} = \frac{1}{2} + \frac{1}{2}\cdot\frac{3}{4} + \frac{1}{2}\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4} + \frac{1}{2}\cdot\frac{1}{4}\cdot 1\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4} + \cdots$$
$$= \frac{7}{8} + \frac{3}{32}\cdot\frac{4}{3} = 1.$$

Thus, the state 0 is persistent (recurrent).

The Mean Recurrence Time for state 0 will be

$$\mu_{00} = 1.f_{00}^{(1)} + 2.f_{00}^{(2)} + 3.f_{00}^{(3)} + \ldots$$
$$= 1\cdot\frac{1}{2} + 2\cdot\frac{1}{2}\cdot\frac{3}{4} + 3\cdot 0 + 4\cdot\frac{1}{2}\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4} + \cdots = \frac{11}{6}$$

Thus, 0 is a non-null persistent (positive recurrent state). Since the Markov chain is irreducible, all its states must be non-null persistent by Theorem 3. Let us verify this by actual calculation for other states in S.

The probabilities may also be obtained using a digraph, described in **Unit 2**. The digraph for the given transition matrix has been shown below, in Fig 1. To find $f_{11}^{(1)}$,

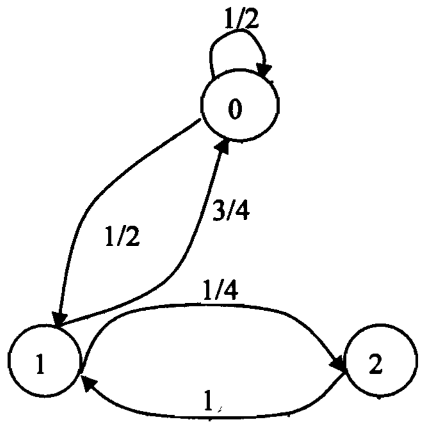

Fig.1

the probability of first return to state 1 in one step, find the paths from node 1 to node 1, traveling any edge only once. Add all the probability labels on the edges of these paths. There is no such path in this example, and the probability $f_{11}^{(1)}$ will be zero. To find $f_{11}^{(2)}$, the probability of first return to state 1 in two step, find the paths from node 1 to node 1 traveling along two distinct edges. We have two paths $1 \to 0 \to 1$ and $1 \to 2 \to 1$. Multiply probability labels on the edges of each path, and add such multiples of all paths to get $f_{11}^{(2)}$. Therefore, $f_{11}^{(2)} = \frac{3}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot 1 = \frac{5}{8}$ and so on. We get the probability of ultimate return to state 1 as

$$f_{11} = \sum_{n=0}^{\infty} f_{11}^{(n)} = 0 + 0 + \frac{3}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot 1 + \frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \cdots = \frac{1}{4} + \frac{3}{4}\cdot\frac{1}{2}\cdot 2 = 1.$$

Similarly, we may obtain

$$f_{22} = \sum_{n=0}^{\infty} f_{22}^{(n)} = 1.$$

Therefore, all the state are persistent (recurrent) as stated above.

The Mean Recurrence Time for state 1 will be

$$\mu_{11} = 1.f_{11}^{(1)} + 2.f_{11}^{(2)} + 3.f_{11}^{(3)} + \ldots$$

$$= 1\cdot 0 + 2\cdot\left(\frac{3}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot 1\right) + 3\cdot\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} + 4\cdot\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \cdots = \frac{1}{2} + \frac{9}{4} = \frac{11}{4}$$

Similarly, we may obtain

$$\mu_{22} = \frac{11}{1}$$

Thus, all the states are non-null since mean recurrence times for all the states are finite, as stated above.

Again, consider the state 0. Since

$$f_{00}^{(1)} = \frac{1}{2} > 0,\ f_{00}^{(2)} = \frac{1}{2}\cdot\frac{3}{4} > 0,\ f_{00}^{(3)} = 0,\ f_{00}^{(4)} = \frac{1}{2}\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4} > 0$$

Therefore, from the definition of periodic recurrent states given in Egn. (7), the period $t = \text{G.C.D.}\{m : f_{ii}^{(m)} > 0\} = \text{G.C.D.}\{1, 2, 4, \ldots\} = 1$.

Therefore, the state 0 is aperiodic. Since the chain is irreducible, all the states will be aperiodic.

Therefore, all the state are persistent (recurrent), aperiodic, and non-null and thus, ergodic. Thus the chain will be ergodic. We have, thus, verified that periodicity, positive or null recurrence, transience, etc., are class properties.

Example 4: Consider a Markov chain with the following transition matrix

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

Since all the states are communicable, it has only one closed set, sample, space $S = \{0, 1, 2, 3\}$. The chain is irreducible.

We can use the following digraph for the given transition matrix to compute the probabilities of first return, as in the previous example.

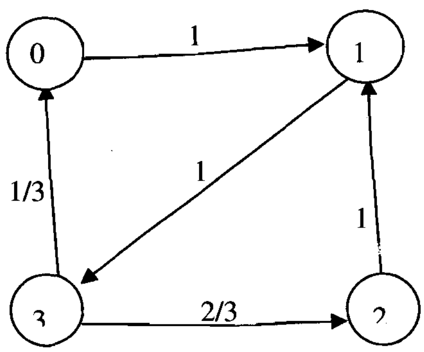

Fig.2

For the state 0,

$$f_{00}^{(1)} = f_{00}^{(2)} = 0,\ f_{00}^{(3)} = 1.1.\frac{1}{3} = \frac{1}{3} > 0,\ f_{00}^{(4)} = f_{00}^{(5)} = 0,\ f_{00}^{(6)} = \frac{2}{3}.\frac{1}{3} = \frac{2}{9} > 0, \cdots$$

therefore, from the definition of periodic recurrent states given in Eqn. (7), the period $t = \text{G.C.D.}\{m : f_{ii}^{(m)} > 0\} = \text{G.C.D.}\{3, 6, 9, \ldots\} = 3$ and probability of ultimate return to the state 0 is

$$f_{00} = \sum_{n=0}^{\infty} f_{00}^{(n)} = \frac{1}{3} + \frac{2}{9} + \frac{4}{.27} + \cdots = 1$$

Thus, the state 1 is recurrent with period 3. Now, since Markov chain is irreducible, all the other states have the same classification, that is, recurrent with period 3.

Therefore the Markov chain is recurrent and periodic with period 3.

***

You may now try the following exercises.

E1) Determine the classes, probability of ultimate return to the states, mean recurrence time of the various states of the Markov chain having the following transition matrix. Is the chain irreducible?

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

E2) Determine the closed set, probability of ultimate return to the states, periodicity of states, mean recurrence time of the states of the Markov chain having the following transition matrix. Is the chain irreducible?

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

So far we have discussed the classification of states and chains. In this section, we will focus on recurrence and transience in details.

3.3 RECURRENCE AND TRANSIENCE

We have discussed formal definitions of recurrence and transience in the previous section. In this section, we will provide some more properties of recurrence and transience without proof which will be used to classify the states of a Markov chain.

Let us begin with the formal definition of a generating function.

Definition 10 (Generating Function): Let $a_0, a_1, a_2, a_3, \ldots$ be a sequence of real numbers, and s be a real number, then a function $A(s)$ defined by

$$A(s) = a_0 + a_1 s + a_2 s^2 + \cdots = \sum_{k=0}^{\infty} a_k s^k$$

is called a **generating function** of the sequence $a_0, a_1, a_2, a_3, \ldots$ provided this power series converges in some interval $-s_0 < s < s_0$. If a non-negative discrete random variable X assumes only integral values $0, 1, 2, 3, \ldots$ and the sequence $\{a_k\}$ represents the probability distribution of X, such that $a_k = P[X = k]$, then $A(s)$ is called the probability generating function of random variable X.

Theorem 4: For a state i of a Markov chain, let $P_{ii}(s)$ be the generating function of the sequence $\{p_{ii}^{(n)}\}$, and $F_{ii}(s)$ be the generating function of the sequence $\{f_{ii}^{(n)}\}$. Then, we have

$$P_{ii}(s) = \frac{1}{1-F_{ii}(s)}, |s|<1 \tag{10}$$

***

Theorem 5: For state i, j of a Markov chain, let $P_{ij}(s)$ be the generating function of the sequence $\{p_{ij}^{(n)}\}$, $P_{jj}(s)$ be the generating function of the sequence $\{p_{jj}^{(n)}\}$, and $F_{ij}(s)$ be the generating function of the sequence $\{f_{ij}^{(n)}\}$. Then, we have for $|s|<1$

(i) $P_{ij}(s) = F_{ij}(s)P_{jj}(s)$ (11)

(ii) $P_{ij}(s) = F_{ij}(s)\,(1-F_{jj}(s))^{-1}$ (12)

***

Let us illustrate the following example to understand.

Example 5: Consider a Markov chain with the following transition matrix

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

We can verify that the matrix is periodic.

$$P = P^4 = \ldots = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \end{pmatrix} \end{array}, \quad P^2 = P^5 = \ldots = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

$$P^3 = P^6 = \ldots = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 1/3 & 0 & 2/3 & 0 \\ 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

For a state 0 of the Markov chain, the generating function of the sequence of the transition probabilities $\{p_{00}^{(n)}\}$ is given by

$$P_{00}(s) = \sum_{k=0}^{\infty} p_{00}^{(k)} s^k = 1 + 0.s + 0.s^2 + \frac{1}{3}s^3 + \cdots, \text{ since } p_{00}^{(0)} = 1$$

$$= 1 + \frac{s^3/3}{1-s^3} \text{ for } |s| < 1$$

and the generating function of the sequence of the probabilities of first return $\{f_{00}^{(n)}\}$ (as obtained in Example 4), will be $F_{00}(s)$ as given below

$$F_{00}(s) = \sum_{k=0}^{\infty} p_{00}^{(k)} s^k = 1 + 0.s + 0.s^2 + \frac{1}{3}s^3 + \cdots \text{ since } p_{00}^{(0)} = 1$$

$$= 1 + \frac{s^3/3}{1-s^3} \text{ for } |s| < 1$$

Therefore,

$$1-F_{00}(s)=\frac{1-s^3}{1-\frac{2}{3}s^3}$$ and thus, we may verify Egn. 10, that

$$P_{ii}(s)=\frac{1}{1-F_{ii}(s)},\ |s|<1$$ for state $i=0$. Similarly, we can verify the relations given in Eqn. (11) and (12) for the states of the Markov chain.

Here, we may note that $F_{00}(1)=\sum_{k=0}^{\infty} f_{00}^{(k)}=f_{00}$. Thus, the probability of ultimate return to the state 0 may be obtained from the generating function $F_{00}(s)$ by setting $s=1$.

***

Now, we are providing a theorem which states that a state i of the Markov chain will be recurrent if, and only if $\sum_{n=0}^{\infty} p_{ii}^{(n)}=\infty$. The result is immediate from the Eqn. (10), since for a recurrent state i as $s\uparrow 1$, $1-F_{ii}(s)\downarrow 0$ and therefore, the left hand side equation $P_{ii}(s)\to\sum_{n=0}^{\infty} p_{ii}^{(n)}$ and the right hand side $\frac{1}{1-F_{ii}(s)}$ tends to infinity as $s\uparrow 1$.

Theorem 6: A state i of a Markov chain will be persistent (recurrent) if

$$\sum_{n=0}^{\infty} p_{ii}^{(n)}=\infty \tag{13}$$

and state i will be transient if

$$\sum_{n=0}^{\infty} p_{ii}^{(n)}<\infty. \tag{14}$$

***

The following theorem gives some limiting results for recurrent states of a Markov chain.

Theorem 7: If the state i is persistent (recurrent) then as $n\to\infty$

(i) $p_{ii}^{(nt)}\to t/\mu_{ii}$ provided state i is non-null and periodic with period t (15)

(ii) $p_{ii}^{(n)}\to 1/\mu_{ii}$ provided state i is non-null and aperiodic (16)

(iii) $p_{ii}^{(n)}\to 0$ provided state i is null (whether periodic or aperiodic) (17)

***

Remark

(i) If state i is **transient**, then $p_{ii}^{(n)}\to 0$ as $n\to\infty$, since $\sum_{n=0}^{\infty} p_{ii}^{(n)}<\infty$. This result directly follows from the property of convergent series.

(ii) If state j is **transient** and i is any state of the Markov chain, then $\sum_{n=0}^{\infty} p_{ij}^{(n)}<\infty$ and $p_{ij}^{(n)}\to 0$ as $n\to\infty$.

(iii) If j is a **transient state**, then no matter where the Markov chain starts, it makes only a finite number of visits to state j, and the expected number of visits to j is finite. It may enter into a recurrent class in a number of steps and when it enters there, then it remains there for ever. On the other hand, if j is a **recurrent state**, then if the chain starts at j, it is guaranteed to return to j infinitely often and will eventually remain forever in the closed set containing state j. If the chain starts at some other state i, it might not be possible for it to ever visit state

j. If it is possible to visit the state j at least once, then it does so infinitely often (See Bhat, 2000).

Theorem 8: Let i and j be any state in state space s of a Markov chain.

(i) If the state j is persistent, null or transient, then

$$\lim_{n\to\infty} p_{ij}^{(n)} \to 0 \tag{18}$$

(ii) If the state j is persistent, non-null, and aperiodic, then

$$\lim_{n\to\infty} p_{ij}^{(n)} \to \frac{f_{ij}}{\mu_{jj}} \tag{19}$$

***

Example 6: Consider a countable state Markov chain having states $0, 1, 2, 3, \ldots$ with, transition probability matrix having (i, j) element $(i, j = 0, 1, 2, 3, \ldots)$ given by

$$p_{i0} = \frac{i+1}{i+2}, \; p_{ii+1} = \frac{1}{i+2} \text{ and } p_{ij} = 0, \;\; j \neq i+1 \text{ or } j \neq 0$$

Therefore, the transition probability matrix is an infinite matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ - \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & - \end{array} \\ \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & - \\ 2/3 & 0 & 1/3 & 0 & 0 & - \\ 3/4 & 0 & 0 & 1/4 & 0 & - \\ 4/5 & 0 & 0 & 0 & 1/5 & - \\ 5/6 & 0 & 0 & 0 & 0 & - \\ - & - & - & - & - & - \end{pmatrix} \end{array}$$

For the state 0, the probabilities of first return will be

$$f_{00}^{(1)} = \frac{1}{2}, \; f_{00}^{(2)} = \frac{1}{2}\cdot\frac{2}{3}, \; f_{00}^{(3)} = \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{3}{4}, \; f_{00}^{(4)} = \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{4}\cdot\frac{4}{5}, \cdots$$

Therefore, the probability of ultimate return to state 0, will be

$$f_{00} = \sum_{n=0}^{\infty} f_{00}^{(n)} = \frac{1}{2} + \frac{1}{2}\cdot\frac{2}{3} + \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{3}{4} + \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{4}\cdot\frac{4}{5} + \cdots$$

$$= \frac{1}{2} + \frac{2}{\lfloor\underline{3}} + \frac{3}{\lfloor\underline{4}} + \frac{4}{\lfloor\underline{5}} + \cdots = \frac{1}{2} + \left(\frac{3}{\lfloor\underline{3}} - \frac{1}{\lfloor\underline{3}}\right) + \left(\frac{4}{\lfloor\underline{4}} - \frac{1}{\lfloor\underline{4}}\right) + \left(\frac{5}{\lfloor\underline{5}} - \frac{1}{\lfloor\underline{5}}\right) + \cdots$$

$$= \frac{1}{2} + \left(\frac{1}{\lfloor\underline{2}} - \frac{1}{\lfloor\underline{3}}\right) + \left(\frac{1}{\lfloor\underline{3}} - \frac{1}{\lfloor\underline{4}}\right) + \left(\frac{1}{\lfloor\underline{4}} - \frac{1}{\lfloor\underline{5}}\right) + \cdots = 1$$

and, thus, the state 0 of the Markov chain is recurrent. Since all states can be reached from any state, hence, the Markov chain is irreducible. Again, the state 0 is aperiodic since the G.C.D. of times with positive probabilities of first return to the state 0 is one. From the class property of recurrence stated above, the Markov chain will be recurrent and aperiodic.

The mean recurrence time to the state 0, can be obtained as

$$\mu_{00} = \sum_{n=0}^{\infty} n\, f_{00}^{(n)} = 1\frac{1}{2} + 2\frac{1}{2}\cdot\frac{2}{3} + 3\frac{1}{2}\cdot\frac{1}{3}\cdot\frac{3}{4} + 4\frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{4}\cdot\frac{4}{5} + \cdots$$

$$= \frac{1}{2} + 2\left(\frac{3}{\lfloor\underline{3}} - \frac{1}{\lfloor\underline{3}}\right) + 3\left(\frac{4}{\lfloor\underline{4}} - \frac{1}{\lfloor\underline{4}}\right) + 4\left(\frac{4}{\lfloor\underline{5}} - \frac{1}{\lfloor\underline{5}}\right) + \cdots$$

$$= \frac{1}{2} + \left(\frac{2}{\lfloor\underline{2}} - \frac{2}{\lfloor\underline{3}}\right) + \left(\frac{3}{\lfloor\underline{3}} - \frac{3}{\lfloor\underline{4}}\right) + \left(\frac{4}{\lfloor\underline{4}} - \frac{4}{\lfloor\underline{5}}\right) + \cdots$$

$$= 1 + \frac{1}{\lfloor\underline{1}} + \frac{1}{\lfloor\underline{2}} + \frac{1}{\lfloor\underline{3}} + \frac{1}{\lfloor\underline{4}} + \cdots - 1 = e - 1$$

Thus, the state 0 is positive recurrent and, hence, the Markov chain is positive recurrent. Further, from the Eqn. (16), we have as $n \to \infty$,

$$p_{00}^{(n)} \to 1/\mu_{00} = 1/(e-1)$$

***

States in a Finite Markov Chain

The results obtained above in this section have some essential implications for the finite Markov chain. The state space of a finite Markov chain must contain at least one persistent state. Otherwise, if all the states of a Markov chain become transient then, the transition probabilities, $p_{ij}^{(n)} \to 0$ as $n \to \infty$ for all i and j in the state space S and it is impossible since for all $i \in S$ we must have $\sum_{j \in S} p_{ij}^{(n)} = 1$. Therefore, a Markov chain with a finite state space, S, cannot be a transient chain. Again, a finite Markov chain cannot have any null persistent state. Since the states of the closed set having this null persistent state will form a stochastic sub-matrix (say $\mathbf{P}_1$) of transition matrix P and as $n \to \infty$, we will have $\mathbf{P}_1^n \to 0$ and, hence, P will not remain stochastic. This is not possible. Thus, a finite Markov chain cannot have a null persistent state.

The following theorem is now easy to visualize.

Theorem 9: In a finite irreducible chain, all the states are non-null persistent.

Example 7: In Example 2, the state 3 is transient. We will find $\lim_{n\to\infty} p_{34}^{(n)}$ and $\lim_{n\to\infty} p_{35}^{(n)}$.

Let us find the probability of ultimate passage time from state 3 to state 4, and to state 5, i.e., f_{34} and f_{35}.

$$f_{34} = \sum_{n=0}^{\infty} f_{34}^{(n)}$$
$$= \frac{1}{2} + \frac{1}{2}\frac{1}{4} + \frac{1}{2}\frac{1}{4}\frac{1}{4} + \cdots = \frac{2}{3}$$

and

$$f_{35} = \sum_{n=0}^{\infty} f_{35}^{(n)}$$
$$= 0 + \frac{1}{2}\frac{3}{4} + \frac{1}{2}\frac{1}{4}\frac{3}{4} + \frac{1}{2}\frac{1}{4}\frac{1}{4}\frac{3}{4} + \cdots = \frac{3}{8}\frac{4}{3} = \frac{1}{2}$$

Again, since we have, from Example 2,

$$\mu_{44} = \frac{7}{4}, \; \mu_{55} = \frac{7}{3}.$$

Since, state 4 is aperiodic, non-null, persistent. Therefore, using Eqn. (19), we have as $n \to \infty$

$$p_{34}^{(n)} \to \frac{f_{34}}{\mu_{44}} = \frac{2}{3}\cdot\frac{4}{7} = \frac{8}{21}$$

and $$p_{35}^{(n)} \to \frac{f_{35}}{\mu_{55}} = \frac{1}{2}\cdot\frac{3}{7} = \frac{3}{14}.$$

***

Example 8 (Unrestricted Random Walk): Let us consider a particle moving along a straight line, and assuming integer values only. Let the random variable X_n denotes the position of the particle at time n. Then, X_n satisfies the relation, $X_n = X_{n-1} + Z_n$, where Z_n denotes the displacement of the particle at time n. We assume that the random variables Z_n, are identically and independently distributed with $P[Z_n = 1] = p$ and $P[Z_n = -1] = q$. It means that the particle either moves a unit in left direction with probability q, or a unit in right direction with probability p at each time. Therefore, { X_n } will be a Markov chain with the state space $\{0, \pm 1, \pm 2, \pm 3 \cdots\}$. Its transition probability matrix $\mathbf{P}$ can be expressed as

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

Since all the states are communicating with every other state, therefore the matrix and the chain is irreducible.

From Corollary 1, the chain is either transitive, or persistent null, or persistent non-null.

Consider the state 0. It is clear that we cannot return to 0 in an odd number of steps. Let it return to state 0 in time $2n$, then during this period it must have moved in right direction n times, and in left direction n times. Therefore, using binomial distribution we have

$$p_{00}^{(2n)} = {}^{2n}C_n p^n q^n$$
$$p_{00}^{(2n+1)} = 0$$

Now from Stirling's formula, a large n approximation to $\underline{|n}$ is

$$\underline{|n} = \sqrt{2\pi n}\,(n/e)^n$$

Using this approximation, we have

$$\begin{aligned} p_{00}^{(2n)} &= {}^{2n}C_n\, p^n q^n \\ &= \frac{\underline{|2n}}{\underline{|n}\,\underline{|n}}\, p^n q^n \\ &\approx \frac{\sqrt{4\pi n}\,(2n/e)^{2n}}{\sqrt{2\pi n}\,(n/e)^n \sqrt{2\pi n}\,(n/e)^n}(pq)^n \\ &\approx \frac{1}{\sqrt{\pi n}}(4pq)^n \end{aligned}$$

Now, as $\sum_{n=0}^{\infty} p_{00}^{(2n)} < \infty$ when $4pq < 1$ i.e., if $p \neq q$, in that case, the state 0 is transient.

Hence, the chain will be transient for $p \neq q$.

In the symmetric case, $p = q = 1/2$, so that $4pq = 1$, it follows that

$\sum_{n=0}^{\infty} p_{00}^{(2n)} = \sum_{n=0}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty$ and the state 0 is recurrent. Hence, the chain will be recurrent if $p = q$. Further, since, $p_{00}^{(2n)} = \frac{1}{\sqrt{\pi n}}(4pq)^n = \frac{1}{\sqrt{\pi n}} \to 0$ as $n \to \infty$ and the state 0 is recurrent, then by using Theorem 6, we may conclude that the chain will be recurrent null when $p = q = 1/2$.

***

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

E3) If n-step transition probabilities of a given state i of a Markov chain is defined as, $p_{ii}^{(n)} = \frac{1}{2}$, for $n > N_0$ then find the mean recurrence time for the state i.

E4) Consider a countable state Markov chain having a transition probability matrix as follows

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ - \end{array}\begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & - \end{array} \\ \begin{pmatrix} p & q & 0 & 0 & 0 & - \\ p & 0 & q & 0 & 0 & - \\ p & 0 & 0 & q & 0 & - \\ p & 0 & 0 & 0 & q & - \\ p & 0 & 0 & 0 & 0 & - \\ - & - & - & - & - & - \end{pmatrix} \end{array}, \text{ where } p+q=1 \text{ and } p>0.$$

Show that the chain is recurrent.

E5) Obtain the limiting value of $p_{ii}^{(n)}$ as $n \to \infty$ for $i = 0, 1, 2, 3$ for the Markov chain given in **E1**).

E6) Obtain the limiting value of $\mathbf{P}^n$ as $n \to \infty$ for the Markov chain given in **E3**).

In this section, we shall discuss stationary distribution.

3.4 STATIONARY DISTRIBUTION

In this section, we will study the behaviour of the Markov chain that has been running for a long time. In many real systems, the long term behaviour of the system requires much attention. Mostly, we need to study, whether the effect of the initial state diminishes in the long run. We may also want to study the proportions of time that the chain will be found in various states of the chain in a long run. Mathematically, this means studying the limit, $\lim_{n\to\infty} \mathbf{u}^{(n)}$, where $\mathbf{u}^{(n)}$ is the probability distribution of the chain at time n. We like to determine whether or not this limit exists, and, if it does then, to analyze the various related conditions and their implications. We will state here a few basic, but important facts regarding the limiting behaviour, mostly without proof, but with suitable examples.

Before discussing limits of $\mathbf{u}^{(n)}$, it is better to describe the notion of a stationary distribution of a Markov chain. We will say that the Markov chain $\{X_n\}$ possesses stationary distribution if the distribution $\mathbf{u}^{(n)}$ is the same for all n, that is, $\mathbf{u}^{(n)} = \mathbf{u}^{(0)} = \mathbf{u}$ the initial probability vector, for all $n \geq 1$. Thus, the probability that the chain is in, say, state i is the same for all time; although X_n is moving from one state to another, it looks statistically the same at any time. Since a stationary distribution of the chain does not depend on n, we drop the superscript and denote it merely by $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$. In general, if $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$ is a probability mass function, giving stationary distribution of a Markov chain $\{X_n\}$ with initial distribution $\mathbf{u} = \{u_1\ u_2 \cdots u_i \cdots\}$ where $u_i = P[X_0 = i]$ for each i and with the transition matrix $\mathbf{P} = (p_{ij})$, on the state space $S = \{1, 2, \ldots\}$, then $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$ is called a stationary distribution for the transition matrix $\mathbf{P}$. Here, we will make the study for countable state space. We will describe for the finite state space separately when the behaviour becomes different from the countable state space.

Definition 11 (Stationary Distribution): Let a Markov chain $\{X_n, n = 0, 1, 2, \ldots\}$ with transition probability matrix $\mathbf{P} = (p_{ij})$ (where p_{ij} denotes the transition probability from state i to j), state space $S = \{1, 2, \ldots\}$ and a sequence of non-negative numbers $\{\pi_j\}$ satisfying the following equations.

(i) $\pi_j = \sum_{i \in S} \pi_i p_{ij}$, for all $j \in S$ or $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ where $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots)$ (20)

(ii) $\sum_{i \in S} \pi_j = 1$ (21)

then the sequence $\{\pi_j\}$ will define a probability distribution over states of the Markov chain. This distribution is called the stationary distribution of the Markov chain.

Theorem 10: If the initial distribution of a Markov chain $\{X_n\}$ is the same as its stationary distribution, then all the random variables in the sequence, $\{X_n\}$, will have identical distributions.

Remark: Let π_j denote the probability that the system is in state j. The condition in Eqn.(20) is often called a **balancing equation**, or **equilibrium equation.** The stationary distribution $\boldsymbol{\pi}$ on S is such that if our Markov chain starts out with the initial distribution $\mathbf{u} = \boldsymbol{\pi}$, then we also have $\mathbf{u}^1 = \boldsymbol{\pi}$, since by Theorem 7 of Unit 2, and Eqn.(20) above, we have $u^{(1)} = uP = \pi P = \pi$. That is, if the distribution at time 0 is $\boldsymbol{\pi}$, then the distribution at time 1 is still $\boldsymbol{\pi}$. In general, $\mathbf{u}^{(n)} = \boldsymbol{\pi}$ for all n (for both finite as well as countable state space). Due to this reason $\boldsymbol{\pi}$ is called a **stationary distribution.**

Let us now discuss the stationary distribution for an Irreducible Aperiodic Markov Chain:

In this the existence of stationary distributions for the irreducible aperiodic Markov chains, and the long term behaviour of the distribution of these chains. The following theorems describe the related conditions. These theorems are applicable for both, finite as well as countable state space chains.

Theorem 11 (Ergodic Theorem): An irreducible aperiodic Markov chain $\{X_n, n = 0, 1, 2, 3, \ldots\}$ can be classified in two classes: (1) all the states are transient or null recurrent (2) all the states are non-null (also called positive) recurrent, that is all the states are ergodic. In the case of (1), there does not exist a stationary distribution. For all states i, j we get $p_{ij}^{(n)} \to 0$ as $n \to \infty$. In the case of (2), $p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$ for all the states i, j . Here, π_j will be the reciprocal of the Mean recurrence Time of j i.e. $\pi_j = \dfrac{1}{\mu_{jj}} > 0$ for all states j (22)

and $\{\pi_j\}$ is the unique stationary distribution of the Markov chain. In this case, as $n \to \infty$ the distribution of the Markov chain at time n tends to the stationary distribution, not depending on the initial distribution of the chain. In other words, if the Markov chain $\{X_n, n = 0, 1, 2, 3, \ldots\}$ an irreducible, aperiodic, and non-null Markov chain, and X_n have the distribution $\mathbf{u}^{(0)}$, an arbitrary initial distribution and $\mathbf{u}^{(n)}$, be its distribution, at time n $(n = 0, 1, 2, 3, \ldots)$, then $\lim_{n\to\infty} \mathbf{u}^{(n)} = \boldsymbol{\pi}$ exists for all states i.

Theorem 12: An irreducible aperiodic Markov chain $\{X_n, n = 0, 1, 2, 3, \ldots\}$ will be ergodic if the **balancing equation**

$$x_j = \sum_{i \in S} x_i p_{ij}, \; j \in S \tag{23}$$

has a solution $\{X_k\}$ (X_k not all zero) satisfying $\sum_{j \in S} |x_j| < \infty$.

Conversely, if the chain be ergodic then every non-negative solution $\{x_j\}$ of the balancing Eqn. (23) satisfies $\sum_{j \in S} |x_j| < \infty$.

Remark

(i) The limiting probability distribution given by $\lim_{n \to \infty} \mathbf{u}^{(n)} = \pi$ is called a steady state distribution of the Markov chain.

(ii) If the probability transition matrix $\mathbf{P}$ is symmetric for a Markov chain having finite state space $S = \{1, 2, 3, \ldots, s\}$, then the uniform distribution [$\pi_j = 1/s$ for all $j = 1, 2, 3, \ldots, s$] is stationary. More generally, the uniform distribution is stationary if the matrix $\mathbf{P}$ is **doubly stochastic**, that is, the column-sums of $\mathbf{P}$ are also 1 (we already know the row-sums of any transition matrix $\mathbf{P}$ are all 1).

(iii) A finite aperiodic irreducible chain is necessarily ergodic, thus, any finite aperiodic irreducible chain has a stationary distribution.

(iv) For an irreducible Markov chain, the existence of a stationary distribution implies that the chain is recurrent in fact positive recurrent, but the converse is not true. That is, there are irreducible, recurrent Markov chains that do not have stationary distributions. Such chains will be null recurrent.

Example 9: Find all stationary distributions for the transition matrix given below.

$$\mathbf{P} = \begin{pmatrix} 0.3 & 0.7 \\ 0.2 & 0.8 \end{pmatrix}$$

The given chain is finite, irreducible, aperiodic since all the transition probabilities are positive and hence, non-null. It must have a unique stationary distribution.

Let $\pi = (\pi_1, \pi_2)$ be the stationary distribution. From Eqn.(20), we have the balancing equations

$$\begin{aligned} \pi_1 &= 0.3\pi_1 + 0.2\pi_2 \\ \pi_2 &= 0.7\pi_1 + 0.8\pi_2 \end{aligned} \tag{24}$$

one equation is redundant; they both lead to the equation $0.7\pi_1 = 0.2\pi_2$. From above, we have an infinite number of solutions. Using the second condition from Eqn.(21),

$$\pi_1 + \pi_2 = 1. \tag{25}$$

We get unique solution $\pi_1 = \frac{2}{9}, \; \pi_2 = \frac{7}{9}$

if we take $\mathbf{u}^{(0)} = \pi = \left(\frac{2}{9} \; \frac{7}{9}\right)$

$$\mathbf{u}^{(1)} = \pi \mathbf{P} = \left(\frac{2}{9} \; \frac{7}{9}\right)\begin{pmatrix} 0.3 & 0.7 \\ 0.2 & 0.8 \end{pmatrix} = \left(\frac{2}{9} \; \frac{7}{9}\right) = \pi$$

and, proceeding recursively, we get

$$\mathbf{u}^{(n)} = \left(\frac{2}{9} \; \frac{7}{9}\right) = \pi$$

since the given Markov chain is ergodic. We, may also verify that

$$\pi_1 = \frac{1}{\mu_{11}} = \frac{2}{9},$$

where μ_{11} is the mean recurrence time of state 1 that may be obtained, as follows

$$\mu_{11} = 0 \cdot f_{11}^{(0)} + 1 \cdot f_{11}^{(1)} + 2 \cdot f_{11}^{(2)} + 3 \cdot f_{11}^{(3)} + 4 \cdot f_{11}^{(4)} + \cdots$$
$$= 0 + 1 \cdot (0.3) + 2 \cdot (0.7)(0.2) + 3 \cdot (0.7)(0.2)(0.8) + 4 \cdot (0.7)(0.2)(0.8)(0.8) + \cdots$$
$$= 0.3 + 4.2 = 4.5 = \frac{9}{2}$$

Similarly, we may verify that

$$\pi_2 = \frac{1}{\mu_{22}} = \frac{7}{9}$$

***

Example 10: Consider the transition matrix given below, which is doubly stochastic.

$$\mathbf{P} = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.3 & 0.2 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$

We use Remark, about doubly stochastic transition matrix.
Since $s = 3$, we will get $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$.
Let us check that it satisfies the balancing equations.

$$\boldsymbol{\pi}\mathbf{P} = (1/3, 1/3, 1/3)\begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.3 & 0.2 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix} = (1/3, 1/3, 1/3) = \boldsymbol{\pi},$$

verifying, $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$ is stationary distribution for $\mathbf{P}$.

***

Let us now discuss the criterion for transience.

Criterion for Transience

Here, we will state a condition for a countable state space Markov chain to be transient. It may be mentioned here, again, that any finite Markov chain cannot have all the states as transient. If a finite state space Markov chain is irreducible then it will necessarily be recurrent. We will also present an example to find the stationary distribution for an irreducible chain having a countable state space.

Theorem 13: An irreducible aperiodic Markov chain with a countable state space $S = \{0, 1, 2, \ldots\}$ and a transition matrix $\mathbf{P} = (p_{ij})$ will be **transient** (all the states will be transient) if, and only if

$$x_i = \sum_{j=1}^{\infty} p_{ij} x_j \quad \text{for all states } i = 1, 2, 3, \ldots \tag{26}$$

has a non-zero bounded solution.

***

Example 11: Consider a Markov chain having a random walk with the countable state space $\{0, 1, 2, \ldots\}$, and having an elastic barrier at state 0. The transition probabilities may be defined as follows

$$p_{ii+1} = p, \quad p_{ii-1} = q, \quad p + q = 1, \quad 0 < p < 1, \quad (i \geq 1)$$
$$p_{00} = q, \quad p_{01} = p$$
$$p_{ij} = 0 \quad \text{elsewhere}$$

As the chain is irreducible, we will study the nature of the solution of the following equations as given by the Eqn. (26), to determine the nature of states of the Markov chain

$$x_i = \sum_{j=1}^{\infty} p_{ij} x_j \quad \text{for all states } i = 1, 2, 3, \ldots$$

Therefore, we get the system of equations as

$$x_i = px_{i+1} + qx_{i-1}, \text{ for } i > 1$$
$$x_1 = px_2$$

and, these may be simplified to

$$p(x_{i+1} - x_i) = q(x_i - x_{i-1})$$
$$p(x_2 - x_1) = qx_1$$

These equations reduce recursively to

$$x_{i+1} - x_i = \left(\frac{q}{p}\right)^i x_1 \ (i \geq 1), \text{ thus we have}$$

$$x_i - x_1 = \sum_{r=1}^{i-1} (x_{r+1} - x_r) = x_1 \left\{ \left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1} \right\}$$

and, we get the solution as

$$x_i = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)} x_1 \quad \text{for } i \geq 1.$$

From the above solution, we see that x_i will be bounded if $p > q$. Therefore, according to Theorem 13, the Markov chain will be transient when $p > q$, and recurrent when $p \leq q$.

Let us find the stationary distribution of the chain when p£ q. The balancing equation to solve will be

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij} = p\pi_{j-1} + q\pi_{j+1} \quad j \geq 1$$
$$\pi_0 = q\pi_0 + q\pi_1$$

which may be written as

$$\pi_1 - \pi_0 = \frac{p}{q}\pi_0 - \pi_0$$

$$\pi_j - \pi_{j-1} = \frac{p}{q}(\pi_{j-1} - \pi_{j-2}) = \left(\frac{p}{q}\right)^{j-1} (\pi_1 - \pi_0)$$

Therefore,

$$\pi_j - \pi_{j-1} = \left(\frac{p}{q}\right)^j \pi_0 - \left(\frac{p}{q}\right)^{j-1} \pi_0 \text{ and, thus}$$

$$\pi_j - \pi_0 = \sum_{r=0}^{j-1} (\pi_{r+1} - \pi_r) = \left(\frac{p}{q}\right)^j \pi_0 - \pi_0 \text{ which gives}$$

$$\pi_j = \left(\frac{p}{q}\right)^j \pi_0 \quad \text{for } j \geq 0$$

Using the condition in Eqn.(21)

$$\sum_{j=0}^{\infty} \pi_j = 1$$

we get $\pi_0(1+\left(\frac{p}{q}\right)^1+\left(\frac{p}{q}\right)^2+\left(\frac{p}{q}\right)^3+\left(\frac{p}{q}\right)^4+\cdots)=1$ (27)

When $p=q$, then the infinite series in the Eqn. (27) will be divergent stationary distribution will not exist in this case, and the chain will be null recurrent. When $p<q$, then the Eqn. (27) gives $\pi_0=\left(1-\left(\frac{p}{q}\right)\right)$ and we have a stationary distribution as

$\pi_j=\left(1-\left(\frac{p}{q}\right)\right)\left(\frac{p}{q}\right)^j$ for $j\geq 0$, which is a geometric distribution with parameter $\frac{p}{q}$.

The chain is positive recurrent for $p<q$.

***

Existence of Stationary Distribution

Till now, we have considered only irreducible aperiodic chains and discussed the problem of the existence of stationary distributions. In general, a Markov chain may have no stationary distribution, one stationary distribution, or infinitely many stationary distributions. We have given the conditions for the existence of unique stationary distribution, along with examples. The chains presented were ergodic-finite or countable. We have also presented a Markov chain which does not possess any stationary distribution. The chains of this type were transient or null recurrent, however, they must be countable (since we cannot have a finite chain as transient, or null recurrent). As an example of the chain having infinitely many stationary distributions, we may take a transition matrix **P** to be the identity matrix, in which case all distributions on the state space will be stationary. Such chains may be finite, or countable. Example 12 illustrates the case. When the Markov chain has finite state space then it will have at least one stationary distribution whether it is reducible or irreducible, periodic or aperiodic.

Example 12: Consider a Markov chain having the following identity transition matrix.

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

Let the stationary distribution be $\boldsymbol{\pi}=(\pi_1,\pi_2,\pi_3)$. Then, the balancing equation of the chain will be

$$(\pi_1,\pi_2,\pi_3)=(\pi_1,\pi_2,\pi_3)\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}=(\pi_1,\pi_2,\pi_3)$$

Clearly, all arbitrary vector with non-negative components, $\boldsymbol{\pi}=(\pi_1,\pi_2,\pi_3)$ satisfying $\pi_1+\pi_2+\pi_3=1$, will be stationary distributions. For example, $\boldsymbol{\pi}=(0.1\ 0.3\ 0.6)$. Thus, for this chain there will exist infinite number of stationary distributions. Here, we may easily observe that a countable identity transition matrix also possesses an infinite number of stationary distributions.

***

Example 13: Consider the Markov chain having the following transition matrix

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

Solving the balancing equation

$$(\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3)\begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$$

subject to condition $\pi_1 + \pi_2 + \pi_3 = 1$, we may get a unique stationary distribution $(\pi_1, \pi_2, \pi_3) = (1/4, 1/2, 1/4)$ but $\lim\limits_{n\to\infty} p_{ij}^{(n)} \neq \pi_j$ since $p_{ii}^{(n)} = 0$ for all odd n.

The chain is irreducible, persistent, but periodic.

For this chain, the mean recurrence times may be obtained as

$$\begin{aligned}\mu_{00} &= 0.f_{11}^{(0)} + 1.f_{11}^{(1)} + 2.f_{11}^{(2)} + 3.f_{11}^{(3)} + 4.f_{11}^{(4)} + \cdots \\ &= 0 + 1.0 + 2.(1)\left(\frac{1}{2}\right) + 3.0 + 4.(1)\left(\frac{1}{2}\right)(1)\left(\frac{1}{2}\right) + 5.0 + 6.(1)\left(\frac{1}{2}\right)(1)\left(\frac{1}{2}\right)(1)\left(\frac{1}{2}\right) + \cdots \\ &= 4\end{aligned}$$

Similarly, we may get $\mu_{11} = 2, \mu_{22} = 4$.

Here, we also observe that $(\pi_1, \pi_2, \pi_3) = (1/\mu_{00}, 1/\mu_{11}, 1/\mu_{22})$.

However, the long run equilibrium probabilities, Theorem 6, is applicable.

***

Remark: In the example above we encountered a Markov chain that is irreducible, persistent, but periodic, has a unique stationary distribution having probabilities reciprocal to the mean recurrence time. We have a theorem which explains such behaviour. It says that if a Markov chain is irreducible and non-null (positive), then there will exist a stationary distribution. The result is based on the Cesaro limit. This tells us that if $\{a_n\}$ is a sequence, such that $\lim\limits_{n\to\infty} a_n = 1$, then the partial sum $\frac{1}{n+1}\sum\limits_{i=0}^{n} a_i$ also converges to the same limit, i.e., $\lim\limits_{n\to\infty} \frac{1}{n+1}\sum\limits_{i=0}^{n} a_i = 1$. This limit also exists, even when $\lim\limits_{n\to\infty} a_n$ does not exist.

Theorem 14: An irreducible, positive recurrent Markov chain has a unique stationary distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \cdots)$, given by

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{j=0}^{n} p_{ij}^{(n)} = \pi_j = \frac{1}{\mu_{jj}} \text{ for all } j, \text{ whatever state } i \text{ may be.}$$

***

You may now try the following exercises.

E7) A Markov chain has an initial distribution $\mathbf{u}^{(0)} = \{1/6 \ 1/2 \ 1/3\}$, and the following transition matrix.

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

Find its stationary distribution. Is it unique? Verify that the limiting distribution of the chain is stationary.

E8) A Markov chain has the following transition matrix $\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Find its stationary distribution. How many distributions are possible?

E9) Consider the **Ehrenfest** chain, presented in Example 12 in Unit 2, with only 3 balls. Then, the transition matrix will be

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

(i) Test the irreducibility of the chain. (ii) Find its stationary distribution.

E10) Consider a Markov chain $\{X_n\}$ with a countable state space having the following transition probabilities

$$p_{ii+1} = p_i, \quad p_{ii-1} = q_i, \quad p_i + q_i = 1, \; p_i, q_i > 0 \qquad (i > 1)$$

$$p_{00} = q_0, \quad p_{01} = p_0, \quad p_0, q_0 > 0.$$

$$p_{ij} = 0 \qquad \text{elsewhere}$$

When will the chain be transient, and when will it be recurrent?

Now let us summarise what we have done in this unit.

3.5 SUMMARY

We are furnishing, in the following, a summary of the discussions in this unit:

1. We introduced the concept of communicating classes, and, closed sets.
2. We defined the irreducible Markov chain.
3. We obtained the distribution of the first passage time to the states, and first recurrence time of states. We also defined mean time of first passage and mean recurrence time.
4. We acquainted you with the concept of recurrence and transience.
5. We investigated the limiting behaviour of the Markov chain.
6. We defined stationary distribution, and illustrated the procedures to find stationary distributions.
7. We investigated some situations when the stationary distributions of the chains exist, and is also the equilibrium distribution.

3.6 SOLUTIONS/ANSWERS

E1) The states {0, 1, 2} form a communicating class. State 3 does not communicate with any state. The chain is reducible.

The probability of ultimate return of the states is

$$f_{00} = \sum_{n=0}^{\infty} f_{00}^{(n)} = 0 + 0 + 1 \cdot \frac{3}{4} + 1 \cdot \frac{1}{4} \cdot 1 + 0 \cdots = 1$$

$$f_{11} = \sum_{n=0}^{\infty} f_{11}^{(n)} = 0 + 0 + 0 + 1 \cdot 1 \cdot \frac{1}{4} + 0 + 1 \cdot 1 \cdot \frac{3}{4} \cdot 1 \cdot \frac{1}{4} + 0 + \cdots$$

$$= \frac{1}{4} \cdot \frac{4}{1} = 1$$

$$f_{22} = \sum_{n=0}^{\infty} f_{22}^{(n)} = 0 + 0 + \frac{3}{4} \cdot 1 + \frac{1}{4} \cdot 1 \cdot 1 + 0 + \cdots = 1$$

$$f_{33}=\sum_{n=0}^{\infty}f_{33}^{(n)}=0$$

Therefore, the states 0, 1, 2 are recurrent, and state 3 is transient.

The mean recurrence times for the recurrent states are given below.

$$\mu_{00}=\sum_{n=0}^{\infty}nf_{00}^{(n)}=0\cdot 0+1.0+2\cdot\frac{3}{4}+3.\frac{1}{4}+0=\frac{9}{4}$$

$$\mu_{11}=\sum_{n=0}^{\infty}nf_{11}^{(n)}=0\cdot 0+1.0+2\cdot 0+3.\frac{1}{4}+4\cdot 0+5.\frac{3}{16}+6\cdot 0+7.\frac{9}{64}+\cdots=9$$

$$\mu_{22}=\sum_{n=0}^{\infty}nf_{22}^{(n)}=0\cdot 0+1.0+2\cdot\frac{3}{4}+3.\frac{1}{4}+0=\frac{9}{4}$$

E2) The state space is a closed set. The chain is irreducible.
The probability of ultimate return of the states are

$$f_{00}=\sum_{n=0}^{\infty}f_{00}^{(n)}=0+0+1\cdot\frac{3}{4}+0+1\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4}+0+1\cdot\frac{1}{4}\cdot 1\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4}+\cdots=1$$

$$f_{11}=\sum_{n=0}^{\infty}f_{11}^{(n)}=0+0+1\cdot\frac{1}{4}+1\cdot\frac{3}{4}+0+\cdots=1$$

$$f_{22}=\sum_{n=0}^{\infty}f_{22}^{(n)}=0+0+1\cdot\frac{1}{4}+0+1\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4}+0+1\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4}\cdot 1\cdot\frac{3}{4}+\cdots=1.$$

All the states are persistent.
Again, we have

$$\mathbf{P}^{2n+1}=\begin{pmatrix}0&1&0\\3/4&0&1/4\\0&1&0\end{pmatrix}=\mathbf{P}\text{ and }\mathbf{P}^2=\begin{pmatrix}.75&0&.25\\0&1&0\\.75&0&.25\end{pmatrix}=\mathbf{P}^{2n}$$

Therefore,

$p_{ii}^{(2n)}>0$ and $p_{ii}^{(2n+1)}=0$ and i.

Therefore, all the states are periodic, with period 2.

The mean recurrence periods are obtained, as follows

$$\mu_{00}=\sum_{n=0}^{\infty}nf_{00}^{(n)}=0\cdot 0+1.0+2\cdot\frac{3}{4}+3\cdot 0+4.\frac{3}{16}+5\cdot 0+6\cdot\frac{3}{64}+\cdots=\frac{8}{3}$$

$$\mu_{11}=\sum_{n=0}^{\infty}nf_{11}^{(n)}=0\cdot 0+1\cdot 0+2\left(\frac{1}{4}+\frac{3}{4}\right)+0=2$$

$$\mu_{22}=\sum_{n=0}^{\infty}nf_{22}^{(n)}=0\cdot 0+1.0+2\cdot\frac{1}{4}+3.0+4.\frac{3}{16}+5\cdot 0+6\cdot\frac{9}{64}+\cdots=8$$

E3) Since $p_{ii}^{(n)}=\frac{1}{2}$ for $n>N_0$. Therefore, $\sum_{n=0}^{\infty}p_{ii}^{(n)}=\infty$ and hence, the state i will be recurrent. Again, it will be aperiodic, since $p_{ii}^{(1)}=\frac{1}{2}>0$. Further, the state i is non-null since $p_{ii}^{(n)}\to\frac{1}{2}\neq 0$.

Using Theorem 6, $p_{ii}^{(n)}\to\frac{1}{\mu_{ii}}=\frac{1}{2}$, we get mean recurrence time of state i, as $\mu_{ii}=2$.

E4) The given Markov chain is irreducible since all the states can be reached from every other state of the chain.

For state 0, the probabilities of first return will be

$$f_{00}^{(1)} = p,\ f_{00}^{(2)} = q \cdot p,\ f_{00}^{(3)} = q \cdot q \cdot p,\ f_{00}^{(4)} = q \cdot q \cdot q \cdot p, \cdots$$

Clearly, the state 0 is aperiodic, since the period of the state is one. The probability of ultimate return to state 0 will be

$$f_{00} = \sum_{n=0}^{\infty} f_{00}^{(n)} = 0 + p + qp + q^2p + q^3p + q^4p + \cdots$$

$$= p(1-q)^{-1} = 1$$

and, thus, the state 0 of the Markov chain is recurrent. From the class property of recurrence it follows that the Markov chain will be recurrent, and aperiodic.

E5) See the solution of **E1)**. In the given problem, we have found that states 0, 1, 2 are non-null, aperiodic, and recurrent, and state 3 is transient. The mean recurrence Times for the recurrent states are found as follows.

$$\mu_{00} = \frac{9}{4},\ \mu_{11} = 9 \text{ and } \mu_{22} = \frac{9}{4}.$$

Using Theorem 6 and Remark 3 we have,

$$p_{00}^{(n)} \to \frac{4}{9},\ p_{11}^{(n)} \to \frac{1}{9},\ p_{22}^{(n)} \to \frac{4}{9},\ p_{33}^{(n)} \to 0 \text{ , as } n \to \infty.$$

E6) See the solution of Example 3. All the states were aperiodic, non-null persistent. The mean recurrence time for the state $0, 1, 2$ were obtained as

$$\mu_{00} = \frac{11}{6},\ \mu_{11} = \frac{11}{4} \text{ and } \mu_{22} = \frac{11}{1}.$$

Using Theorem 6, we have

$$p_{00}^{(n)} \to \frac{1}{\mu_{00}} = \frac{6}{11},\ p_{11}^{(n)} \to \frac{4}{11},\ p_{22}^{(n)} \to \frac{1}{11}, \text{ as } n \to \infty.$$

The limits of $p_{ij}^{(n)}$ for other i, j may be obtained using Theorem 8. According to this theorem, when state j is non-null aperiodic persistent $\lim_{n\to\infty} p_{ij}^{(n)} \to \frac{f_{ij}}{\mu_{jj}}$

We may find ultimate probability first passages f_{ij} from the transition matrix given in the example, as follows

$$f_{01} = \sum_{n=0}^{\infty} f_{12}^{(n)} = 0 + \frac{1}{2} + \frac{1}{2}\cdot\frac{3}{4} + 0 + \frac{1}{2}\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4} + 0 + \frac{1}{2}\cdot\frac{1}{4}\cdot 1\cdot\frac{1}{4}\cdot 1\cdot\frac{3}{4} + \cdots$$

$$= \frac{1}{2} + \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{4}{3} = 1.$$

Similarly, we may obtain $f_{02} = f_{10} = f_{12} = f_{20} = f_{21} = 1$.

$$\lim_{n\to\infty} p_{10}^{(n)} \to \frac{1}{11/6} = \frac{6}{11},\quad \lim_{n\to\infty} p_{20}^{(n)} \to \frac{1}{11/6} = \frac{6}{11}$$

$$\lim_{n\to\infty} p_{01}^{(n)} \to \frac{1}{11/4} = \frac{4}{11},\quad \lim_{n\to\infty} p_{21}^{(n)} \to \frac{1}{11/4} = \frac{4}{11}$$

$$\lim_{n\to\infty} p_{02}^{(n)} \to \frac{1}{11/1} = \frac{1}{11},\quad \lim_{n\to\infty} p_{12}^{(n)} \to \frac{1}{11/1} = \frac{1}{11}$$

Therefore,

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} \frac{6}{11} & \frac{4}{11} & \frac{1}{11} \\ \frac{6}{11} & \frac{4}{11} & \frac{1}{11} \\ \frac{6}{11} & \frac{4}{11} & \frac{1}{11} \end{pmatrix}.$$

E7) The chain is irreducible since all the states are communicable. We may also verify that the chain is aperiodic recurrent. Solving the balancing equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, i.e.

$$(\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3)\begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$

we get

$$\pi_1 = 0.5\pi_2 + 0.5\pi_3$$
$$\pi_2 = 0.5\pi_1 + 0.5\pi_3$$
$$\pi_3 = 0.5\pi_2 + 0.5\pi_1$$

Solving these equations along with the condition $\pi_1 + \pi_2 + \pi_3 = 1$, we get, unique solution $(\pi_1, \pi_2, \pi_3) = (1/3, 1/3, 1/3)$. This is obvious since $\mathbf{P}$ is doubly stochastic.

From **Theorem 7** of Unit 2, we have $\mathbf{u}^{(n)} = \mathbf{u}^{(0)}\mathbf{P}^{(n)}$. From **Theorem 11**, we have $p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$. In the matrix form, we have as $n \to \infty$.

$$\mathbf{P}^{(n)} \to \begin{pmatrix} \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{pmatrix}$$

Therefore, as

$$\mathbf{u}^{(n)} = \mathbf{u}^{(0)}\mathbf{P}^{(n)} \to (1/6\ 1/2\ 1/3)\begin{pmatrix} \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{pmatrix} = \boldsymbol{\pi}$$

The limiting distribution is stationary and independent of the initial distribution.

E8) The chain

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

is reducible. It has two closed sets $\{1, 2\}$ and $\{3\}$.
Here, we may get infinite solutions to the balancing equation, given below

$$\pi_1 = .5\pi_1 + \pi_2$$
$$\pi_2 = .5\pi_1$$
$$\pi_3 = \pi_3$$

Removing redundancy, we get $\pi_2 = .5\pi_1$ and the condition $\pi_1 + \pi_2 + \pi_3 = 1$. Then, two equations provide an infinite number of solutions $(\pi_1, \pi_2, \pi_3) = (2(1-x)/3,\ (1-x)/3,\ x)$ where $0 \le x \le 1$

E9) The Markov chain is irreducible, but periodic, with period 2. However, we may solve the general balance equations to get its stationary distribution which is $\boldsymbol{\pi} = (1/8, 3/8, 3/8, 1/8)$, and is unique. Note, however, that $\lim_{n\to\infty} p_{ij}^{(n)} \ne \pi_j$ since $p_{ii}^{(n)} = 0$ for all odd n.

E10) The chain is irreducible as all states communicate. To determine the nature of states of the Markov chain, we will study the nature of the solution of the following equations, as given by Eqn. (26)

$$x_i = \sum_{j=1}^{\infty} p_{ij} x_j \quad \text{for all states } i = 1, 2, 3, \ldots\ T$$

Therefore, we get

$$x_i = p_i x_{i+1} + q_i x_{i-1} \text{ for } i > 1$$
$$x_1 = p_1 x_2$$

and these may be simplified as

$$p_i(x_{i+1} - x_i) = q_i(x_i - x_{i-1})$$
$$p_1(x_2 - x_1) = q_1 x_1$$

and, hence

$$\frac{x_{i+1} - x_i}{x_i - x_{i-1}} = \frac{q_i}{p_i}$$

We get, recursively

$$x_{i+1} - x_i = \frac{q_i}{p_i} \cdots \frac{q_2}{p_2} \frac{q_1}{p_1} x_1$$

Adding the above equations, we get

$$x_i = x_1 \sum_{k=1}^{i-1} L_k \text{ where, } L_k = \frac{q_k}{p_k} \cdots \frac{q_2}{p_2} \frac{q_1}{p_1}$$

Therefore, the above equations will have a bounded solution if, and only if, $\sum L_i < \infty$. Therefore, the states will be transient if $\sum L_i < \infty$ and recurrent, if the infinite series is divergent.

—x—

NOTES