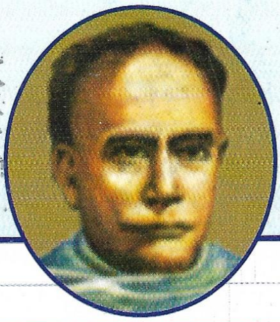
DIRECTORATE OF DISTANCE EDUCATION



VIDYASAGAR UNIVERSITY
MIDNAPORE-721 102

M. Sc. in Mathematics

Part - II

Paper - X

Module No. 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120

CONTENTS

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X **Special Paper – OR**

Module No. - 109

Advanced Optimization and Operations Research-II

(Dynamic Programming)

STRUCTURE :

109.1. Introduction

In the dynamic programming technique, we usually start with the smallest and hence the simplest subinstances. By combining their solutions, we obtain the answers to subinstances of increasing size, until finally we arrive at the solution of the original problem.

109.2. Objectives

After completion of this module you will be able to :

- Explain the concept of dynamic programming.
- Explain the method to solve optimization problem with single and multiple constraints.
- Solve inventory problems using dynamic programming method.
- Solve linear programming problem using dynamic programming method.

109.3. Key words :

Dynamic programming method, optimization problems with single or multiple constraints, LPP, inventory control.

109.4. Elements of Dynamic Programming Problem :

In the following we discuss the three basic characteristic of dynamic programming technique.

1. **Stage.** The dynamic programming problem can be decomposed or divided into a sequence of smaller sub-problems called stages of the original problem. At each stage there are a number of decision alternatives and a decision is made by selecting the most suitable out of the given list. Stages generally represent different time periods associated with the planning period of the problem, places, people or other objects.

2. **State.** Each stage in a dynamic programming problem has a certain number of states associated with it. These states represent various conditions of the decision process at a stage. The variables which specify the condition of the decision process or describe the status of the system at a particular stage are called state variables. These variables provide information for analyzing the possible effects that the current decision could have upon future courses of action. At any stage of the decision-making process there could be a finite finite or infinite number of states.

3. **Return function.** At each stage, a decision is made which can affect the state of the system at the next stage

and help in arriving at the optimal solution at the current stage. Every decision that is made has its own merit in terms of benefit associated with it and is described in an algebraic equation form.

This equation is generally called a return function, since for every set of decisions, a return on each decision is obtained. This return function in general depends on the state variable as well as the decision made at a particular stage. An optimal policy or decision at a stage yields optimal i.e., maximum or minimum return for a given value of the state variable.

Bellman's Principle of Optimality :

An optimal policy has the property that whatever the initial state and decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

Mathematically,

$$f_n(x) = \underset{d_n \in X}{\text{optimize}} \left[r(d_n) * f_{n-1}\{T(x, d_n)\} \right]$$

where

$f_n(x)$ = the optimal return from an n-stage process when initial state is x;

$r(d_n)$ = immediate return due to decision d_n;

$T(x, d_n)$ = the transfer function which gives the resulting state;

X = set of admissible decisions.

It may be remembered that if a problem does not satisfy the principle of optimality cannot be solved by the dynamic programming technique.

The solution of problems by dynamic programming is usually done in two stages:

- The development of functional equations for the problem,
- To solve functional equations for determining the optimal policy.

109.5. Mode I : Shortest Path Problem

Example 1 : (Multi-stages graph) Find the shortest path from the vertex A to the vertex B along the edges joining various vertices lying between A and B shown in Figure 1. The length of each edge is given, with each edge.

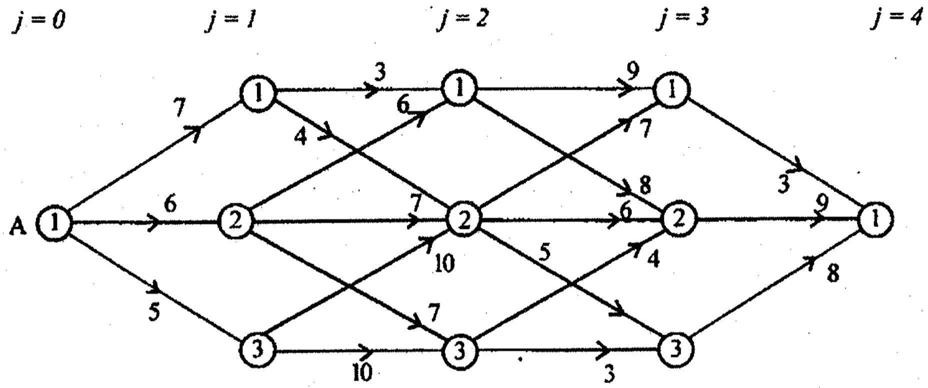


Figure 1

Solution :

Step 1. Formulation of the problem :

We divide the vertices into five stages 0, 1, 2, 3 and 4 denoted by j shown in Figure 1. The vertices of each stage are numbered as 1, 2, 3, ... It may be noted the vertices of stage j are adjacent to the vertices of stage $(j+1)$ only.

The state of the system is denoted by s_j. Thus, s_0 is the state in which the node A lies. The state s_2 has three vertices possible values, say, 1, 2 and 3 corresponding to three vertices in stage 2, and so on. The possible alternative paths from one stage to the other will be called decision variables denoted by d_j, the decision which takes from state s_{j-1} to state s_j. The return is the function of decision will be denoted by $f_i(d_j)$. Here d_j can be identified with the length of the corresponding edge. Here we consider $f_j(d_j) = d_j$.

The minimum path from state s_0 to any vertex in state s_j will be denoted by $F_j(s_j)$. For this problem we have to calculate the minimum path $F_4(s_4)$ and the values of the decisions variables d_1, d_2, d_3 and d_4.

Setp 2. Obtain the functional Equations

We start from the vertex B backward. Obviously, d_4 can either be 3 or 9 or 8, as the possible paths are

$$1 \xrightarrow{3} 1, 2 \xrightarrow{9} 1, 3 \xrightarrow{8} 1.$$

If $d_4 = 3$, then $s_3 = 1$. Similarly, if $d_4 = 9$ then $s_3 = 2$;

$d_4 = 8$, then $s_3 = 3$. Hence the minimum path from A to B is either through $s_3 = 1$ or 2 or 3 according as d_4

is 3, 9 or 8.

Thus,

$$F_4(s_4) = \min[3 + F_3(1), 9 + F_3(2), 8 + F_3(3)] = \min_{d_4}[d_4 + F_3(s_3)].$$

In similar way,

$$F_3(1) = \min[9 + F_2(1), 7 + F_2(2)]$$

$$F_3(2) = \min[8 + F_2(1), 6 + F_2(2), 4 + F_2(3)]$$

$$F_3(3) = \min[5 + F_2(2), 3 + F_2(3)].$$

In general,

$$F_3(s_3) = \min_{d_3}[d_3 + F_2(s_2)], s_3 = 1, 2, 3.$$

Similarly,

$$F_2(s_2) = \min_{d_2}[d_2 + F_1(s_1)].$$

Lastly,

$$F_1(s_1) = d_1.$$

Step 3. Determination of the minimum path.

Now, we can determine the value of $F_4(s_4)$ recursively. The states $s_j, j = 0, 1, 2, 3$ are tabulated in the following:

State s_0

d_1	7	6	5
s_1	1	2	3

State s_1

s_2 \ d_2	3	4	6	7	10
1	1	–	2	–	–
2	–	1	–	2	3
3	–	–	–	2	3

State s_2

s_3 \ d_3	3	4	5	6	7	8	9
1	–	–	–	–	2	–	1
2	–	3	–	2	–	1	–
3	3	–	2	–	–	–	–

State s_3

s_4 \ d_4	3	9	8
1	1	2	3

It is observed that s_{j-1} may not be defined for all combinations of s_j and d_j. The dash indicates there is no transformation required.

The recursive operations are shown in the following tables.

j	s_1	s_1	s_1
1	1	7	7
	2	6	6
	3	5	5

		$F_1(s_1)$					$d_2+F_1(s_1)$					$F_2(s_2)$
j	d_2 \ s_2	3	4	6	7	10	3	4	7	7	10	Min.
2	1	7	–	6	–	–	10	–	12	–	–	10
	2	–	7	–	6	5	–	11	–	13	15	11
	3	–	–	–	6	5	–	–	–	13	15	13

		$F_2(s_2)$							$d_3+F_2(s_2)$							$F_3(s_3)$
j	d_3 \ s_3	3	4	5	6	7	8	9	3	3	5	6	7	8	9	
3	1	–	–	–	–	11	–	10	–	–	–	–	18	–	19	18
	2	–	13	–	11	–	10	–	–	17	–	17	–	18	–	17
	3	13	–	11	–	–	–	–	16	–	16	–	–	–	–	16

		$F_3(s_3)$			$d_4+F_3(s_3)$			$F_4(s_4)$
j	d_4 \ s_4	3	9	8	3	9	8	Min.
4	1	18	17	16	21	26	24	21

Thus, $F_4(s_4) = 21$, i.e., the length of the minimum path is 21 unit. By tracing the minimum path and decision backwards, successive distances are 7, 4, 7, 3 through the states $s_0 = 1, s_1 = 1, s_2 = 2, s_3 = 1$ and $s_4 = 1$.

109.6. Model II. Single Additive Constraint, Multiplicative Separable Return

The problem of this type is stated below:

Maximize $z = f_1(y_1) f_2(y_2) .... f_n(y_n)$

subject to $a_1 y_1 + a_2 y_2 + ... + a_n y_n = b,$

$$y_j \geq 0, a_j \geq 0,$$

where y_j is the decision variable at *j*th stage and a_j is the constant, for all $j = 1, 2, ..., n$.

We introduce the state variables, i.e.,

$$s_n = a_1 y_1 + a_2 y_2 + ... + a_n y_n = b,$$

$$s_{j-1} = s_j - a_j y_j, j = 2, 3, ..., n.$$

Let

$$F_j(s_j) = \max_{y_1, y_2, ..., y_n} \{f_1(y_1) f_2(y_2) .... f_n(y_n)\}$$

the general recursion formula is

$$F_j(s_j) = \max_{y_j} [f_j(y_j) F_{j-1}(s_{j-1})], j = n, n-1, ..., 2$$

and $F_1(s_1) = f_1(y_1)$.

Example 1. Find the value of z = maximize $(y_1 y_2 y_3)$ subject to $y_1 + y_2 + y_3 = 5;\ y_1, y_2, y_3 \geq 0$.

Solution : Let s_1, s_2, s_3 be the state variables, where

$$s_3 = y_1 + y_2 + y_3 = 5;\ s_2 = s_3 - y_3 = y_1 + y_2, s_1 = s_2 - y_2 = y_1.$$

Also, $F_3(s_3) = \max_{y_3} [y_3 F_2(s_2)]$

$$F_2(s_2) = \max_{y_2} [y_2 F_1(s_1)]$$

$$F_1(s_1) = y_1 = s_2 - y_2.$$

Hence, $F_2(s_2) = \max_{y_2} [y_2 (s_2 - y_2)]$.

Let $A = y_2 s_2 - y_2^2$. $\frac{dA}{dy_2} = s_2 - 2y_2$.

At extrema, $\frac{dA}{dy_2} = 0$, gives $y_2 = s_2/2$.

Also, $\frac{d^2A}{dy_2^2} = -2 < 0$.

Therefore, A is maximum when $y_2 = s_2/2$.

Hence, using the Bellman's principle of optimality

$$F_3(s_3) = \max_{y_3}\left[y_3 s_2^2/4\right] = \max_{y_3}\left[y_3(s_3 - y_3)^2/4\right].$$

Again, using calculus, we get

$$y_3 = s_3/3 = 5/3.$$

Also, $y_2 = 5/3, y_1 = 5/3$.

Hence, $\max\ y_1 y_2 y_3 = \frac{125}{27}$.

109.7. Model III: Single Additive Constraint, Additively Separable Return

Let us consider the problem

$$\text{minimize } z = \sum_{j=1}^{n} f_j(y_j)$$

subject to the constraints

$$\sum_{j=1}^{n} a_j y_j \geq b, \quad a_j \text{ and } b \text{ are real numbers,}$$

$$a_j \geq 0, y_j \geq 0, b > 0, j = 1, 2, \ldots, n.$$

This is an n-stage problem where the suffix j indicates the stage. Since values of y_j are to be decided, y_j is called decision variable. The return at the j th stage is the function $f_j(y_j)$. Thus, each decision y_j is associated with a return function $f_j(y_j)$.

Let us introduce the state variables $s_0, s_1, \ldots, s_n$, as

$$s_n = a_1 y_1 + a_2 y_2 + \ldots + a_n y_n \geq b,$$

$$s_{n-1} = a_1y_1 + a_2y_2 + ... + a_{n-1}y_{n-1} = s_n - a_ny_n,$$

$$s_{n-2} = a_1y_1 + a_2y_2 + ... + a_{n-2}y_{n-2} = s_{n-1} - a_{n-1}y_{n-1},$$

....

....

$$s_1 = a_1y_1 = s_2 - a_2y_2.$$

Also, $s_{j-1} = T_j\left(s_j, y_j\right), j = 1, 2, ..., n,$ is the stage transformation function and indicates that each state variable is a function of next state and decision variables.

$F_n(s_n)$ denotes the minimum value of z for any feasible value of s_n, where s_n being the function of all decision variables. Thus.

$$F_n(S_n) = \min_{y_1, y_2, ..., y_n}\left[f_1(y_1) + f_2(y_2) + ... + f_n(y_n)\right], \quad s_n \geq b.$$

Now, we choose a particular value of y_n and minimize z over the remaining $(n-1)$ variables. Hence

$$F_n(s_n) = \min_{y_1, y_2, ..., y_{n-1}}\left[\sum_{j=1}^{n-1} f_j\left(y_j\right)\right] = f_n(y_n) + F_{n-1}(s_{n-1}).$$

The values $y_1, y_2, ..., y_{n-1}$ for which $\sum_{j=1}^{n-1} f_j\left(y_j\right)$ is minimum keeping y_n fixed thus depend on s_{n-1} which is a function of s_n and y_n. Therefore, the minimum over all y_n for any feasible s_n would now become

$$F_n(s_n) = \min_{y_n}\left[f_n(y_n) + F_{n-1}(s_{n-1})\right].$$

If the value of $F_{n-1}(s_{n-1})$ is known for all y_n, the function to be minimized would involve only a single variable y_n. This mimimization now becomes easy and can be done by simple methods. Similarly, the recursion formula is

$$F_j(s_j) = \min_{y_j}\left[f_j\left(y_j\right) + F_{j-1}\left(s_{j-1}\right)\right], 1 \leq j \leq n.$$

and $F_1(s_1) = f_1(y_1)$.

Starting with $F_1(s_1)$ and recursively optimizing to obtain $F_2(s_2), F_3(s_3), ...$ we obtain $F_n(s_n)$ for each variable s_n.

Example 1. Find the values of y_1, y_2, y_3 to

Minimize $z = y_1^2 + y_2^2 + y_3^2$

subject to $y_1 + y_2 + y_3 \geq 15, y_1, y_2, y_3 \geq 0.$

Solution. Here the decision variables are y_1, y_2, y_3 and let the state variables be s_1, s_2, s_3. They are defined as

$$s_3 = y_1 + y_2 + y_3 \geq 15$$

$$s_2 = y_1 + y_2 = s_3 - y_3$$

and $s_1 = y_1 = s_2 - y_2$.

$$F_3(s_3) = \min_{y_3}\left[y_3^2 + F_2(s_2)\right]$$

$$F_2(s_2) = \min_{y_2}\left[y_2^2 + F_1(s_1)\right]$$

$$F_1(s_1) = y_1^2 = (s_2 - y_2)^2.$$

Thus, $F_2(s_2) = \min_{y_2}\left[y_2^2 + (s_2 - y_2)^2\right]$.

Let $A = y_2^2 + (s_2 - y_2)^2 . \dfrac{dA}{dy_2} = 2y_2 - 2(s_2 - y_2)$.

At extrema, $\dfrac{dA}{dy_2} = 0$, i.e., $2y_2 - 2(s_2 - y_2) = 0$ or, $y_2 = s_2/2$.

Hence, $F_2(s_2) = s_2^2/2$.

Now, $F_3(s_3) = \min_{y_3}\left[y_3^2 + F_2(s_2)\right]$

$$= \min_{y_3}\left[y_3^2 + \frac{(s_3 - y_3)^2}{2}\right].$$

Again, using calculus, for minimum of the function of single variable y_3,

$$2y_3 - (s_3 - y_3) = 0 \text{ or } y_3 = \frac{s_3}{3}.$$

Hence, $F_3(s_3) = \dfrac{s_3^2}{3}, \qquad s_3 \geq 15.$

Since $F_3(s_3)$ is minimum for $s_3 = 15$, the minimum value of $y_1^2 + y_2^2 + y_3^2$ is 75, where $y_1 = y_2 = y_3 = 5$.

Example 2. Show that $\sum_{i=1}^{n} p_i \log p_i$, subject to $\sum_{i=1}^{n} p_i = 1$, is minimum when $p_1 = p_2 = ... = p_n = 1/n$.

Solution. Let $f_n(1)$ denote the minimum attainable sum regarded as a function of discrete variable n.

For $n = 1, f_1(1) = p_1 \log p_1 = 1 \log 1$. ....(i)

Now, consider the case for $n = 2$.

Let $p_1 = z$ and $p_2 = 1 - z$, then

$$f_2(1) = \min_{0 \le z \le 1} [p_1 \log p_1 + p_2 \log p_2]$$

$$= \min_{0 \le z \le 1} [z \log z + (1-z) \log(1-z)]. \quad \text{.......(ii)}$$

Since $f_1(1-z) = (1-z)\log(1-z)$

$$f_2(1) = \min_{0 \le z \le 1} [z \log z + f_1(1-z)] \quad \text{.......(iii)}$$

It can be easily verified that minimum value of

$F(z) = z \log z + (1-z)\log(1-z)$ is attained for $z = \frac{1}{2}$.

Thus for n=2, optimal policy is given by

$$\left. \begin{array}{l} p_1 = p_2 = \frac{1}{2} \\ f_1(1) = 2\left(\frac{1}{2}\log\frac{1}{2}\right) \end{array} \right\} \quad \text{.......(iv)}$$

Similarly, for n=3, take one of the three parts as z leaving an amount (1–z) for further division into two parts. Using Bellman's principle,

$$f_3(1) = \min_{0 \le z \le 1} [z \log z + f_2(1-z)]. \quad \text{.......(v)}$$

Since $f_2(1) = 2\left(\frac{1}{2}\log\frac{1}{2}\right)$,

$$f_2(1-z) = 2\frac{1-z}{2}\log\frac{1-z}{2}.$$

Hence the equation (v) becomes

$$f_3(1) = \min_{0 \le z \le 1} \left[z \log z + 2\left(\frac{1-z}{2}\right)\log\frac{1-z}{2} \right] \quad \text{.......(vi)}$$

in which $z \log z + 2\left(\frac{1-z}{2}\right)\log\frac{1-z}{2} = F(z)$, which is a known function of the variable z. Again, it is easy to observed that the minimum value of $F(z)$ is attained for $z = \frac{1}{3}$ where $0 \le z \le 1$.

Since, $f_2(1-z) = f_2\left(1-\frac{1}{3}\right) = f_2\left(\frac{2}{3}\right) = 2\left(\frac{2}{6}\log\frac{2}{6}\right)$

$$= \frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3}.$$

Thus for $n=3$, optimal policy is given by

$$\left.\begin{aligned} p_1 = p_2 = p_3 = \frac{1}{3} \\ f_3(1) = 3\left(\frac{1}{3}\log\frac{1}{3}\right). \end{aligned}\right\} \qquad \text{.....(vii)}$$

Further, suppose that the optimal policy

$$p_1 = p_2 = \ldots = p_n = \frac{1}{n} \text{ for which } f_n(1) = n\left(\frac{1}{n}\log\frac{1}{n}\right)$$

holds for $n = 2, 3, 4, \ldots, m$. Thus, we have to show that this result will also hold for $n = m+1$. Now, applying the induction on n.

By the principle of optimality,

$$f_{m+1} = \min_{0\le z\le 1}\left[z\log z + f_m(1-z)\right]$$

$$= \min_{0\le z\le 1}\left[z\log z + m\left(\frac{1-z}{m}\log\frac{1-z}{m}\right)\right]$$

in which the function

$$F(z) = z\log z + (1-z)\log\frac{1-z}{m}$$

is a function of the variable z. The minimum value of this function is attained for $z = \frac{1}{m+1}$. Since

$$f_m(1-z) = f_m\left(\frac{m}{m+1}\right) = m\left(\frac{\frac{m}{m+1}}{m}\right)\log\left(\frac{\frac{m}{m+1}}{m}\right)$$

$$= \frac{m}{m+1}\log\frac{1}{m+1}$$

$$= \frac{1}{m+1}\log\frac{1}{m+1} + \frac{1}{m+1}\log\frac{1}{m+1} + \ldots m \text{ times.}$$

So the optimality policy for $n = m+1$ will be

$$p_1 = p_2 = \ldots = p_{m+1} = \frac{1}{m+1}$$

for which $f_{m+1}(1) = (m+1)\left[\frac{1}{m+1}\log\frac{1}{m+1}\right]$.

Hence the result is true for $n = m+1$.

Thus, the optimal policy is

$$p_1 = p_2 = \ldots\ldots = p_n = \frac{1}{n}$$

will be true for general n.

109.8. Model IV : Single Multiplicative Constraint, Additively Separable Return.

Let us consider the problem

Minimize $z = f_1(y_1) + f_2(y_2) + \ldots\ldots + f_n(y_n)$

subject to $y_1 y_2 .. y_n \geq p, p > 0, y_j \geq 0$ for all j.

State variables are defined as

$$s_n = y_n y_{n-1} \ldots y_2 y_1 \geq p$$

$$s_{n-1} = s_n / y_n = y_{n-1} y_{n-2} \ldots\ldots y_2 y_1$$

....

....

$$s_2 = s_3 / y_3 = y_2 y_1$$

$$s_1 = s_2 / y_2 = y_1$$

These state variables are stage transformations of the type $s_{j-1} = T_j(s_j, y_j)$.

Let $F(s_n)$ be the minimum of value of the objective function for any feasible s_n. Thus, the recursion formula is $F_j(s_j) = \min_{y_j}\left[f_j(y_j) + F_{j-1}(s_{j-1})\right], 2 \leq j \leq n$

which will lead to the required situation.

Example 1. Use Bellman's principle of optimality to

Minimize $z = y_1 + y_2 + \ldots\ldots + y_n$

subject to $y_1 y_2 .. y_n = a, y_j \geq 0$ for all $j=1, 2, \ldots, n$.

Solution. Let $f_n(a)$ denote the minimum attainable sum $y_1 + y_2 + y_3 + \ldots + y_n$ when the quality a is factorized into n factors.

For n=1, a is factorized into one factor only, so

$$f_1(1) = \min_{y_1=a}\{y_1\} = a.$$

For n=2, a is factorized into two factors y_1, y_2.

If $y_1 = z, y_2 = a/z$, then

$$f_2(a) = \min\{y_1 + y_2\}$$
$$= \min_{0\le z\le a}\{z + a/z\}$$
$$= \min_{0\le z\le a}\{z + f_1(a/z)\} \quad [\because f_1(a/z) = a/z]$$

For $n = 3$, a is factorized into three factors y_1, y_2 and y_3.

If $y_1 = z, y_2 y_3 = a/z$, then

$$f_3(a) = m\{y_1 + y_2 + y_3\}$$
$$= \min_{0\le z\le a}\{z + f_2(a/z)\}.$$

Proceeding likewise, the recurrence relation for $n = i$ becomes

$$f_i(a) = \min_{0\le z\le a}\{z + f_{i-1}(a/z)\}, i = 2,3,....,n.$$

Now, proceed to solve this functional equation as follows:

$$f_1(a) = a,$$
$$f_2(a) = \min_{0\le z\le a}\{z + a/z\}$$
$$= \sqrt{a} + \frac{a}{\sqrt{z}} = 2\sqrt{a} \quad \text{[Using Calculus]}$$
$$f_3(a) = \min_{0\le z\le a}\{z + f_2(a/z)\}$$
$$= \min_{0\le z\le a}\{z + 2\sqrt{a/z}\}$$
$$= a^{1/3} + 2\sqrt{\left(a/a^{1/3}\right)} = 3a^{1/3}$$

and so on.

By induction hypothesis, assume for $n=m$,

$$f_m(a)=ma^{1/m}.$$

Now, the result can be proved for $n=m+1$ as follows:

$$f_{m+1}(a)=\min_{0\le z\le a}\{z+f_n(a/z)\}$$

$$=\min_{0\le z\le a}\left\{z+n\sqrt{(a/z)^{1/n}}\right\}$$

$$=(n+1)a^{1/(n+1)}. \quad \text{[Using Calculus]}$$

Hence the optimal policy is

$$\left(a^{1/n},a^{1/n},\ldots\ldots,a^{1/n}\right) \text{ with } f_n(a)=na^{1/n}.$$

Example 2. Solve the following problem using dynamic programming.

Maximize $z=y_1^2+y_2^2+y_3^2$

subject to $y_1y_2y_3\le 4$, where y_1,y_2 and y_3 are positive integers.

Solution. The state variables are $s_3=y_1y_2y_3\le 4, s_2=s_3/y_3=y_1y_2, s_1=s_2/y_2=y_1$.

Stage returns : $f_j(y_j)=y_j^2, j=1,2,3$

y_j :	1	2	3	4
$f_j(y_j)$:	1	4	9	16

Stage returns : $s_{j-1}=s_j/y_j, j=2,3$

s_j \ y_j	1	2	3	4
1	1	–	–	–
2	2	1	–	–
3	3	–	1	–
4	4	2	–	1

Recursive operations

s_1	$F_1(s_1)$
1	1
2	4
3	9
4	16

	$f_2(y_2)$				$F_1(s_1)=f_1(y_1)$				$F_2(s_2)$
s_2 \ y_2	1	2	3	4	1	2	3	4	
1	1	–	–	–	1	–	–	–	2
2	1	4	–	–	4	1	–	–	5
3	1	–	9	–	9	–	1	–	10
4	1	4	–	16	16	4	–	1	17

	$f_3(y_3)$				$F_2(s_2)$				$F_3(s_3)$
s_3 \ y_3	1	2	3	4	1	2	3	4	
1	1	–	–	–	2	–	–	–	3
2	1	4	–	–	5	2	–	–	6
3	1	–	9	–	10	–	2	–	11
4	1	4	–	16	17	5	–	–	18

Thus, the required solution is given by

$y_3 = 1, y_2 = 1, y_1 = 4$ and $max\ z = 18$.

109.9. Model V: System Involving More than One Constraint

In the previous sections we consider the problems containing only one constraint. However, the dynamic programming method can be applied to problems involving more than one constraint also. In single constraint problems, there has to be single state variable for each stage, while in multi-constraint problems there has to be one state variable per constraint per stage. The structure of problems is of such type that sometimes it is possible to reduce the number of state variables. The stage transformation becomes more and more complicated with the increase in number of constraints and consequently the state variables. So, the dynamic programming method is not suitable for problems contain large number of constraints.

Example 1. Solve the following problem using dynamic programming method :

Maximize $z = y_1^3 + y_2^3 + y_3^3$,

subject to $y_1 + y_2 + y_3 \le 6$,

$y_1 y_2 y_3 \le 6$, where y_1, y_2 and y_3 are positive integers.

Solution. Here we define two sets of state variables as follows:

$s_3 = y_1 + y_2 + y_3$ $t_3 = y_1 y_2 y_3$

$s_2 = s_3 - y_3 = y_1 + y_2$ $t_2 = t_3 / y_3 = y_1 y_2$

$s_1 = s_2 - y_2 = y_1$ $t_1 = t_2 / y_2 = y_1$.

Clearly, the solution is y_i, $i = 1, 2, 3, 4$.

For stage j=1, stage formations will give following possible values of s_1 and t_1.

y_1	s_1	t_1
1	1	1
2	2	2
3	3	3
4	4	4

For $j = 2, 3$ following table gives transformations

$s_{j-1} = T_{j-1}(s_j, y_j), t_{j-1}(t_j, y_j)$.

(s_j, t_j) \ y_j	1	2	3	4
	(s_{j-1}, t_{j-1})			
(1, 1)	(0,1)	(–,–)	(–,–)	(–,–)
(2, 2)	(1, 2)	(0, 1)	(–,–)	(–,–)
(3, 3)	(2, 3)	(1,–)	(0, 1)	(–,–)
(4, 4)	(3, 4)	(2, 2)	(1,–)	(0, 1)
(5, 5)	(4, 5)	(3,–)	(2,–)	(1,–)
(6, 6)	(5, 6)	(4, 3)	(3, 2)	(2,–)

In order to preserve the validity of constraints it is not necessary to consider $s_j, t_j > 6$. Since fractional and negative integral values are not considered, so these are denoted by dash (–) in above table.

Stage 1.

$$F(s_1, t_1) = y_1^3$$

y_1	s_1	t_1	$F_1(s_1, t_1)$
1	1	1	1
2	2	2	8
3	3	3	27
[4]	4	4	**[64]**

Stage 2.

$$F_2(s_2, t_2) = \max_{y_2}\left[y_2^3 + F_1(s_1, t_1)\right]$$

y_2	s_1	t_1	$F_1(s_1, t_1)$	$y_2^3 + F_1(s_1, t_1)$	s_2	t_2	$F_2(s_2, t_2)$
1	1	1	1	2	2	1	2
	2	2	8	9	3	2	9
	3	3	27	28	4	3	28
	4	4	64	65	5	4	65
2	1	1	1	9	3	2	–
	2	2	8	16	4	4	16
	3	3	27	35	5	6	35
3	1	1	1	28	4	3	–
	2	2	8	35	5	6	–
4	1	1	1	65	5	4	–

Stage 3.

$$F_3(s_3,t_3)=\max_{y_3}\left[y_3^3+F_2(s_2,t_2)\right]$$

y_3	s_2	t_2	$F_2(s_2,t_2)$	$y_3^3+F_2(s_2,t_2)$	s_3	t_3	$F_3(s_3,t_3)$
[1]	2	1	2	3	3	1	3
	3	2	9	10	4	2	10
	4	3	28	29	5	3	29
	4	4	16	17	5	4	17
	5	5	65	66	6	4	[66]
	5	6	35	36	6	6	36
2	2	1	2	10	4	2	–
	3	2	9	17	5	4	–
	4	3	28	36	6	6	–
3	2	1	2	29	5	3	–
	3	2	9	36	6	6	–
4*	2	1	2	66*	6	4	66*

Now, proceeding in the backward direction, optimal decisions are

$$(y_1,y_2,y_3)=(4,1,1) \text{ or } (1,1,4) \text{ or } (1,4,1).$$

Hence max $F_3(s_3,t_3)=66$ for $(s_3,t_3)=(6,4)$.

109.10. Application in Inventory Control

We already studied inventory models for constant demand of an item. If models are considered in which the demand is known exactly but different in each period, the solution of such models become somewhat more complicated. Such inventory models may be easily solved by using the dynamic programming technique.

Example 1. A man is engaged in buying and selling identical items. He operates from a warehouse that can hold 500 items. Each month he can sell any quantity that he chooses upto the stock at the beginning of the month. Each month, he can buy as much as he wishes for delivery at the end of the month so long as his stock does not exceed 500 items. For the next four months, he has the following error-free forecasts of cost sales prices:

Month :	i	1	2	3	4
Cost :	c_i	27	24	26	28
Sale Price :	p_i	28	25	25	27

If he currently has a stock of 200 units, what quantities should he sell and buy in next four months? Find the solution using dynamic programming.

Solution. We devoted first, second, third and fourth month as period 1, 2, 3 and 4 respectively.

Let

x_i = amount to be sold during the month i

y_i = amount to be ordered during the month i

b_i = stock level in the beginning of month i

p_i = sale price in the month i

c_i = purchase price in the ith month, and

H = warehouse capacity.

Let $f_n(b_n)$ be the maximum possible return when there are n months to precede and initial stock is b_n.

We consider $i = 4$ first and $i = 1$ last.

Thus,

$$f_1(b_n) = \max_{x_n, y_n}[p_1 x_n - c_n y_n]$$

where $b_n \geq x_n, b_n - x_n + y_n \leq H$.

Also, $f_n(b_n) = \max_{x_n, y_n}[p_n x_n - c_n y_n + f_{n-1}(b_n - x_n + y_n)]$.

For $n = 1$,

$$f_1(b_1) = \max_{x_1, y_1}[p_1 x_1 - c_1 y_1].$$

Obviously, $y_1 = 0, x_1 = b_1$.

Therefore, $f_1 = (b_1) = p_1 b_1 = 27 b_1$ and $b_1 = b_2 - x_2 + y_2$.

For, $n=2$, $f_2(b_2)=\max\limits_{x_2,y_2}\left[p_2x_2-c_2y_2+f_1(b_2-x_2+y_2)\right]$

where $y_2 \le H-b_2+x_2$

$\le 500-b_2+x_2$.

Therefore, $f_2(b_2)=\max\limits_{x_2}\left[26b_2-x_2+500\right]$

$=26b_2+500$ [taking $x_2=0$ for maximum]

and $b_2=b_3-x_3+y_3$.

For $n=3$,

$$f_3(b_3)=\max_{x_3,y_3}\left[p_3x_3-c_3y_3+f_2(b_3-x_3+y_3)\right]$$

$$=\max_{x_3,y_3}\left[25x_3-24y_3+26(b_3-x_3+y_3)+500\right]$$

$$=\max_{x_3,y_3}\left[26b_3-x_3+2y_3+500\right] \text{ where } y_3 \le 500-b_3+x_3$$

$$=\max_{x_3}\left[26b_3-x_3+2(500-b_3+x_3)+500\right]$$

$$=\max_{x_3}\left[24b_3+x_3+1500\right]$$

$=25b_3+1500$ [since $b_3 \ge x_3$, therefore $b_3=x_3$ for maximum]

But, $b_3=b_4-x_4+y_4$.

Now, taking $n=4$,

$$f_4(b_4)=\max_{x_4,y_4}\left[p_4x_4-c_4y_4+f_3(b_4-x_4+y_4)\right]$$

$$=\max_{x_4,y_4}\left[28x_4-27y_4+25(b_4-x_4+y_5)+1500\right]$$

$$=\max_{x_4,y_4}\left[25b_4+3x_4-2y_4+1500\right]$$

$=25b_4+3b_4+1500$ [since $y_4=0, x_4=b_4$ for maximum]

$=28b_4+1500$.

It is given that $b_4=200$.

Therefore, $b_3=200-200+0=0$

$b_2=0-0+500=500$

$b_1=500-0-0=500$

$x_4=200, y_4=0$

$x_3 = 0, y_3 = 500$

$x_2 = 0, y_2 = 0$

$x_1 = 500, y_1 = 0.$

Thus, the required solution is given below:

Month:	1	2	3	4
Purchase:	0	500	0	0
Sales:	700	0	500	500.

Maximum possible return $= 28 \times 200 + 1500 = 7100.$

109.11. Application in Linear Programming Problem

The general linear programming problem is

Maximize $z = c_1x_1 + c_2x_2 + ... + c_nx_n$

subject to

$$a_{11}x_1 + a_{12}x_2 + ... + a_{1x}x_n \le b_1$$

$$a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2$$

....

....

$$a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le b_m$$

$$x_1 \ge 0, x_2 \ge 0, ..., x_n \ge 0.$$

Let each activity, $j (= 1, 2, ..., n)$ be a stage. The level of activty, $x_j (\ge 0)$, represents decision variable at stage j. Since x_j is continuous, each stage possesses an infinite number of alternatives within the feasible region.

Since the linear programming problem is an allocation problem, states may be defined as the amounts of resources to be allocated to the current stage and suceeding stages. This will result in a backward functional (recursive) equation. Since there are m resources, states, must be represented by m-dimensional vector.

Further, let $(\beta_{1j}, \beta_{2j}, ..., \beta_{mj})$ be the states of the system at stage j in accordance with the definition, i.e.

amounts of resources 1, 2, 3, ..., m, respectively, allocated to stage j, j+1, ..., n. Using backward recursive equation, let $f_j(\beta_{1j}, \beta_{2j}, \ldots, \beta_{mj})$ be the optimum value of the objective function for stages j, j+1, ... for given states $\beta_{1j}, \beta_{2j}, \ldots, \beta_{mj}$.

Thus

$$f_n(\beta_{1n}, \beta_{2n}, \ldots, \beta_{mn}) = \max_{0 \le a_{in} x_n \le \beta_{in}} [c_n x_n].$$

$$i = 1, 2, \ldots, m.$$

$$f_j(\beta_{1j}, \beta_{2j}, \ldots, \beta_{mj}) = \max_{0 \ge ax \ge \beta} \left[c_j x_j + f_{j+1}(\beta_{1j} - a_{1j}x_j, \ldots, \beta_{mj} - a_{mj}x_j)\right]$$

$$(i = 1, 2, \ldots m)$$

for $j = 1, 2, 3, \ldots, n-1$.

Thus, a recursive equation is obtained and can be used to solve the linear programming problem by the dynamic programming approach.

Example 1. Solve the following linear programming problem by dynamic programming method.

Maximize $z = 2x_1 + 5x_2$

subject to

$$2x_1 + x_2 \le 43$$

$$2x_2 \le 46$$

$$x_1, x_2 \ge 0.$$

Solution. Since there are two resources, the state of equivalent dynamic programming problem can be described by two variables only.

Let (u_j, v_j) describe state j (=1, 2). Thus,

$$f_2(u_2, v_2) = \max_{\substack{0 \le x_2 \le u_2 \\ 0 \le 2x_2 \le v_2}} [5x_2]$$

Since $x_2 \le \min(u_2, v_2/2)$ and $f_2(x_2/\text{given } u_2, v_2)$ then

$$f_2(u_2, v_2) = \max f_2(x_2/\text{given } u_2, v_2) = 5\min(u_2, v_2/2)$$

and $\quad x_2^* = \min(u_2, v_2/2)$.

Now, $f_1(u_1, v_1) = \max\limits_{\substack{0 \le 2x_1 \le u_1 \\ 0 \le 0x_1 \le v_1}} \left[2x_1 + f_2(u_1 - 2x_1, v_1 - 0)\right]$

$$= \max_{0 \le 2x_1 \le u_1} \left[2x_1 + 5\min(u_1 - 2x_1, v_1/2)\right]$$

[by definition of f_2]

Since this is the last stage, then $u_1 = 43, v_1 = 46$.

Thus, $x_1 \le u_1/2 = 21.5$

and $f_1(x_1/\text{given } u_1, v_1) = f_1(x_1/\text{given } u_1 = 43, v_1 = 46)$

$$= 2x_1 + 5\min(43 - 2x_1, 46/2)$$

$$= 2x_1 + \begin{cases} 5(23), \text{for } 0 \le x_1 \le 10 \\ 5(43 - 2x_1) \text{for } 10 \le x_1 \le 21.5 \end{cases}$$

$$= \begin{cases} 2x_1 + 15, 0 \le x_1 \le 10 \\ -8x_1 + 215, 10 \le x_1 \le 21.5 \end{cases}.$$

Hence for given range of x_1,

$$(u_1, v_1) = f_1(43, 46) = \max_{x_1}(2x_1 + 115, -8x_1 + 215)$$

$$= \max[2\times10 + 115, -8\times10 + 215]$$

$$= 135$$

which is achieved at $x_1^* = 10$.

To obtain x_2^*, it is observed that

$$u_2 = u_1 - 2x_1 = 43 - 20 = 23, v_2 = v_1 - 0 = 46$$

and $x_2^* = \min(u_2, v_2/2) = \min(23, 46/2) = 23$.

Thus, the optimal solution is given by

$$z^* = 135, x_1 = 10, x_2 = 23.$$

Note: This example shows that the dynamic programming method to solve linear programming problem is much complicated than simplex method.

Example 2. Use dynamic programming method to solve the L.P.P.

Maximize $z = x_1 + 9x_2$

subject to

$2x_1 + x_2 \le 25,$

$x_1 \le 11$

$x_1, x_2 \ge 0.$

Solution. The problem has two resources and two decision variables. The states of the equivalent dynamic programming are β_{1j} and β_{2j} for $j = 1, 2$. Thus,

$$f_2(\beta_{12}, \beta_{22}) = \max\{9x_2\}$$

where maximum is taken over $0 < x_2 \le 25$ and $0 \le x_2 \le 11$.

That is,

$$f_2(\beta_{12}, \beta_{22}) = 9\max\{x_2\}$$
$$= 9\max\{25, 11\}.$$

Since the maximum of x_2 satisfying the conditions of $x_2 \le 25$ and $x_2 \le 11$ is the minimum of 25 and 11.

Therefore, $x_2^* = 11$.

Now, $f_1(\beta_{11}, \beta_{21}) = \max\{x_1 + f_2(\beta_{11} - 2x_1, \beta_{21} - 0)\}$

where maximum is taken over $0 \le x_1 \le 25/2$.

At this last stage, substitute the values of $\beta_{11} = 25$ and $\beta_{21} = 11$.

Therefore, $f_1(25, 11) = \max\{x_1 + 9\min(25 - 2x_1, 11)\}$.

Now, $\min(25 - 2x_1, 11) = \begin{cases} 11, \text{ for } 0 \le x_1 \le 7 \\ 25 - 2x_1, \text{ for } 7 \le x_1 \le 25/2. \end{cases}$

Therefore, $x_1 + 9\min(25 - 2x_1, 11) = \begin{cases} x_1 + 99 \text{ for } 0 \le x_1 \le 7 \\ 225 - 17x_1, \text{ for } 7 \le x_1 \le 25/2. \end{cases}$

Since the maximum of both $x_1 + 99$ and $225 - 17x_1$ occurs at $x_1 = 7$, therefore

$$f_1(25, 11) = 7 + 9\min(11, 11)$$
$$= 106 \text{ at } x_1^* = 7$$

$$x_2^* = \min(25 - 2x_1^*, 11) = \min(11, 11) = 11.$$

Hence the optimal solution is

$x_1^* = 7, x_2^* = 11$ and $max\ z = 106$.

109.12. Module Summary

In this module, we have introduced a very powerful mathematical technique-dynamic programming method which is used to solve different kinds of optimization problems. Using this method we have solve different types of optimization problems containing one or more constraints. This method is also used to solve problems in inventory control and also linear programming problems. The module is ended by an exercise and references.

109.13. Self Assessment Questions

1. State the 'Principle of optimality' in dynamic programming and give a mathematical formulation of a dynamic programming problem.
2. State Bellman's principle of optimality in dynamic programming.
3. Use Bellman's principle of optimality to solve the following problem:

   Maximize $z = x_1 x_2 x_3 .... x_n$

   subject to $x_1 + x_2 + x_3 + ... + x_n = c$

   $x_1, x_2, x_3, ..., x_n \geq 0.$
4. Obtain the functional equation for maximizing

   $$z = g_1(x_1) + g_2(x_2) + ... + g_n(x_n)$$

   subject to $x_1 + x_2 + ... + x_n = c$

   $x_j \geq 0, j = 1, 2, 3, ..., n.$
5. Find the minimum value of $x_1 + x_2 + ... + x_n$, when $x_1 x_2 x_3 ... x_n = d, x_1, x_2, x_3 ...., x_n \geq 0.$
6. Use the principle of optimality to solve the problem

   Minimize $z = \sum_{j=1}^{N} x_j \alpha$

   subject to $x_1 x_2 x_3 ... x_n = r,$

   $x_i \geq 1, i = 1, 2, ..., N,$

   where $r \geq 1$ and $\alpha > 0$ are given fined numbers.

7. Find the maximum value of

$$z = x_1^2 + 2x_2^2 + 4x_3$$

subject to $x_1 + 2x_2 + x_3 \le 8$

$$x_1, x_2, x_3 \ge 0.$$

8. Find the maximum value of

$$z = -x_1^2 - 2x_2^2 + 3x_2 + x_3$$

subject to the conditions

$$x_1 + x_2 + x_3 \le 1, x_1, x_2, x_3 \ge 0.$$

9. Use method of dynamic programming to minimize

$$u_1^2 + u_2^2 + u_3^2$$

subject to $u_1 + u_2 + u_3 \ge 10, u_1, u_2, u_3 \ge 0.$

10. Solve the following linear programming problems using dynamic programming method.

(a) Maximize $z = 8x_1 + 7x_2$

subject to $2x_1 + x_2 \le 8, 5x_1 + 2x_2 \le 15, x_1, x_2 \ge 0.$

(b) Maximize $z = 3x_1 + 5x_2$

subject to the constraints

$$x_1 \le 4, x_2 \le 6, 3x_1 + 2x_2 \le 18, x_1, x_2 \ge 0.$$

(c) Max $z = 3x_1 + x_2$

subject to $2x_1 + x_2 \le 6, x_1 \le 2, x_2 \le 4, x_1, x_2 \ge 0.$

109.14. Suggested Further Readings

1. S.D. Sharma, Operations Research, Kedar Nath Ram Nath & Co.
2. J.K. Sharma, Operations Research, Macmillan.
3. G Hadley, Non-Linear and Dynamic Programming, Addison-Wesley, Reading Mass.
4. G Nembauser, Introduction to Dynamic Programming, Wiley, New York.
5. K. Swarup, P.K. Gupta and M. Mohan. Operations Research, Sultan Chand.

---- 0 ----

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-VI Group–B

Module No. - 110
OPTIMAL CONTROL

Structure :

1. Introduction : Most mathematical models in physical science, mechanics or economics can be characterised by a set of functions $x_1(t), x_2(t), ..., M_x(t)$ which satisfy a set of first order differentical equations of the form

$$\frac{dxi}{dt} = fi(x_i, x_2..., x_n; u_1, u_2, ..., un; t) \quad (1)$$

Where fi is a function of $x_1, x_2, ..., x_n; u_1, u_2, ..., un$ and t. The variables $u_1, u_2, ..., un$ are called the control variables and $x_1, x_2, ..., xn$ are called state variables.

Optimal control deals with the problem of finding a control law for such a given system of differential equations describing he paths of the control variables that minimizes a cost functional which is a function of the state variables and control variables.

For a constrained set of equations, there will be some oundary conditons. If to $<t<t_1$ is the time interval under consideration, then x_i's are normally specified at t = to i.e., x (to) are given for $i = 1, 2, ..., n$. Thus for $u;j = 1, 2, ..., n$, the above set of equations (1) determines xi's uniquely. Normal control problem is to choose the controls $u_1, u_2, ..., un$ so that the system moves from its unitial state $\{x_1(0), x_2(0), ..., x_n(0)\}$ to some prescribed state at $t - t_1$. If this can be achieved, we say that the system is controllable.

If there exist more than one set of controls which control the system, then our problems is to find the set of controls which control the system in the best way, i.e., we must minimize or maximize the functional of the form

$$J = \int_0^{t} F(x_1, x_2..., x_n, u_1, u_2; t)\, dt$$

We organize this module into the following sections :

2. Performance indices : In modern control theory, the optimal control problem is to find a control which causes the dynamical system to reach a target or follow a state variable (ortrajectory) and at the same time extremize a performance index which may take several forms as described below.

(i) performance index for time-optimal control system"

Here, we try to transfer a system from an arbitrary initial state x_0 at time t_0 to a specified final state x_1 at time t_1 in minimum time

The corresponding performance index is $J = \int_{t_0}^{t_1} d_t = t_1 - t_0 = t^*(say)$

(ii) Performance index for fuel optimal control system :

Consider a spacecrat problem. Let $u(t)$ be the thrust of a rocket engine and assume that the magnitude $|u(t)|$ of the thrust is proportional to the rate of fuel consumption. In order to minimize the total expenditure of fuel, we may formulate the performance index as

$$J = \int_{t_0}^{t_1} |u(t)| dt$$

For several controls, we may write it as

$$J = \int_{t_0}^{t_1} \sum_{i=1}^{m} Ri|ui(t)| dt,$$ where R is a weighting function.

(iii) performance index for minimum energy control system :

Consider *ui* (*t*) as the current in the *i*th loop of an electric network then $\sum_{i=1}^{m} ui^2(t)$ (Where, *ri* is the resistance of the *i*-th loop) is the total power or total rate of energy expenditure of the network, then for minimization of the total expended energy, we have a performance criterion as $J = \int_{t_0}^{t_1} \sum_{i=2}^{m} ui^2(t) r_i dt$

or, in general $J = \int_{t_0}^{t_1} u'(t) R u(t) dt$

Where *R* is a positive definite matrix and prime (') denotes the transpose.

Similarly, we can think of minimization of the integral of the squared error of a tracking system. We then have $J = \int_{t_0}^{t_1} x'(t) Q(x) t\, dt,$

Where $x_d(t)$ is the desired value, $x_a(t)$ is the actually obtained value and $x(t) = x_a(t) - x_d(t)$, is the error. Here, *Q* is a weighting matrix which can be positive semi-difinite.

(iv) performance index for terminal control system : In a terminal target problem we are interested in minimizing the error between the desired target position $x_d(ti)$ and the actual target position *xa)ti)* and the end of the manever or at the final time *ti*. The terminal (final) error is $x(t1) = uxa(t) - xd(ti)$. Taking care of positive and negative values or error and weghting factors, we construct the cost functin as $J = x'(t_1\; V\; x(t_1)$

Which is also called the terminal cost function. Here *V* is positive semi definite matrix.

(v) performance index for general optimal control systems "Combining the above formulations, we have a performance index in general form as

$$J = x'(t_1) V x(t_1) + \int_{t_0}^{t_1} [x'(t) Qx)t) + u'(t) R u(t)] dt \qquad (2)$$

or, $J = G(x(t_1), t_1) + \int_{t_0}^{t_1} f(x(t), u(t), t) dt$

Where *R* is a positive definite materix, and *Q* and *V* are positive semi-definite matrices, respectively. Note that the matrics *Q* and *R* may be time varying. The particular form of performance index given by equation (2) is called a quadratic (in terms of states and controls) form.

The problems arising in optimal controls are classified based on the structure of the performance index J. If the performance index (PI) contains the terminal cost function $G(x(t),...u(t), t)$ only, it is called the Mayer problem, if the PI has only the integrl cost term, it is called the lagrage problem and the problem is of the Bolaz type if the PI contains both the terminal cost term and the integral cost terms.

However, in this coure module, we intend to study the control problems of Lagrange type only.

It is also to be noted that the performance indices are sometimes referred to as payoff functionals.

3. Calculus of variations : To optimize a quantity which appears as an integral, we use the calculus of variations. In other words, calculus of variation can be applied to obtain the necessary condition for a quantity appearing as an integral has either a minimum or a maximum value, i.e., stationary value.

Let us consider the simplest integral.

$$J = \int_{t_0}^{t_1} F(x,\dot{x},t)\,dt, \text{where } \dot{x} = \frac{dx}{dt} \qquad (3)$$

Here, F is a known function of x, m. and t, but the function $x(t)$ is unknown. The problem is to find a path $x = x(t), t0 \le t \le t1$ which optimize the functional J.

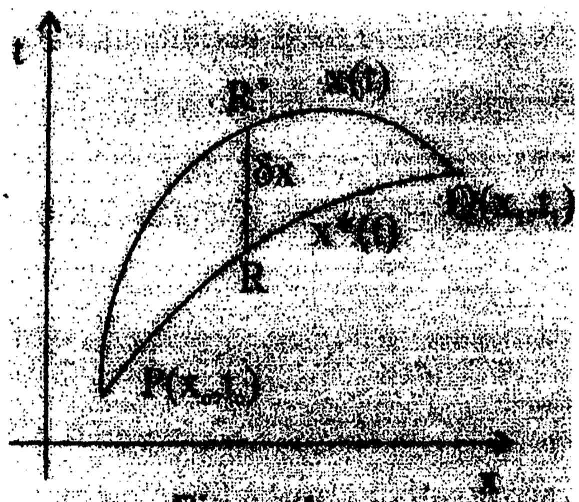


Figure-1

The value of J is different along different paths connecting the points $P(x0, t0)$ and $Q(x1, t1)$. We have to choose the path of integration $x(t)$ such that J has stationary value. We consider two paths out of infinite number of possible paths. The difference between the values of x for thee two paths for a given value of t, is the variation of x, which is denoted by δx and may be described by introducing a new function $\eta(t)$ and ε to describe the arbitrary deformation of the path and the magnitude of variation respectively.

The function $\eta(t)$ must satisfy the following two conditions :

(i) All varied paths must pass through the fixed points P & Q, i.e. $\eta(to) = \eta(ti) = 0$

(ii) $\eta(t)$ must be differentiable.

Let *PRQ* be the optimum path, i.e., *J* is optimum along *PRQ* and let it be given by $x = x^*(t)$. Also let a varied path is given by $x = x^*(t) + \delta x = x^*(t) + h(t).\varepsilon$.

Thus the value of *J* on *PR'Q* is $J = \int_{t_0}^{t_1} F\left(x^*(t)+\eta(t)\varepsilon, \dot{x}^*(t)\varepsilon, t\right)dt$ where $\dot{\eta}(t) = \dfrac{d\eta}{dt}$

Hence, for a given $\eta(t)$, *J* is a function of ε only.

Therefore, $J(E) = \int_{t_0}^{t_1} F\left(x^*(t)+\eta(t)e, x^*(t)+\eta\cdot(t)\varepsilon, t\right)dt.$

The condition for extremum of *J*(ε) is $\dfrac{dJ(\varepsilon)}{d\varepsilon} = 0$

Also from the figure, it is clear that *J*(*e*) will be optimum where *e*=o.

Thus the necessary condition for optimization of *J* is $\dfrac{dJ(\varepsilon)}{d\varepsilon} = 0$

Also from the figure, it is clear that *J*(ε) will be otpimum where ε= o.

Thus the necessary condition for optimization of *J* is $\dfrac{dJ(\varepsilon)}{d\varepsilon} = 0$ for ε= o.

$$\text{Now } \frac{dJ}{d\varepsilon} = \int_{t_0}^{t_1}\left(\frac{\delta F}{\delta x}\frac{dx}{d\varepsilon} + \frac{\delta F}{\delta x}\frac{dx}{d\varepsilon} + \frac{\delta F}{\delta t}\frac{dx}{d\varepsilon}\right)dt$$

$$= \int_{t_0}^{t_1}\left(\frac{\delta F}{\delta x}\eta(t) + \frac{\delta F}{\delta x}\eta(t) + \frac{\delta F}{\delta t}\cdot 0\right)dt$$

$$= \int_{t_0}^{t_1}\left\{F_x\left(x^*(t)+\eta(t)\varepsilon, \dot{x}^*(t)+\dot{\eta}(t)\varepsilon, t\right)\cdot\eta(t)\right.$$

$$\left. + F_{\dot{x}}\left(x^*(t)+\eta(t)\varepsilon, \dot{x}^*(t)+\dot{\eta}(t)\varepsilon, t\right)\cdot\dot{\eta}(t)\right.$$

Since $\dfrac{dJ}{d\varepsilon} = 0$ for $e = 0$, therefore,

13. PROJECT TIME-COST TRADE OFF

In this section, the cost of resources consumed by activities are taken into consideration. The project completion time can be reduced (crshing) by reducing the normal completion time of critical activities. The reduction in normal time of completion will increase the total budget of the project. However, the decision-maker will always look for trade off between the total cost of the project and the total time required to complete it.

Project cost

In order to include the cost aspect in project scheduling we have to find out the cost duration relationships for various activities in the project. The total cost of any project comprises direct and indirect costs.

Direct Cost

This cost is directly dependent upon the amount of resources in the execution of individual activiteis such as manpower loading, material consumed etc. The direct cost increases if the activity duration is to be reduced.

Indirect Cost

This cost is associated with expenditures which can not be allocated to invidiual activities of the project. This cost may include managerial services, loss of revenue, fixed overheads etc. The indirect cost is computed oon per day, per week or per month basis. This cost decreases if the activity duration is to be reduced.

The network diagram can be used to identify the activities whose duration should be shortened so that the completion time of the project can be shortened in the most economic manner. The process of reducing the activity duration by putting on extra effort is called crashing the activity.

The crash time (T_c) represents the minimum activity duration time that is possible and any attempts to further crash would only raise the activity cost without reducing the time. The activity cost corresponding to the crash time is called the crash cost (C_c) which is the minimum direct cost required to achieve the crash performance time.

The normal cost (C_n) is equal to the absolute minimum of the direct cost required to perform an activity. The corresponding time duration taken by an activity is known as the normal time (T_n).

12. DIFFERENCE BETWEEN PERT AND CPM

PERT	CPM
1. The technique was developed in connection with R & D (Research & Development) works, therefore it had to scope with the uncertainty which is associated with R & D activities. In this cae, the total project duration is regarded as a random variable. As a result, multiple time estimates are made to calculate the probability of completing the project within the schedule time. Therefore, it is a probabilistic model.	1. This technique was developed in connection with a construction and maintenance project which consists of routine tasks or jobs whose resource requirement and duration is known with certainty. Therefore, it is basically a deterministic model.
2. It is used for projects involving activities of non-repetitive nature.	2. It is used for projects involving activities of repetitive nature.
3. It is event oriented technique because the results of analysis are expressed in terms of events or distinct points in time indicative of progress.	3. It is actively oriented technique as the results of calculations are considered in terms of activities.
4. It incorporates statistical analysis and there by enables the determination of probabilities concerning the time by which each activity and the entire project would be completed.	4. It does not incorporate statistical analysis in determining the estimates because time is precise and known.
5. It serves a useful control device as it assists the management in controlling a project by calling attention through constant review to such delays in activities which might lead to a delay in the project completion date.	6. It is difficult to use this technique as a controlling device for the simple reason that one must repeat the entire evaluation of the project each time the changes introduced into the network.

(i) Now, the probability that the project will be completed at least 4 weeks earlier than expected is given by

$p(Z \le D)$ where $D = \frac{(17-4)-17}{3} = \frac{-4}{3} = -1.33$ (approx.)

$p(Z \le 1.33)$

$= -0.5 - P(0 < Z \le 1.33)$

$= 0.5 - \varphi(1.33)$

$= 0.5 - 0.4082$ [From the table of area under standard normal curve]

$= 0.0918$

(ii) Again, the probability that the project will be completed not more than 4 weeks later than expected = the probability that the project will be completed within 17 + 4 i.e., within 21 weeks

$p(Z \le D)$ where $D = \frac{21-17}{3} = \frac{4}{3} = -1.33$ (approx.)

$= p(Z \le 1.33)$

$= 0.5 + \varphi(1.33) = 0.5 + 0.4082 = 0.9082$

(f) When the due date is 19 weeks, $D = \frac{19-17}{3} = \frac{2}{3} = 0.67$

Then the probability of meeting the due date is given by

$p(\dot{Z} \le 0.67) = 0.5 + \varphi(0.67) = 0.5 + 0.2514 \backslash 0.7514$

(g) Since the probability for the completion of the project is 0.90,

$p(Z \le D) = 0.90$ where $D = \frac{D_s - T_e}{\sigma}$ and T_s be the schedule time

As $p(Z \le 1.29) = 0.90$

$\therefore D = 1.29$ which implies $\frac{T_s - 17}{3} = 1.29$ i.e., $T_s = 17 + 3 \times 1.29$ *i.e.* $T_s = 20.87$

(b) Forward pass calculatons

Let E_i be the earliest occurence time of event i.

Set $E_1 = 0$

$E_2 = E_1 + t_{12} = 0 + 2 = 2$

$E_3 = E_1 + t_{13} = 0 + 4 = 4$

$E_4 = E_1 + t_{14} = 0 + 3 = 3$

$$E_5 = \max_{i=2,3}\{E_i + t_{i5}\} = \max\{E_2 + t_{25}, E_3 + t_{35}\} = \max\{2+1, 4+6\} = 10$$

$$E_6 = \max_{i=4,5}\{E_i + t_{i6}\} = \max\{E_4 + t_{46}, E_5 + t_{56}\} = \max\{3+5, 10+7\} = 17$$

Backward pass calculations

Let L_j be the latest occurrence time of event j

Set $L_6 - E_6 = 17$

$L_5 = L_6 - t_{56} = 17 - 7 = 10$

$L_4 = L_6 - t_{46} = 17 - 5 = 12$

$L_3 = L_5 - t_{35} = 10 - 6 = 4$

$L_2 = L_5 - t_{25} = 10 - 1 = 9$

$$L_1 = \min_{j=2,3,4}\{L_j - t_{1j}\} = \min\{L_2 - t_{12}, L_3 - t_{13}, L_4 - t_{14}\} = \min\{9-2, 4-4, 12-3\} = 0$$

From the above calculations, it is seen that

$E_1 = L_1, E_3 = L_3, E_5 = L_5, E_6 = L_6$

Hence the critical events are 1, 3, 5, 6 and the critical path is 1 – 3 – 6 – 6. Also, the expected project length $= E_6 = L_6 = 17$ weeks.

(d) Variance of the project length is given by

$\sigma_e^2 = 1 + 4 + 4 = 9$, or, $\sigma = 3$

(e) The standard normal deviate is given by

$$D = \frac{\text{schedule time - expected time effect impletion}}{\sqrt{\text{variance or sttandard deviation}}}$$

Solution :

(a)

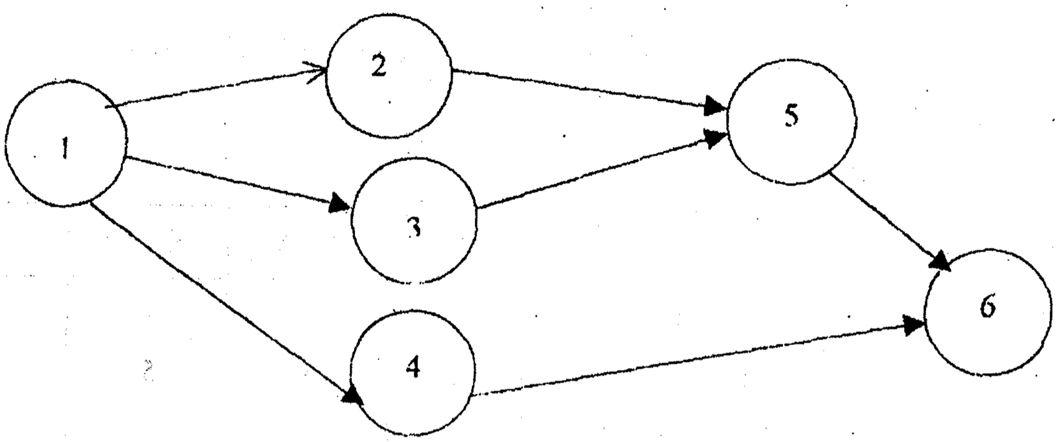
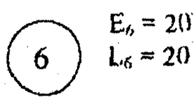


Fig. 13

(a) The expected time and variance of each activity is computed and displayed in the following table.

Activity	t_0	t_m	t_p	$t_e = \dfrac{t_0 + 4t_m + t_p}{6}$	$\sigma^2 = \left(\dfrac{t_p - t_0}{6}\right)^2$
1 – 2	1	1	7	2	1
1 – 3	1	4	7	4	1
1 – 4	2	2	8	3	1
2 – 5	1	1	1	1	0
3 – 5	2	5	14	6	4
4 – 6	2	5	8	5	1
5 – 6	3	6	15	7	4



The values of the probabilities for a normal distribution curve corresponding to the different values of normal deviate are available from the table of standard normal curve.

Example 3

A small project is composed of seven activities whose time estimates (in weeks) are listed in the following table:

Activity	1 – 2	1 – 3	1 – 4	2 – 5	3 – 5	4 – 6	5 – 6
t_0	1	1	2	1	2	2	3
t_m	1	4	2	1	5	5	6
t_p	7	7	8	1	14	8	15

(a) Draw the project network.

(b) Find the expected duration and variance of each activity.

(c) Calculate the earliest and latest occurrence time for each event and the expected project length.

(d) Calculate the variance and standard deviation of project length.

(e) What is the probability that the project will be completed?

(i) At least 4 weeks earlier than expected?

(ii) Not more than 4 weeks later than expected?

(f) If the project due date is 19 weeks, what is the probability of meeting the due date?

(g) Find also the schedule time on which the project will be completed with a probability 0.90.

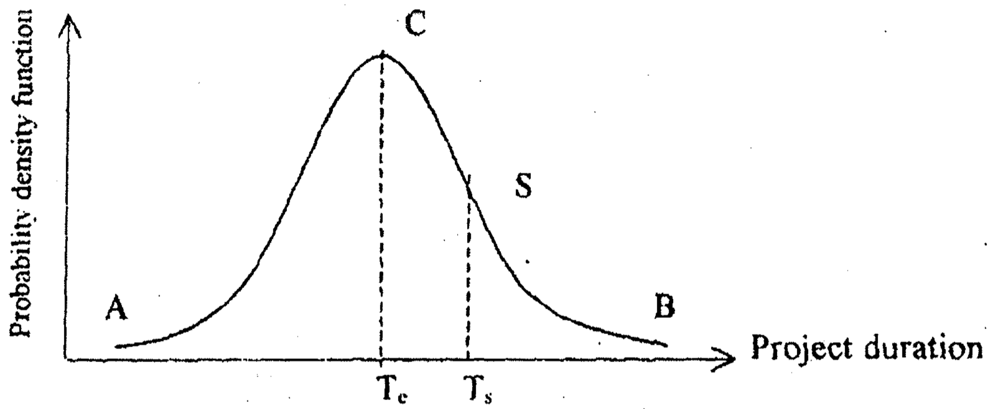


Fig.-12

The probability of completing a project in time T_s is given by

$$p(T_s) = \frac{Area\ under\ ACS}{Area\ under\ ACB}$$

The $p(T_s)$ depends upon the location of T_s, Taking T_e as reference point the distance $T_s - T_e$ can be expressed in terms of standard deviation for a network is calculated as

Standard deviation for a network $= \sigma_e = \sqrt{\text{sum of the variances along the critical path}}$

i.e., δ for a network $= \sqrt{\sum \sigma_{ij}^2}$ where σ_{ij}^2 for an activity $(i-j) = \left(\frac{t_p - t_0}{6}\right)^2$

Since the standard deviation for a standard normal curve is unity, the standard deviation σ_e, calculated above is used as scale factor for calculation the normal deviate.

The normal deviation $D = \frac{T_s - T_e}{\sigma_e}$

Hence the probability of completing the project by scheduling time (T_s) is given by

$p(Z \le D)$ where $D = \frac{T_s - T_e}{\sigma_e}$ and Z is the standard normal variate.

Let us recall that the expected time of an activity is the weighted average of three time estimates t_0, t_p and t_m,

i.e. $t_e = (t_0 + 4t_m + t_p)/6$

The probability that the activity $(i-j)$ will be completed in time t_e is 0.5 i.e., the chance of completion of that activity is 50%. In the frequency distribution curve, for the activity (i,j) the vertical line through t_e will divide the area under the curve in two equal parts as shown in the following figure.

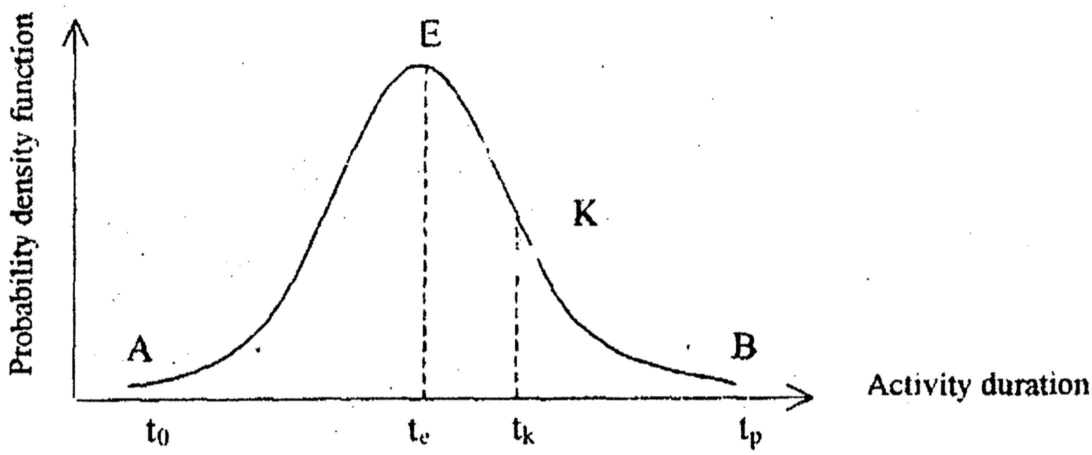


Fig.-11

For completing the activity in any other time t_k, the probability will be

$$p = \frac{Area\ under\ AEK}{Area\ under\ AEB}$$

A project consists of a number of activities. All activities as we know are independent random variables and hence the length of the project upto a certain event through a certain path is also a random variable. But the point of different is that expected project length T_e does not have the same probability distribution as the expected activity time t_e. While a Beta distribution curve approximately represents the activity time probability distribution, the project expected time T_e follows approximately a standard normal distribution. This standard normal distribution curve has an area equal to unity and standard deviation 1 and is symmetrical about the mean as follows:

The range specified by the optimistic time (t_0) and pessimistic time (t_p) estimates supposedly must close every possible estimate of duration of the activity. The most likely time (t_m) estimate may not coincide with the mid point $t_{mid} = (t_0 + t_p)/2$ and may occur to its left or right as shown in **Fig. 10.**.

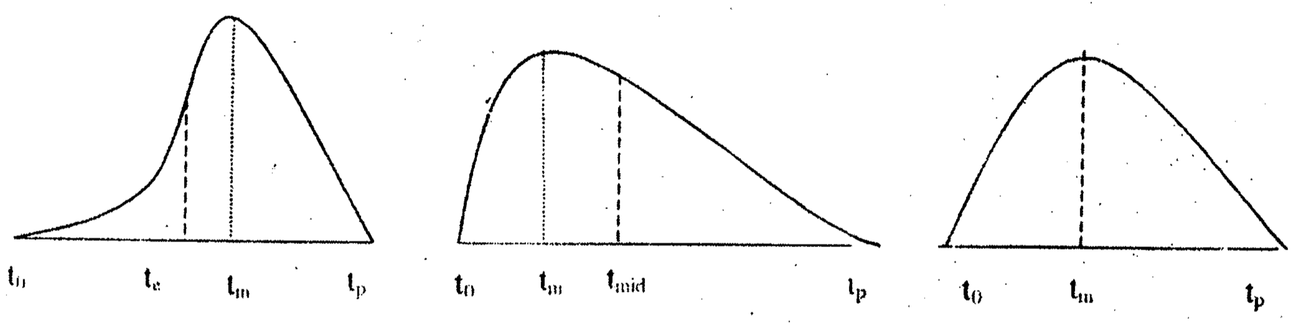


Fig. 10

Keeping in view of the above mentioned properties, it may be justified to assume that the duration of each activity may follow Beta (β) distribution with its unimodal point occurring at t_m and its end point at t_0 and t_p.

The expected or mean value of an activity duration can be approximated by a linear combination of three time estimates or by the weighted average of three time estimates t_0, t_p and t_m, i..e., $t_e = (t_0 + 4t_m + t_p)/6$

Again, to determine the activity duration variance in PERT, the unimodal property of β-distribution is used. However, in PERT, the standard deviation is expressed as

$$\sigma = \frac{1}{6}(t_p - t_0) \text{ or, variance } \sigma^2 = \left(\frac{t_p - t_0}{6}\right)^2$$

It is noted that in PERT analysis, Beta distribution is assumed because it is unimodal, has non-negative end points and is approximately symmetric.

Probability of meeting the schedule time

After identifying the critical path and the occurrence time of all activities, there arises a quetion - what is the probability that a particular event will occur on or before the schedule data? This particular event may be any event in the network.

(2, 5)	19	20	39	38	57	18	0	0	
(3, 4)	16	23	39	23	39	0	0	0	(3, 4)
(3, 7)	24	23	47	43	67	20	20	20	
(4, 5)	0	39	39	57	57	18	0	0	
(4, 6)	18	39	57	39	57	0	0	0	(4, 6)
(5, 6)	0	39	39	57	57	18	18	0	
(5, 7)	4	39	43	63	67	24	24	46	
(6, 7)	10	57	67	57	67	0	0	0	(6, 7)

From the aboe table it is clear that the critical activities (zero total float) are (1, 3), (3, 4), (4, 6) and (6, 7). hence the critical path is 1 – 3 – 4 – 6 – 7 and the duration of the project is 67 time limits (as $E_7 = L_7 = 67$).

11. PERT ANALYSIS

Time estimates

It is difficult to estimate time required for the execution of each activity or because of various uncertainties. Taking the uncertainties into account, three types of time estimates are generally obtained.

The PERT system is based on these three time estimates of the performance time of an activity.

(i) *Optimistic time* (t_0) : This is the estimate of the shortest possible time in which an activity can be completed under ideal conditions.

(ii) *Pessimistic time* (t_p) : This is the maximum time which is required to perform the activity under extremely bad conditions. However such conditions do not include labour strike or acts of nature (like flood, earthquake, tornado etc.)

(iii) *Most – likely time* (t_m) : This is the estimate of the normal time in which an activity would take. This time estimate lies between the optimistic and pessimistic time estimates. Statistically, it is the modal value of duration of the activity.

At node 5 : $E_4 = \max_{i=2,4}\{E_i + t_{i5}\} = \max\{E_2 + t_{25}, E_4 + t_{45}\} = \max\{20+19, 39+0\} = 39$

At node 6 : $E_6 = \max_{i=4,5}\{E_i + t_{i6}\} = \max\{E_4 + t_{46}, E_5 + t_{56}\} = \max\{39+18, 39+0\} = 57$

At node 7 :

$$E_7 = \max_{i=3,5,6}\{E_i + t_{i7}\} = \max\{E_3 + t_{37}, E_5 + t_{57}, E_6 + t_{67}\} = \max\{23+24, 39+4, 57+10\} = 67$$

Backward Pass calculations

At node 7 : Set $L_7 = E_7 = 67$

At node 6 : $L_6 = L_7 - t_{67} = 67 - 10 = 57$

At node 5 : $L_5 = \min_{j=6,7}\{E_j + t_{5j}\} = \min\{L_6 - t_{56}, t_7 - t_{57}\} = \min\{57-0, 67-4\} = 57$

At node 4 : $L_4 = \min_{j=5,6}\{L_j - t_{4j}\} = \min\{L_5 - t_{45}, L_6 - t_{46}\} = \min\{57-0, 57-18\} = 39$

At node 3 : $L_3 = \min_{j=4,7}\{L_j - t_{3j}\} = \min\{L_4 - t_{34}, L_7 - t_{37}\} = \min\{39-16, 67-24\} = 23$

At node 2 : $L_2 = L_5 - t_{25} = 57 - 19 = 38$

At node 1 : $L_1 = \min_{j=2,3,4}\{E_j - t_{1j}\} = \min\{L_2 - t_{12}, L_3 - t_{13}, L_4 - t_{14}\}$

$$= \min\{38-20, 23-23, 39-8\} = 0$$

To find the critical activities and different floats, we construct the following table.

Activity	Duration of activity	Earliest time		Latest time		Float			Critical Activity
		Start (E_i)	Finish (E_i+t_{ij})	Start (L_j-t_{ij})	Finish (L_j)	Total $L_j-E_j-t_{ij}$	Free $E_j-E_i-t_{ij}$	Independent $E_j-L_i-t_{ij}$	
(1, 2)	20	0	20	18	38	18	0	0	
(1, 3)	23	0	23	0	23	0	0	0	(1, 3)
(1, 4)	8	0	8	31	39	31	31	31	

(i) If the project has to be shortened, then some of the activities on that path must be shortened. The application of additional resources on other activities will not give the desired results unless that critical path is shortened first.

(ii) The variation in actual performance from the expected activity duration time will be completely reflected in one-to-one fashion in the anticipated completion of the whole project.

Example 2

Determine the critical path, minimum time of completion of the project whose network diagram is shown below. Find also the different floats.

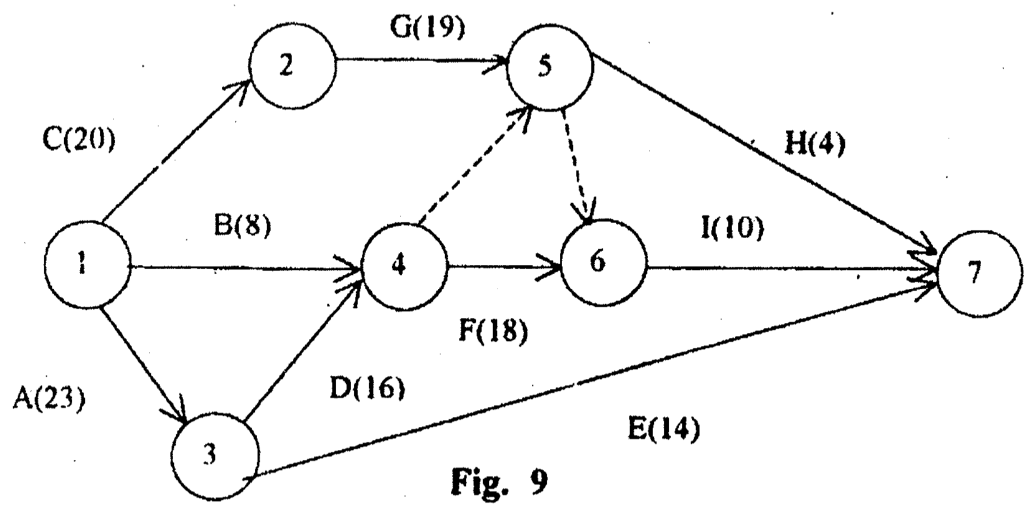


Fig. 9

Solution :

Forward Pass Calculations

At node 1 : Set $E_1 = 0$

At node 2 : $E_2 = E_1 + t_{12} = 0 + 20 = 20$

At node 3 : $E_3 = E_1 + t_{13} = 0 + 23 = 23$

At node 4 : $E_4 = \max_{i=1,3}\{E_i + t_{i4}\} = \max\{E_1 + t_{14}, E_3 + t_{34}\} = \max\{0+8, 23+16\} = 39$

Independent float

In some cases, the delay in the completion of an activity neither affects its predecessor nor the successor activities. This amount of delay is called independent float. Mathematically, independent of an activity (i, j) denoted by IF_{ij} is computed by the formula.

$$IF_{ij} = E_j - L_i - t_{ij}$$

The negative independent float is always taken as zero.

Event slack or Event float

The slack of an event is the difference between its latest time and its earliest time. Hence for an event i,

$$\text{Slack} = L_i = E_i$$

Critical Event : An event is said to be critical if its slack is zero i.e. $L_i = E_i$ for i-th event.

Critical activity : An activity is critical if its total float is zero i.e. $LS_{ij} = ES_{ij}$ or, $LF_{ij} = EF_{ij}$ for an activity (i, j).

Otherwise, an activity is called non-critical.

10.4 Critical Path

The continuous chain or sequence of critical activities in a network diagram path in the network from starting event to ending event and is shown by a dark line or double lines to make distinction from other non-critical path.

The length of the critical path is the sum of the individual times of all critical activities lying on it and defines the minimum time required to complete the project.

The critical path on a network diagram can be identified as

(*i*) For all activities (i, j) lying on the critical path, the E-values and L-values for tail and head events are equal *i.e.* $E_j = L_j$ and $E_i = L_i$.

(*ii*) On the critical path, $E_j - E_i = L_j - L_i - t_{ij}$

Main features of the critical path

The critical path has two main features :

10.3 Determination of floats and slack times

When the network is completely drawn, properly labeled, earliest and latest event times are computed and then the next object is to determine the floats of each activity and slack and slack time of each event.

The float of an activity is the amount of time by which it is possible to delay its completion time without affecting the total project completion time. There are three types of activity floats :

(i) Total float

(ii) Free float.

(iii) Independent float

Total float

The total float of an activity of an activity represents the amount of time by which an activity can be delayed without delay in the project completion time.

Mathematically, the total float of an activity (i,j) is the difference between the latest start time and earliest start time of that activity (or the different between the earliest finish time and latest finish time). Hence the total float for an activity (i,j) is denoted by TF_{ij} and is computed by the formula.

$$TF_{ij} = LS_{ij} - ES_{ij} \text{ or, } TF_{ij} = LF_{ij} - EF_{ij} \text{ or, } TF_{ij} = L_j - \left(E_i + t_{ij}\right) \text{as } L_j = LR_{ij} \text{ and } EF_{ij} = E_i + t_{ij}$$

Free float

Sometimes, it may be needed to know how much an activity's completion time may be delayed without causing any delay in its immediate successor activities. This amount of float is called free float. Mathematically, the free float for an activity (i,j) is denoted by EF_{ij} and is computed by

$$EF_{ij} = E_j - E_i - t_{ij}$$

As $TF_{ij} = L_i - E_i - t_{ij}$ and $L_i \geq E_j$

$$\therefore TF_{ij} \geq E_j - E_i - t_{ij} \text{ i.e., } TF_{ij} \geq FF_{ij}$$

Hence for all activities, free float can take values from zero upto total float but it does not exceed total float.

Again, free float is very useful for rescheduling the activities with minimum disruption of earlier plans.

Step - 4 : Go to next event (node), say event j ($j > i$) and compute the earliest occurrence time for event j. This is the maximum of the earliest finishing times of all activities ending into that event i.e., $E_j = Max\{EF_{ij}\} = \underset{i}{Max}\{E_i + t_{ij}\}$ for all immediate predecessor activities.

Step - 5 : If $i = n$ (final event number), then the earliest finishing time for the project is given by $E_n = \underset{i}{Max}\{EF_{ij}\} = \underset{i}{Max}\{E_i + t_{ij}\}$ for all terminal activities.

10.2 Backward pass calculations method

In this method, calculations begin from the terminal event, proceed through the events in a decreasing order of event numbers and end at the initial event of the network. At each node (event), the latest starting and finishing times are calculated for each activity. The method may be summarized as follows :

Step - 1 : Set $L_n = E_n, j = n$

Step - 2 : Calculate the latest finishing time LF_{ij} for each activity that ends at event j i.e., $LF_{ij} = L_j$ for all activities (i, j) that end at node j.

Step - 3 : Calculate the latest starting time LS_{ij} of each activity that ends at event j by subtracting the duration of each activity from the latest finishing time of the activitiy. Thus

$$LS_{ij} - t_{ij} = L_j - t_{ij}$$

Step - 4 : Proceed backward to the node in the sequence that decrease j by 1. Also compute the latest occurrence time of node $i(i < j)$. This is the minimum of the latest starting times of all activitieis starting from that event i.e.

$L_i = \underset{i}{Min}\{LS_{ij}\} = \underset{j}{Min}\{L_j - t_{ij}\}$ for all immediate successor activities.

Step - 5 : If $i = 1$ (initial node), then

$L_1 = \underset{i}{Min}\{LS_{ij}\} = \underset{j}{Min}\{L_j - t_{ij}\}$ for all initial activities.

Notations

The following notations are used in this analysis.

E_j = Earliest occurrence time of event i i.e., it is the earliest time at which the event i can occur without affecting the total project duration.

L_i = Latest allowable occurrence time of event i. It is the latest allowable time at which an event can occur without affecting the total project duration.

t_{ij} = Duration of activity (i,j)

ES_{ij} = Earliest starting time of activity (i,j)

LS_{ij} = Earliest starting time of activity (i,j)

EF_{ij} = Earliest finishing time of activity (i,j)

LF_{ij} = Latest finishing time of activity (i,j)

The critical path calculations are done in the following two ways :

(a) Forward pass calculations method.

(b) Backward pass calculations method.

10.1 Forward pass calculations method

In this method, caclculation begins from the initial event, proceed through the events in an increasing order of event numbers and end at the final event of the network. At each node (event), the earliest starting and finishing times are calculated for each activity. The method may be summarized as follows:

Step - 1 : Set $E_1 = 0, i = 1$

Step - 2 : Calculate the earliest starting time ES_{ij} for each activity that begins at event i i.e., $ES_{ij} = E_i$ for all activities (i,j) that start at node i

Step - 3 : Calculate the earliest finishing time EF_{ij} of each activity that begins at event i by adding the earliest starting time of the activity with the duration of the activity. Thus $EF_{ij} = ES_{ij} + t_{ij} = E_i + t_{ij}$

Example 1

Construct a network of a project whose activities and their precedence relationships are given below. Then number the events

Activity	A	B	C	D	E	F	G	H	I
Immediate Predecessor	-	A	A	-	D	B,C,E	F	D	G,H

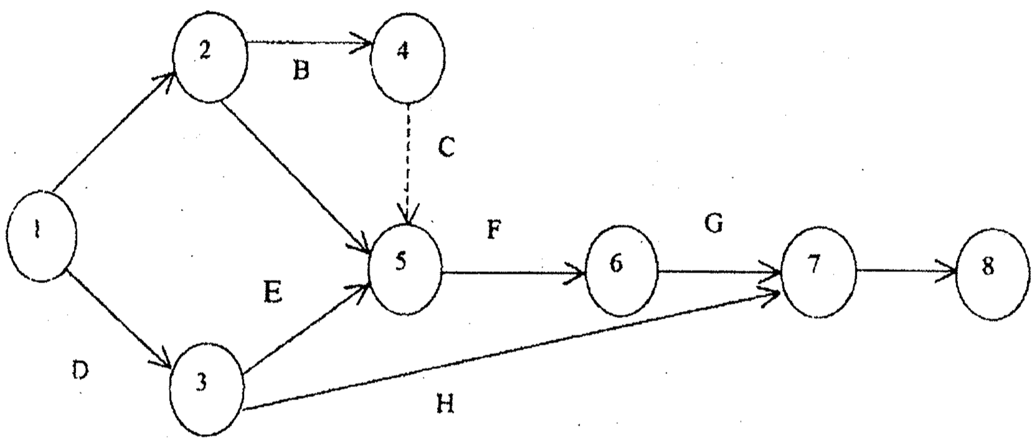


Fig. 8

10. CRITICAL PATH ANALYSIS

Once the network of a project is constructed, the time analysis of the network becomes essential for planning various activities of the project. The main objective of the time analysis is to prepare a planning schedule of the project. The planning schedule should include the following factors :

(i) Total completion time for the project.

(ii) Earliest time when each activity can start.

(iii) Latest time when each activity can be started without delaying the total project.

(iv) Float for each activity i.e., the duration of time by which the completion of an activity can be delayed without delaying the total project completion.

(v) Identification of critical activities and critical path.

(i) Each activity is represented by one and only one arrow.

(ii) Crossing an arrow and curved arrows should be avoided, only straight arrows are to be used.

(iii) Each activity must be identified by its starting and ending node.

(iv) No event can occur until every activity preceding it has been completed.

(v) An event cannot occur twice i.e., there must be no loops.

(vi) An activity succeeding an event can not be started until that event has occurred.

(vii) Events are numbered to identify an activity uniquely. The number of tail event (starting event) should be lower than that of the head (ending) event of an activity.

(viii) Between any pair of nodes (event), there should be one and only one activity. However, more than one activity may emanate from a node or terminate to a node.

(ix) Dummy activities should be introduced if it is extremely necessary.

(x) The network has only one entry point called the starting event and one point of emergence called the end or terminal event.

8. NUMBERING THE EVENTS

After the network is drawn in a logical sequence every event is assigned a number. The number sequence must be such so as to reflect the flow of the network. In numbering the events. Fulkerson's (D. R. Fulkerson) rules are used : These rules are as follows:

(i) Event number should be unique.

(ii) Event numbering should be carried out on a sequential basis from left to right.

(iii) An initial event is one which has all outgoing arrows with no incoming arrow. In any network, there will be one such event. Number it as 1.

(iv) Delete all arrows emerging from event 1. This will create at least one more initial event.

(v) Number these initial events as 2, 3, ... etc.

(vi) Delete all emerging arrows from these numbered events which will create new initial events.

(vii) Repeat steps (v) and (vi) until the last event is obtained which has no arrows emerging from it. Number the last event.

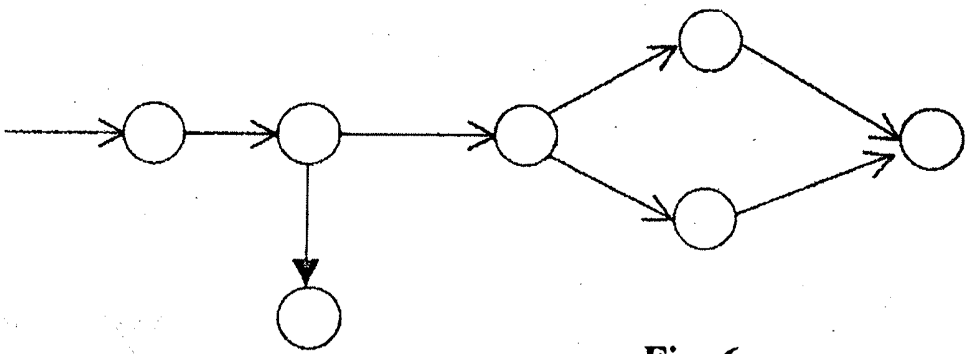


Fig. 6

Dangling

To disconnect an activity before the completion of all the activitis in a network diagram is known as dangling. It should be avoided. In that case, a dummy activity is introduced in order to maintain the continuity of the system (See **Fig. 6**).

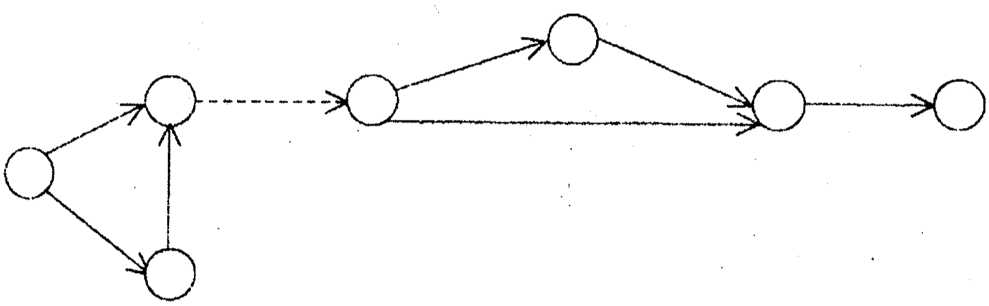


Fig. 7

Redundancy

If a dummy activity is the only activity emanating from an event and which can be eliminated is known as redundancy (see **Fig. 7**).

8. RULES OF NETWORK CONSTRUCTION

For the construction of a network, generally, the following rules are followed :

It is assumed that inside cleaning and outside cleaning can be done concurrently by two assistants of the garage. Activities *B* and *C* represent these cleaning operations. What do activitis *D* and *E* stand for? Their time consumptions are zero but they express the condition that events 3 and 4 must occur before the event 5 can take place. Activities *D* and *E* are called the dummy activities

Network

It is the graphical representation of logically and sequentially connected arrows and nodes representing activities and events of a project Network are also called arrow diagram.

Path

An unbroken chain of actvity arrows connecting the initial the initial event to some other event is called a path.

7. COMMON ERRORS

There are three common errors in a network construction.

Looping (cycling) : In a network diagram looping error is also known as cycling error. Drawing of an endless loop in a network is known as error of looping. A looping network is given in **Fig. 5.**

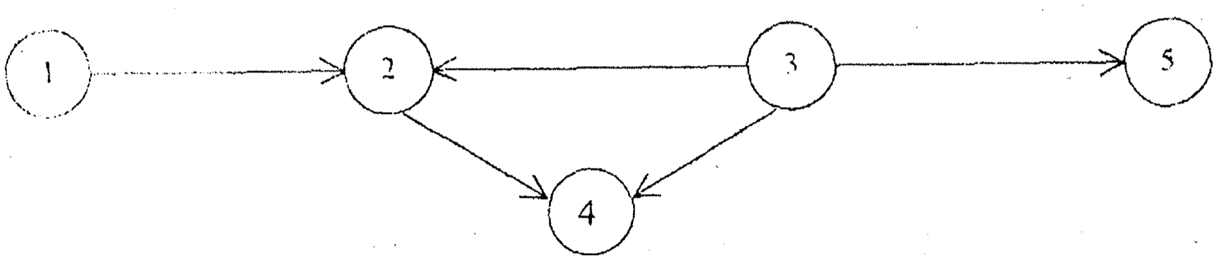


Fig. 5

(i) *Predecessor activity :* An activity which must be completed before one or more other activities start is known as predecessor actvitiy.

(ii) *Successor activity :* An activity which started immediately after the completion of one or more of other activities are completed is known as successor activity.

(iii) *Dummy activity :* In connecting events by activities showing their inter dependencies, very often a situation arises where a certain event *j* can not occur until another event *i* has taken place but, the activity connecting *i* and *j* does not involve any time or expenditure of other resources. In such a case, the activity is called the dummy activity. It is depicted by dotted line in the network diagram.

Let us consider an example of a car taken to a garage for cleaning. Inside as well as outside of the car is to be cleaned before it is taken away from the garage. The events can be put down as follows:

Event 1 : Starting of car from house

Event 2 : Parking of car in garage

Event 3 : Completion of outside cleaning

Event 4 : Completion of inside cleaning

Event 5 : Taking of car from garage

Event 6 : Parking of car in house

The network diagram for the problem is given in **Fig. 4.**

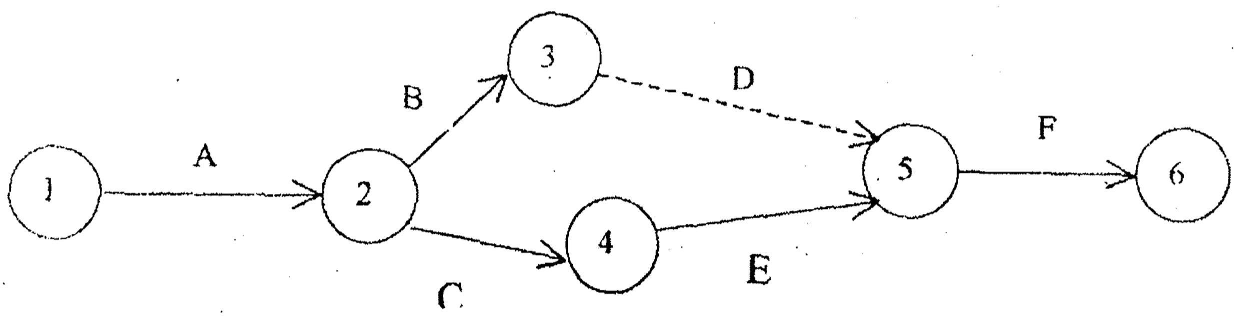


Fig. 4

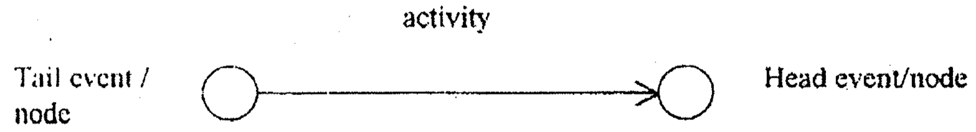


Fig. 1

Merge and burst event

It is necessary for an event to be the ending event of only one activity but can be the ending event of two or more activities. Such event is defined as merge event.

If the event happens to be the beginning event of two or more activities it is defined as a burst event.

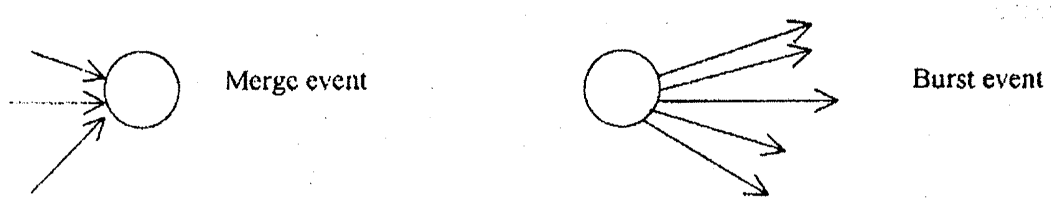


Fig. 2 (a) **Fig. 2 (b)**

Activity

An activity is a task or item of work to be done that consumes time, effort, money or other resources. Activities are represented by arrows.

Activities are identified by the numbers of their starting (tail) event and ending (head) event. Generally, an ordered pair (i, j) represents an activity where events i and j represent the starting and ending of the activity respectively. Activities are also denoted by capital alphabets.

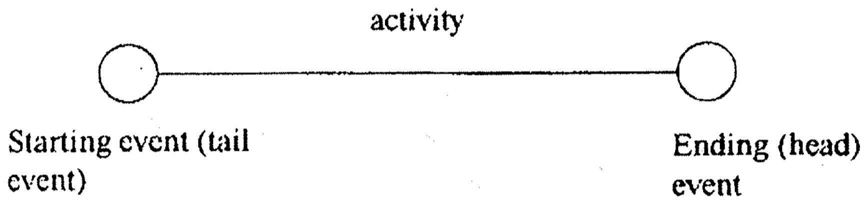


Fig. 3

The activities can further be classifed into different categories :

Hence,

$$\mu_{C_i}^{-1}(\alpha) = C_i - (1-\alpha)P_i, \quad (i=1,3)$$

$$\mu_{C_0}^{-1}(\alpha) = C_0 + (1-\alpha)P_0$$

$$\mu_B^{-1}(\alpha) = B + (1-\alpha)P$$

From (7), we have,

$$L(\alpha, Q, \lambda_1, \lambda_2) = \alpha - \lambda_1\left[\{C_1 - (1-\alpha)P_1\}Q/2 + \{C_3 - (1-\alpha)P_3\}D/Q - C_0 - (1-\alpha)P_0\right]$$
$$-\lambda_2\{AQ - B - (1-\alpha)P\}$$

Now, the Kuhn-Tucker necessary conditions are

$$\frac{\partial L}{\partial \alpha} = 0$$

$$\frac{\partial L}{\partial Q} = 0$$

$$\{C_1 - (1-\alpha)P_1\}Q/2 + \{C_3 - (1-\alpha)P_3\}D/Q - C_0 - (1-\alpha)P_0 = 0$$

$$AQ - B - (1-\alpha)P = 0$$

Solving these equations, the expression for optimal order quantity is

$$Q^* = \{B + (1-\alpha^*)P\}/A$$

where α is a root of

$$P_1P^2(1-\alpha)^3 - (C_1P^2 - 2BPP_1 - 2AP_0P)(1-\alpha)^2 - (2BC_1P - P_1B^2 - 2DA^2P_3 - 2APC_0 - 2AP_0B)$$
$$(1-\alpha) - (C_1B^2 + 2DA^2C_3 - 2AC_0B) = 0$$

Fuzzy goal and storage area represented by parabolic concave membership functions and costs by linear membership functions

In this case, the membership functions $\mu_{C_i}(u)$, $i=1,3$ for fuzzy inventory cost will be the same as defined in the earlier section.

The membership function $\mu_{C_0}(C(Q))$ for fuzzy goal is defined as follows:

$$\mu_{C_0}(C(Q)) = 1 \qquad \text{for } C(Q) < C_0$$

$$= 1 - [\{C(Q) - C_0\} / P_0]^2 \qquad \text{for } C_0 \le C(Q) \le C_0 + P_0$$

$$= 0 \qquad \text{for } C(Q) > C_0 + P_0$$

is graphically represented in the following figure.

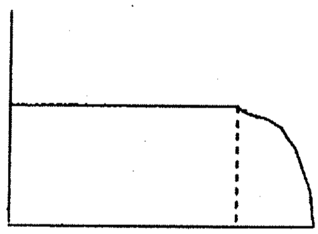


Fig.-12: Parabolic membership function of $\tilde{C}(Q)$

Again, the membership function for storage area is defined as

$$\mu_B(AQ) = 1 \qquad \text{for } AQ < B$$

$$= 1 - [(AQ - B) / P]^2 \qquad \text{for } B \le AQ \le B + P$$

$$= 0 \qquad \text{for } AQ > B + P$$

Proceeding exactly as before, the optimal order quantity is given by

$$Q^* = \left\{B + \sqrt{(1-\alpha^*)}P\right\} / A$$

where α^* is a root of

$$P_1P^2(1-\alpha)^2 + 2BPP_1(1-\alpha)^{\frac{3}{2}} - (C_1P^2 - P_1B^2 - 2A^2DP_3 - 2AP^2)(1-\alpha)$$

$$-(2BPC_1 - 2AC_0P - 2APB)(1-\alpha)^{\frac{1}{2}} - (C_1B^2 + 2A^2DC_3 - 2AC_0B) = 0$$

❑ 112. 7. SELF ASSESSMENT QUESTIONS

1. What is inventory? Give reasons for maintaining the inventory.
2. What are the costs associated with inventory? Discuss the different types of inventory costs.
3. Discuss the different types of demand.
4. Discuss the classification of inventory models.

5. Write short notes on : Ordering/set up cost, holding cost, shortage cost, disposal cost.
6. Explain : Lead time, time horizon, demand, Deterioration.
7. Derive Harris-Wilson's formula.
8. Derive the EOQ formula for the manufacturing model without shortages.
9. In a certain manufacturing situation, the production is instantaneous and the uniform rate of demand is D. Show that the optimal order quantity is

$$Q = \sqrt{\frac{2c_3(c_1 + c_2)}{c_1 c_2}}$$

where C_1, C_2 are the holding and shortage cost per unit per unit time and C_3 is the setup cost per order.

10. The demand rate of a particular item is 12,000 units per year. The ordering cost per order is Rs. 350.00 and the holding cost is Rs. 0.20 per unit per month. If no shortage is allowed and the replenishment is instantaneous, determine (i) the optimal lot size (ii) the optimum scheduling period (iii) minimum total average cost.
11. The annual requirement for a product is 3000 units. The ordering cost is Rs. 100.00 per order. The cost per unit is Rs. 10.00. The carrying cost per unit per year is 30% of the unit cost.
    (a) Find the EOQ.
    (b) If a new EOQ is founded by using the ordering cost as Rs. 80.00, what would be further saving in cost ?
12. The demand rate for an item of a company is 18000 units per year. The company can produce at the rate of 3000 units per month. The set up cost is Rs. 500.00 per order and the holding cost is 0.15 per units per month. Calculate
    (i) Optimum manufacturing quantity, (ii) the time of manufacture, (iii) cycle length, (iv) the optimum annual cost if the cost of an item is Rs. 2.00 per unit.
13. A contractor has to supply 10,000 bearings per day to an automobile manufacturer. He finds that when he starts production run he can produce 25000 bearings per day. The holding cost of a bearing in stock is Rs. 0.02 per year. Setup cost of a production is Rs. 18.00. How frequently should production run be made ?
14. The demand for an item is 18000 units per year. The holding cost is Rs. 1.20 per unit per unit item and the shortage cost is Rs.5.00. The ordering cost is Rs. 400.00. Assuming the replenishment

rate as instantaneous, determine the optimum order quantity along with total minimum cost of the system.

15. Find the optimum order level which minimizes the total expected cost under the following assumptions.
   (i) t is the constant interval between orders.
   (ii) Q is the stock (in discrete units) at the beginning.
   (iii) d is the estimated random instantaneous demand at a discontinuous rate.
   (iv) C_1 and C_2 be the holding and shortage costs per item per t time unit.
   (v) Level time is zero.

16. Let $F(x)$ be the probability density function of the demand x in the prescribed scheduling period T weeks. The demand is assumed to occur with a uniform pattern and the probability distribution is continuous. The unit carrying cost and the shortage cost are respectively C_1 money units and C_2 money units both per unit per week. There is no set up cost for the system. Obtain the expression for optimal control.

17. Let the probability density of demand of a certain item during a week be

$$f(x)=\begin{cases}0.1, & 0\le x\le 10\\ 0\ , & \text{Otherwise}\end{cases}$$

This demand is assumed to occur with a uniform pattern over the week. Let the unit carrying cost of the item in inventory be Rs. 2.00 per week and unit shortage cost be Rs. 8.00 per week. Determine the optimal order level of the inventory and the total minimum cost of the inventory system.

18. The annual demand for a product is 500 units. The cost of storage per unit per year is 10% of the unit cost. The ordering cost is Rs. 180 for each order. The unit cost depends upon the amount ordered. The range of amount ordered and the unit price are as follows:

Range of amount ordered	Unit cost (Rs.)
$1 \le q < 500$	25.00
$500 \le q < 1500$	24.80
$1500 \le q < 3000$	24.60
$3000 \le q$	24.40

Find the optimal order quantity.

19. Show that when determining the optimal inventory level so, which minimizes the total expected cost in case of continuous (non-discrete) quantities, the condition to be satisfied is

$$F(S_0) = \frac{C_2}{C_1 + C_2}$$

where $F(S_0) = \int_0^{S_0} f(r)\,dr$,

$f(r)$ = the probability density function of requirement of r quantity;

C_2 = shortage cost per unit per unit time;

C_1 = holding cost per unit per unit time.

20. An ice-cream company sells one of its types of ice-cream by weight. If the product is not sold on the day it is prepared, it can be sold at a loss of 50 paise per pound. There is, however, an unlimited market for one day old ice-cream. On the other hand, the company makes a profit of Rs. 3.20 on every pound of ice-cream sold on the day it is prepared. Past daily orders form a distribution with

$f(x) = 0.02 - 0.0002x,\ 0 \leq x \leq 100.$

How many pounds of ice-cream should the company prepare every day?

21. The following data describe three inventory items. Determine the economic order quantity for each of the three items to be accommodated with in total available storage area of 25 sq. ft.

Item	Set-up Cost (Rs.)	Demand (Units per day)	Unit holding cost per unit time (Rs.)	Storage area required per unit (sq. ft.)
1	10	2	0.3	1
2	5	4	0.1	1
3	15	4	0.2	1

Find the economic lot size for the inventory model with finite replenishment rate, shortages are allowed but fully backlogged, uniform finite demand and zero lead time so that total average cost is minimum.

22. A shop produces three items in lots. The demand rate for each item is constant and can be assumed to be deterministic. No back orders are to be allowed. The pertinent data for the items is given in the following table:

Item	I	II	III
Carrying cost (Rs.)	20	20	20
Set up cost (Rs.)	50	40	60
Cost per unit (Rs.)	6	7	5
Yearly demand rate (units)	10,000	12,000	7,500

Determine approximately the economic order quantity for three items subject to the condition that the total value of average inventory levels of these items does not exceed Rs. 1,000.00.

23. Find the optimum order quantity for a product for which the price breaks are as follows:

Quantity	Purchasing Cost (per unit)
$0 \le Q_1 < 100$	Rs. 20
$100 \le Q_2 < 200$	Rs. 18
$200 \le Q_3$	Rs. 16

The monthly demand for the product is 400 units. The storage cost is 20% of the unit cost of the product and the cost of ordering is Rs. 25.00 per month.

24. Formulate and solve a single period discrete stochastic inventory model for a single product with instantaneous discrete demand, zero lead time and no replenishment cost. The storage and storage costs are independent of time.

25. A newspaper-boy buys papers for Rs. 1.70 each and sells them for Rs. 2.00 each. He cannot return unsold newspapers. Daily demand has the following distribution:

No. of Customers:	23	24	25	26	27	28	29	30	31	32
Probability :	0.01	0.03	0.06	0.10	0.20	0.25	0.15	0.10	0.05	0.05

If each day's demand is independent of the previous days, how many papers he should order each day?

26. A small shop produces three machine parts 1, 2, 3 in lots. The shop has only 700 sq.mts of storage space. The appropriate data for the three items are presented in the following table:

Item	*1*	*2*	*3*
Demand (unit per year)	2000	5000	10000
Set-up cost (Rs.)	100	200	75
Cost per unit (Rs.)	10	20	5
Floor space required (sq.mt/unit)	0.50	0.60	0.30

The shop uses an inventory carrying charge of 20 percent of average inventory valuation per annum. If no stock-outs are allowed, determine the optimal lot size for each item.

❑ 113.8. References :

[1] Hadley, G. and Whitin, T.M. (1963), Analysis of inventory systems, ***Prentice Hall***, Englewood Cliffs, N.J.

[2] Naddor, E. (1966), Inventory Systems, ***John Wiley***, New York.

[3] Bronson, R and Naadimuthu, G., Theory and problems of Operations Research, Schuam's Outline Series, ***MGH***.

[4] Swarup, K., Gupta, P. K. and Man Mohan, Operations Research, ***Sultan Chand & Sons***.

[5] Sharma, J. K., Operations Research - Theory and Applications, ***Macmillan***.

[6] Gupta, P. K. and Hira, D. S., Operations Research, ***S. Chand & Co****. Ltd.*

[7] Carlsson, C. and Korhonen, P. (1986), A parametric approach to fuzzy linear programming, ***Fuzzy Sets and Systems***, Vol. -29, 303-313.

[8] Trappy, J.F.C., Liu, C. R. and Chang, T. C. (1988), Fuzzy non-linear programming : Theory and Application in manufacturing, ***International Journal of Production Research***, Vol.-26, 957- 985.

[9] Roy, Tapan Kumar, Some inventory models in fuzzy environment, ***Ph. D. thesis***, Vidyasagar University.

---- 0 -----

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X — Special Paper-OR

Module No. - 113
PROJECT MANAGEMENT PERT AND CPM

1. INTRODUCTION

A project is a well defined set of jobs, tasks or activities, all of which must be completed to finish the project. Construction of a highway, power plant, production and marketing of anew product, research and development work are the examples of project. Such projects involve large number of interrelated activities (or tasks) which must be completed in a specified time, in a specified sequence (or order) and require resources such as personnel, money, materials, facilities and/or space. The main objective before starting any project is to schedule the required activities in an efficient manner so as to

(i) Complete it on or before a specified time limit.

(ii) Minimize the total time.

(iii) Minimize the time for a prescribed cost.

(iv) Minimize the cost for a specified time.

(v) Minimize the total cost.

(vi) Minimize the idle resources.

Therefore, before starting any project, it is very much essential to prepare a plan for scheduling and controlling the various activities involved in the project. The techniques of O.R. used for planning scheduling and controlling large and complex projects are very often referred to as network analysis, network planning or network scheduling techniques. In all these techniques, a project is broken down into various activities which are arranged in logical sequence in the form of network. This approach assists managers to visualize a project as a number of tasks which can easily be defined in terms of its duration, cost, starting time. The sequence of activities is also defined. There are two basic planning and control techniques that utilize a network to complete a predetermined project or schedule. These are PERT (Program Evaluation and Review Technique) and CPM (Critical Path Method). PERT network was developed in 1956-58 by a research team of US Navy's Polaris Nuclear Submarine Missile development project. Since 1958, this technique has been used to plan in all most all types of projects. All the same time but independently, CPM was developed jointly jointly by two companies : E.I. Dupont Company and Remington Rand Corporation. Other network techniques were PEP (Performance Evaluation Programme), LCES (Least Cost Estimating and Scheduling), SCANS (Scheduling and Control by Automated Network System).

Structure :

2. OBJECTIVES

The objectives of this unit are to

* discuss the importance of using PERT and CPM techniques for project management.
* show the difference between PERT and CPM network techniques.
* discuss the different phases of any project and various activities need to be done during these phases.
* draw the network diagrams with single and three time estimates of activities involved in a project.
* determine the critical path and floats associated with non-critical activities and events along with total project completion time.
* determine the probability of completing a project within the schedule date.
* find the crash project schedule time and establish a time-cost trade-off for completion of a project.

3. KEYWORDS

Project, planning, scheduling, controlling, event, activity, network, critical path, critical event, looping, dangling, float, slack, pert, time estimates, optimistic time, pessimistic time, most-likely time, time cost trade off, project cost, indirect cost, crashing, cost slope.

4. PHASES OF PROJECT MANAGEMENT

The work involved in a project can be divided into three phases corresponding to the management functions of planning, scheduling and control.

Planning

This phase involves setting the objectives of the project and the assumptions to be made. Also it involves the listing of tasks or jobsthat must be performed to complete a project under consideration. In this paper, men, machines and materials and materials required for the project in addition to the estimates of costs and duration of the various activities of the project are also determined.

Scheduling

This consists of laying the activities according to the precedence order and determining.

(i) The starting and finishing times for each activity.

(ii) The critical path on which the activities require special attention and

(iii) The slack and float for the non critical paths.

Controlling

This phase is exercised after the planning and scheduling which involves the following :

(i) Making periodical progress reports.

(ii) Reviewing the progress.

(iii) Analysiing the status of the project.

(iv) Management decisions regarding updating, crashing and resource allocation, etc.

5. ADVANTAGES OF NETWORK ANALYSIS

The network analysis

(i) shows inter-relationships of all jobs in the project.

(ii) gives a clear picture of relationship controlling the order of performance of varous activities than a typical Barchart.

(iii) helps in communication of ideas. The pictorial approach helps to clarify the verbal instructions.

(iv) provides time schedule containing much more information than other methods like Barchart etc.

(v) identifies jobs which are critical for a project completion data.

(vi) permits an accurate forecast of resource requirement.

(vii) provides a method of resource allocation to meet the limiting condition and to maintain or to minimize the overall costs.

(viii) integrates all elements of a program to whatever detail is desired by the management.

(ix) relates time to costs which allows a money value to be placed on proposed changes.

6. BASIC COMPONENTS

There are two basic components in network. These are

(i) Event/Node

(ii) Activity

Event / Node

A node / event is a particular instant in time showing the end or beginning of one or more activity. It is a point of accomplishment or decision. The starting and end points of an activity are thus described by two events usually known as the tail event and head event respectively. An event is generally represented by a circle, rectangle, hexagon or some other geometric shapes. These geometric shapes are numbered for distinguishing an activity from another one. The occurrence of an event indicates that the work has been accomplished upto that point.

Again, $\mu_{C_0}(C(Q)) = 1$ for $C(Q) < C_0$

$= 1 - (C(Q) - C_0)/P_0$ for $C_0 \le C(Q) \le C_0 + P_0$

$= 0$ for $C(Q) > C_0 + P_0$

Its pictorial representation is given in the following figure.

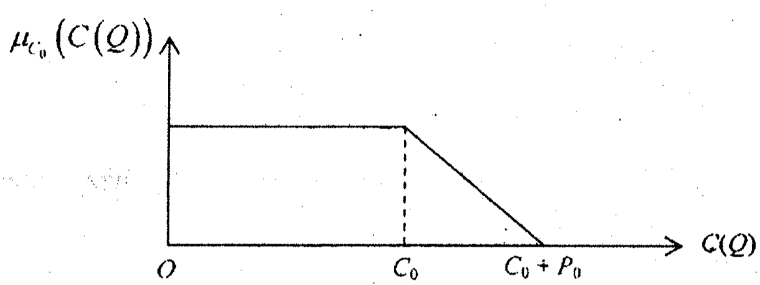


Fig.-10: Linear membership function of $\tilde{C}(Q)$

Again, the linear membership function for storage area is given by

$\mu_B(AQ) = 1$ for $AQ < B$

$= 1 - (AQ - B)/P$ for $B \le AQ \le B + P$

$= 0$ for $AQ > B + P$

This membership function is depicted in the following Figure.

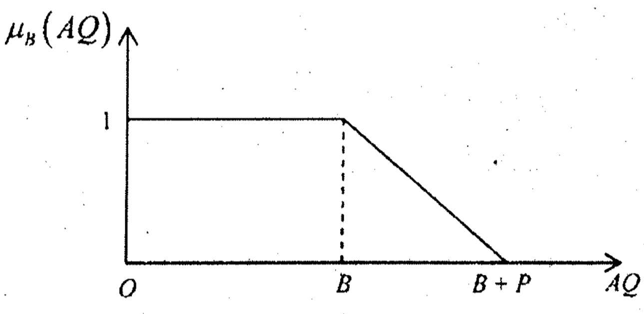


Fig.-11: Linear membership function of AQ

Here P_0, P and P_i's ($i = 1, 3$) are the maximally acceptable violation of the aspiration levels C_0, B and C_i's ($i = 1, 3$). Considering the nature of these parameters, we assume membership function to be non-decreasing for fuzzy inventory costs and non-increasing for fuzzy goal and storage area.

$$\widetilde{Min}\, C(Q) = \widetilde{C}_1 Q/2 + \widetilde{C}_3 D/Q \tag{6}$$

subject to

$$AQ \le \widetilde{B},\ Q > 0$$

So the corresponding fuzzy non-linear programming problem is

Max α

subject to

$$\mu_{C_1}^{-1}(\alpha)Q/2 + \mu_{C_3}^{-1}(\alpha)D/Q \le \mu_{C_0}^{-1}(\alpha)$$

$$AQ \le \mu_B^{-1}(\alpha),\ Q > 0,\ \alpha \in [0,1]$$

where $\mu_{C_3}(x), \mu_{C_1}(x), \mu_{C_0}(x)$ and $\mu_B(x)$ be the membership function for fuzzy ordering cost, holding cost, objective goal and storage area respectively.

Here the Lagrangian function is

$$L(\alpha,Q,\lambda_1,\lambda_2) = \alpha - \lambda_1\{\mu_{C_1}^{-1}(\alpha)Q/2 + \mu_{C_3}^{-1}(\alpha)D/Q - \mu_{C_0}^{-1}(\alpha)\} - \lambda_2\{AQ - \mu_B^{-1}(\alpha)\} \tag{7}$$

where λ_1 and λ_2 are Lagrange multipliers.

Now, we consider three combinations of different types of membership functions to represent the fuzzy goal, costs and warehouse space.

Fuzzy goal, costs and storage area represented by linear membership functions

In this case, $\mu_{C_i}(i = 1,3)$ are given by

$$\mu_{C_i}(u) = 1 \quad \text{for } u > C_i$$

$$= 1-(C_i - u)/P_i \quad \text{for } C_i - P_i \le u \le C_i \ (i = 1, 3)$$

$$= 0 \quad \text{for } u < C_i - P_i$$

Graphically, it is represented in the following Figure.

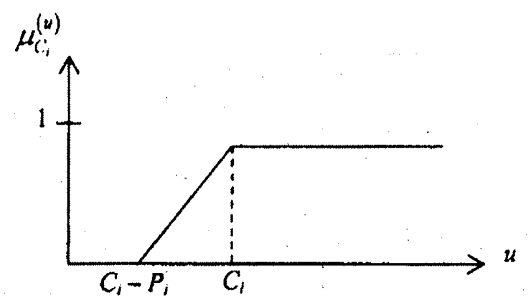


Fig.-9: Linear membership function of C_i

where $\mu_{\tilde{C}_i}(x)=\{\mu_{C_{i1}}(x),\mu_{C_{i2}}(x),\cdots,\mu_{C_{il_i}}(x)\}$ be the membership functions of fuzzy coefficients and $\mu_{b_i}(x)$, $i=1,2,\cdots,m$ be the membership functions of fuzzy objective and fuzzy constraints.

Here the additional variable α is known as the aspiration level.

Therefore, the Lagrangian function $L(\alpha,x,\lambda)$ is given by

$$L(\alpha,x,\lambda)=\alpha-\sum_{i=0}^{m}\lambda_i\{g_i(x,\mu_{\tilde{C}_i}^{-1}(\alpha))-\mu_{b_i}^{-1}(\alpha)\}$$

where $\lambda=(\lambda_0,\lambda_1,\lambda_2,\cdots,\lambda_m)^T$ be the Lagrange multiplier vector.

Now the kuhn-Tucker necessary conditions are

$$\frac{\partial L}{\partial x_j}=0,\quad j=1,2,\cdots,n$$

$$\frac{\partial L}{\partial \alpha}=0$$

$$\lambda_i\{g_i(x,\mu_{\tilde{C}_i}^{-1}(\alpha))-\mu_{b_i}^{-1}(\alpha)\}=0 \tag{4}$$

$$g_i(x,\mu_{\tilde{C}_i}^{-1}(\alpha))-\mu_{b_i}^{-1}(\alpha)\le 0,\quad i=0,1,2,\cdots,m$$

$$\lambda_i\le 0$$

Now, solving the system (4), the optimal solution for the fuzzy non-linear programming problem is obtained.

In a deterministic EOQ model, the problem is to determine the order level $Q(>0)$ which minimizes the average total cost $C(Q)$ i.e.,

$$Min\,C(Q)=C_1Q/2+C_3D/Q \tag{5}$$

subject to

$$AQ\le B,\quad Q>0$$

where C_3 = ordering cost per order, C_1 = holding cost per unit quantity per unit time,

D = demand per unit time, B = maximum available warehouse space (in sq.ft.),

A = the space required by each unit (in sq.unit).

When the above objective goal, costs and available storage area become fuzzy, the said problem is transformed to:

Hence Q^*= 63.5 pounds is the optimum quantity i.e., the company will prepare 63.5 pounds of ice-cream per day.

❍ 6.3 Fuzzy inventory model

First of all, we shall discuss the solution procedure of fuzzy non-linear programming problem. A deterministic or crisp non-linear programming problem may be defined as follows:

Minimize $g_0(x, C_0)$

subject to

$$g_i(x, C_i) \le b_i, \quad i = 1, 2, \cdots, m \tag{1}$$

and $x \ge 0$

where $x = (x_1, x_2, \cdots, x_n)^T$ is a variable vector, f, g_i's are algebraic expressions in x with coefficients $C_0 = (C_{01}, C_{02}, \cdots, C_{0l_0})$ and $C_i = (C_{i1}, C_{i2}, \cdots, C_{il_i})$ respectively.

Introducing fuzziness in the crisp parameters, the system (1) in a fuzzy environment is

Minimize $g_0(x, \widetilde{C_0})$

subject to

$$g_i(x, \widetilde{C_i}) \le \tilde{b}_i, \quad i = 1, 2, \cdots, m \tag{2}$$

and $x \ge 0$

where the wavy bar (~) represents the fuzziness of the parameters.

According to fuzzy set theory, the fuzzy objective, coefficients and constraints are defined by their membership functions which may be either linear or non-linear. According to Bellman and Zadeh (1970) and following the techniques of Carlsson and Kerhonen (1986) and Trappy et. al (1988), the problem (2) is transformed to the new optimization problem as follows:

Maximize α

subject to

$$g_i(x, \mu_{\tilde{C}_i}^{-1}(\alpha)) \le \mu_{\tilde{b}_i}^{-1}(\alpha), \quad i = 0, 1, 2, \cdots, m \tag{3}$$

and $x \ge 0$

Solution : Here C_1 = Re. 1.00, C_2 = Rs. 7.00, C_3 = 0

The required optimum value for Q^* is determined by

$$\int_0^{Q^*} f(r)dr = \frac{C_2}{C_1 + C_2} \qquad (1)$$

Since the distribution is rectangular, the probability density function is given by $f(r) = \frac{1}{b-a}$, $a \le r \le b$. Here a = 4000, b = 5000.

Hence from (1), we have

$$\int_{4000}^{Q^*} \frac{1}{5000-4000} dr = \frac{7}{1+7} \text{ or, } \int_{4000}^{Q^*} \frac{1}{1000} dr = \frac{7}{8}$$

or, $\frac{Q-4000}{1000} = \frac{7}{8}$ or, $Q^* = 4875$

Hence the optimal order quantity is $Q^* = 4875$ units.

Example - 11 :

An ice-cream company sells one of its type of ice-creams by weight. If the product is not sold on the day it is prepared, it can be sold for a loss of Rs. 0.50 per pound. But there is an unlimited market for one day old ice-creams. On the other hand, the company makes a profit of Rs. 3.20 on every pound of ice-creams sold on the day it is prepared. If daily orders form a distribution with $f(x) = 0.02 - 0.0002\,x$, $0 \le x \le 100$, how many pounds of ice-creams should the company prepare every day.

Solution : It is given that C_1 = Rs. 0.50, C_2 = Rs. 3.20

Let Q be the amount of ice-cream prepared every day. The required optimum for Q^* is determined by

$$\int_0^{Q^*} f(x)dx = \frac{C_2}{C_1 + C_2}$$

or, $$\int_0^{Q^*} (0.02 - 0.0002x)dx = \frac{32}{0.5+3.2}$$

or, $\quad 0.0002Q^{*2} + 0.04\,Q^* + 1.730 = 0$

On solving, Q^* = 136.7 and 63.5.

But Q^* = 136.7 is not possible as $0 \le x \le 100$.

Solution :

Let Q be the number of news papers ordered per day and r be the demand for it i.e., the number that are actually sold per day.

In this problem, the holding cost per paper per day is C_1 = Rs. 1.40 and the shortage cost per paper per day is C_2 = Rs. (2.00-1.40) = Rs. 0.60.

Now to obtain the optimal solution, we shall find out the cumulative probability of daily demand of newspapers.

Calculations are given in the following table.

r	23	24	25	26	27	28	29	30	31	32
$p(r)$	0.01	0.03	0.06	0.10	0.20	0.25	0.15	0.10	0.05	0.05
$\sum_{r=23}^{Q} p(r)$	0.01	0.04	0.10	0.20	0.40	0.65	0.80	0.90	0.95	1.00

The required optimum value for Q^* is determined by

$$\sum_{r=23}^{Q^*-1} p(r) < \frac{C_2}{C_1+C_2} < \sum_{r=23}^{Q^*} p(r)$$

$$\text{or, } \sum_{r=23}^{Q^*-1} p(r) < 0.3 < \sum_{r=23}^{Q^*} p(r) \quad \left[\because \frac{C_2}{C_1+C_2} = \frac{0.60}{1.40+0.60} = \frac{3}{10} = 0.3\right]$$

From the table, we have

$$\sum_{r=23}^{27} p(r) = 0.40 > 0.3 \text{ and } \sum_{r=23}^{26} p(r) = 0.20 < 0.3$$

$$\therefore \sum_{r=23}^{26} p(r) < 0.3 < \sum_{r=23}^{27} p(r)$$

Hence Q^*=27 i.e., optimum number of newspaper to be ordered is 27.

Example-10 :

The demand for a certain product has a rectangular distribution between 4000 and 5000. Find the optimal order quantity, if the storage cost is Re. 1.00 per unit and shortage cost is Rs. 7.00 per unit.

$$\therefore \frac{dTEC(Q)}{dQ} = C_1 \int_0^Q f(r)dr - C_2 \int_Q^\infty f(r)dr$$ [Using Leibniz's rule for differentiation

under the integral sign]

$$= C_1 \int_0^Q f(r)dr - C_2 \left[\int_0^\infty f(r)dr - \int_0^Q f(r)dr \right]$$

$$= -C_2 + (C_1 + C_2) \int_0^Q f(r)dr$$

For optimum value of Q, we must have

$$\frac{dTEC(Q)}{dQ} = 0$$

i.e., $(C_1 + C_2) \int_0^Q f(r)dr = C_2$

or, $$\int_0^Q f(r)dr = \frac{C_2}{C_1 + C_2} \qquad (7)$$

Moreover, it can be proved that

$$\frac{d^2TEC(Q)}{dQ^2} = (C_1 + C_2) f(Q) > 0$$

Therefore the optimum value of Q i.e., Q^* is given by (7).

Hence the optimal ordering policy with x as on hand amount before placing an order is as follows :

If $Q^* > x$, then order the amount Q^*-x and if $Q^* < x$, then do not order the quantity.

Example-9:

A newspaper boy buys papers for Rs. 1.40 each and sells them for 2.00. He can not return the unsold news papers. Daily demand has the following distribution.

No. of Customers	23	24	25	26	27	28	29	30	31	32
Probability	0.01	0.03	0.06	0.10	0.20	0.25	0.15	0.10	0.05	0.05

In each day's demand is independent of previous day's demand, how many papers should he ordered each day ?

$$\text{or, } TEC(Q-1)-TEC(Q) = -(C_1+C_2)\sum_{r=0}^{Q-1} p(r)+C_2 \tag{3}$$

For optimal Q^*, we must have

$$TEC(Q^*+1)-TEC(Q^*)>0$$

From (2), we have

$$(C_1+C_2)\sum_{r=0}^{Q^*} p(r)-C_2>0$$

$$\text{or, } \sum_{r=0}^{Q^*} p(r) > \frac{C_2}{C_1+C_2} \tag{4}$$

Similarly for optimal Q^*, TEC(Q^*-1) - TEC(Q^*) > 0

From (3), we have

$$-(C_1+C_2)\sum_{r=0}^{Q^*-1} p(r)+C_2>0$$

$$\text{or, } \sum_{r=0}^{Q^*-1} p(r) < \frac{C_2}{C_1+C_2} \tag{5}$$

Thus combining (4) and (5), we have

$$\sum_{r=0}^{Q^*-1} p(r) < \frac{C_2}{C_1+C_2} < \sum_{r=0}^{Q^*} p(r) \tag{6}$$

$$\text{or, } p(r<Q^*-1) < \frac{C_2}{C_1+C_2} < p(r \le Q^*)$$

where $p(r \le Q^*)$ represents the probability for $r < Q^*$

Continuous Case : *Where r is a continuous variable*

Let x be the on hand amount before placing the order. Also, let the demand r be a continuous variable with the probability density function $f(r)$, then proceeding as before, the total expected cost for this model is

$$TEC(Q) = C_1\int_0^Q (Q-r)f(r)dr + C_2\int_Q^\infty (r-Q)f(r)dr + C_p(Q-x)$$

Case-1 : $r \le Q$

Case-2 : $r > Q$

The corresponding inventory situations are shown in Fig.-7 and Fig.-8.

Discrete case : ***When r is discrete***

Let r be the estimated demand at an instantaneous rate with probabilities $p(r)$. Then there is only holding cost and no shortage cost.
Here the holding cost is $C_1(Q-r)$.

In the second case, the demand r is filled up at the beginning of the period. There is only shortage cost, no holding cost. Therefore, the shortage cost in this case is $C_2(r-Q)$.

Let x be the amount on hand before placing an order.
Therefore, the total expected cost for this model is

$$TEC(Q) = C_1\sum_{r=0}^{Q}(Q-r)p(r) + C_2\sum_{r=Q+1}^{\infty}(r-Q)p(r) \tag{1}$$

Our problem is now to find Q, so that $TEC(Q)$ is minimum. Let an amount Q+1 instead of Q be ordered. Than the total expected cost given in (1) reduces to

$$TEC(Q+1) = C_1\sum_{r=0}^{Q}(Q+1-r)p(r) + C_2\sum_{r=Q+2}^{\infty}(r-Q-1)p(r)$$

on simplification, we have

$$TEC(Q+1) = C_1\sum_{r=0}^{Q}(Q-r)p(r) + C_2\sum_{r=Q+1}^{\infty}(r-Q)p(r) + C_1\sum_{r=0}^{Q}p(r) - C_2\sum_{r=Q+1}^{\infty}p(r) + C(Q-x)$$

$$= TEC(Q) + (C_1 + C_2)\sum_{r=0}^{Q}p(r) - C_2 \quad \left[\text{Since} \sum_{r=Q+1}^{\infty}p(r) = 1 - \sum_{r=0}^{Q}p(r)\right]$$

$$\therefore TEC(Q+1) - TEC(Q) = (C_1 + C_2)\sum_{r=0}^{Q}p(r) - C_2 \tag{2}$$

Similarly, when an amount Q-1 instead of Q is ordered, then we have

$$TEC(Q-1) = TEC(Q) - (C_1 + C_2)\sum_{r=0}^{Q-1}p(r) + C_2$$

(i) T is the constant interval between orders (T may also be considered as unity e.g., daily, weekly, monthly etc.)

(ii) Q is the stock level at the beginning of each period T.

(iii) Lead time is zero.

(iv) The holding cost, C_1 per unit quantity per unit time, the shortage cost, C_2 per unit quantity per unit time are known and constant.

(v) r is the demand at each interval T.

Solution : In this model it is assumed that the total demand is filled up at the beginning of the period.

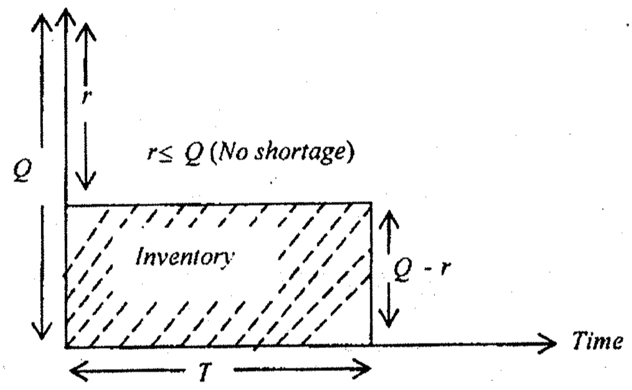


Fig. -7

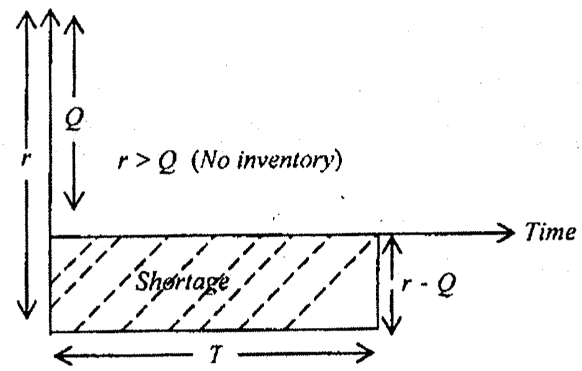


Fig. -8

Thus depending on the amount r demanded, the inventory position just after the demand occurs may be either positive (surplus) or negative (shortage) i.e., there are two cases :

Solution :

Q	r	$p(r)$	$\frac{p(r)}{r}$	$\sum_{r=Q+1}^{\infty}\frac{p(r)}{r}$	$Q+\frac{1}{2}$	$(Q+\frac{1}{2})\sum_{r=Q+1}^{\infty}\frac{p(r)}{r}$	$\sum_{r=0}^{Q}p(r)$	$L(Q)=(Q+\frac{1}{2})\sum_{r=Q+1}^{\infty}\frac{p(r)}{r}+\sum_{r=0}^{Q}p(r)$
0	0	0.40	∞	0.3878	0.5	0.1939	0.40	0.5937
1	1	0.24	0.2400	0.1478	1.5	0.2217	0.64	0.8617
2	2	0.20	0.1000	0.0478	2.5	0.1195	0.84	0.9595
3	3	0.10	0.0333	0.0145	3.5	0.0575	0.94	0.9907
4	4	0.05	0.0125	0.0020	4.5	0.0090	0.99	0.9990
5	5	0.01	0.0020	0.0000	5.5	0.0000	1.0	1.0000
6 *or more*	6 *or more*	0.00	0.0000	0.0000	6.5	0.0000	1.0	1.0000

Here C_1 = Rs. 100.00, C_2 = Rs. 1000.00

$$\therefore \frac{C_2}{C_1+C_2}=\frac{1000}{100+1000}=\frac{10}{11}=0.9090$$

Now for the optimal value of Q (say, Q^*), we must have

$$L(Q^*-1)<\frac{C_2}{C_1+C_2}<L(Q^*)$$

where $L(Q)=\sum_{r=0}^{Q}p(r)+(Q+\frac{1}{2})\sum_{r=Q+1}^{\infty}\frac{p(r)}{r}$

From the table, it is clear that for $Q=2$, the above inequality satisfied i.e., $L(1)<0.9090<L(2)$ or, $L(2\text{-}1)<0.9090<L(2)$.

Hence $Q^*=2$ i.e., optimum stock level of truck is 2.

❍ 6.2.2 Single period inventory model with instantaneous demand (No replenishment cost model)

In this model, we have to find the optimum order quantity which minimizes the total expected cost under the following assumptions :

After simplification, we have

$$\frac{dTEC(Q)}{dQ} = (C_1 + C_2)\int_0^Q f(r)dr + (C_1 + C_2)\int_Q^\infty \frac{Qf(r)}{r}dr - C_2 \tag{8}$$

The necessary condition for TEC (Q) to be optimum is $\frac{dTEC(Q)}{dQ} = 0 \;\; for \;\; Q = Q^*$

i.e., $(C_1 + C_2)\int_0^{Q^*} f(r)dr + (C_1 + C_2)\int_{Q^*}^\infty \frac{Q^* f(r)}{r} - C_2 = 0$

or, $$\int_0^{Q^*} f(r)dr + \int_{Q^*}^\infty \frac{Q^* f(r)}{r} = \frac{C_2}{C_1 + C_2} \tag{9}$$

Again from (8),

$$\frac{d^2TEC(Q)}{dQ^2} = (C_1 + C_2)f(Q) + (C_1 + C_2)\int_Q^\infty \frac{f(r)}{r}dr - (C_1 + C_2)f(Q)$$

After simplification, we have

$$\frac{d^2TEC(Q)}{dQ^2} = (C_1 + C_2)\int_Q^\infty \frac{f(r)}{r}dr \;\; = + \text{ve quantity}$$

Hence the equation (9) gives the optimum value of Q for minimum expected average cost.

Example - 8 :

A contractor of second hand motor trucks uses to maintain a stock of trucks every month. The demand of the trucks occurs at a relatively constant rate but not in a constant size. The demand follows the following probability distributions :

Demand (r)	0	1	2	3	4	5	6 or more
Probability [$p(r)$]	0.40	0.24	0.20	0.10	0.05	0.01	0.0

The holding cost of an old truck in stock for one month is Rs. 100.00 and the penalty for a truck is not supplied on the demand, is Rs. 1000.00. Determine the optimal size of the stock for the contractor.

Using the relation (7), we find the range of optimum value of Q. In these cases, Q^* need not be unique. If $\frac{C_2}{C_1+C_2}=L(Q^*)$ then both Q^* and Q^*+1 be the optimal values. Similarly, if $\frac{C_2}{C_1+C_2}=L(Q^*-1)$ then both Q^* and Q^*-1 be the optimal values.

Continuous case : ***When r is a continuous random variable***

When the uncertain demand is estimated as a continuous random variable, the cost expressions of inventory, holding and shortage costs involve integrals instead of summation signs.

Let *f(r)* be the known probability density function for demand *r*. The discrete point probabilities *p(r)* are replaced by the probability differential *f(r)dr* for small interval, say, $\left(r-\frac{dr}{2}, r+\frac{dr}{2}\right)$. In this case, we have

$$\int_0^\infty f(r)dr=1 \text{ and } f(r)\geq 0.$$

Let x be the amount on hand before placing an order.

Case -1 : ***When*** $r \leq Q$

Proceeding as before for $r \leq Q$, the holding cost is $C_1(Q-\frac{r}{2})t$ and there is no shortage cost.

Case -2 : ***When*** $r > Q$

Proceeding as before for r > Q, the holding cost is $\frac{C_1Q^2t}{2r}$ and the shortage cost is $C_2\frac{(r-Q)^2t}{2r}$.

Proceeding as before, the total expected cost per unit time is given by

$$TEC(Q)=\int_0^Q C_1(Q-\frac{r}{2})f(r)dr+\int_Q^\infty [C_1\frac{Q^2}{2r}+C_2\frac{(r-Q)^2}{2r}]f(r)dr$$

$$\therefore \frac{dTEC(Q)}{dQ}=C_1\int_0^Q f(r)dr+C_1[(Q-\frac{r}{2})f(r)\frac{dr}{dQ}]_0^Q$$

$$+\int_Q^\infty [\frac{C_1}{2r}.2Q-\frac{C_2}{2r}2(r-Q)]f(r)dr+[\{\frac{c_1Q^2}{2r}+\frac{C_2(r-Q)^2}{2r}\}f(r)\frac{dr}{dQ}]_Q^\infty$$

$$= C_2 \sum_{r=Q+1}^{\infty} \frac{(r-Q)^2}{2r} p(r) - C_2 \sum_{r=Q+1}^{\infty} p(r) + C_2 \sum_{r=Q+1}^{\infty} \frac{Q}{r} p(r) + \frac{C_2}{2} \sum_{r=Q+1}^{\infty} \frac{p(r)}{r}$$

Substituting these values in $TEC\ (Q+1)$ and then simplifying, we get

$$TEC(Q+1) = TEC(Q) + (C_1 + C_2)\left[\sum_{r=0}^{Q} p(r) + (Q+\frac{1}{2}) \sum_{r=Q+1}^{\infty} \frac{p(r)}{r}\right] - C_2$$

If we put $\sum_{r=0}^{Q} p(r) + (Q+\frac{1}{2}) \sum_{r=Q+1}^{\infty} \frac{p(r)}{r} = L(Q)$ (2)

then $TEC(Q+1) = TEC(Q) + (C_1 + C_2)\, L(Q) - C_2$

(3)

Similarly, putting Q-1 in place of Q in (1), we have

$TEC(Q-1) = TEC(Q) - (C_1 + C_2)L(Q-1) + C_2$ (4)

For optimal Q, we must have

$TEC(Q+1) - TEC(Q) > 0$

i.e., $(C_1 + C_2)\, L\,(Q^*) - C_2 > 0$ [From (3)]

or, $L(Q^*) > \frac{C_2}{C_1 + C_2}$ (5)

Again, for optimal Q, we have

$TEC(Q-1) - TEC(Q) > 0$

or, $-(C_1 + C_2)\, L\,(Q^*-1) + C_2 - C_p > 0$

or, $L(Q^* - 1) < \frac{C_2}{C_1 + C_2}$ (6)

Combining (5) and (6) for optimal value of Q^*, we have

$$L(Q^* - 1) < \frac{C_2}{C_1 + C_2} < L(Q^*) \quad (7)$$

where $L(Q) = \sum_{r=0}^{Q} p(r) + \left(Q + \frac{1}{2}\right) \sum_{r=Q+1}^{\infty} \frac{p(r)}{r}$

Hence the expected cost in this case ($r > Q$) is

$$\sum_{r=Q+1}^{\infty}\left[\frac{1}{2}C_1 Q t_1 p(r) + \frac{1}{2}(r-Q)C_2 t_2 p(r)\right]$$

$$= \sum_{r=Q+1}^{\infty}\left[C_1\frac{Q^2}{2r}Tp(r) + C_2\frac{(r-Q)^2}{2r}Tp(r)\right]$$

Therefore the average expected cost is given by

$$TEC(Q) = C_1\sum_{r=0}^{Q}(Q-\frac{r}{2})p(r) + C_1\sum_{r=Q+1}^{\infty}\frac{Q^2}{2r}p(r) + C_2\sum_{r=Q+1}^{\infty}\frac{(r-Q)^2}{2r}p(r) \qquad (1)$$

The problem is now to find Q, so as to minimize TEC(Q), Let an amount $Q + 1$ instead of Q be produced. Then the average total expected cost is

$$TEC(Q+1) = C_1\sum_{r=0}^{Q+1}(Q+1-\frac{r}{2})p(r) + C_1\sum_{r=Q+2}^{\infty}\frac{(Q+1)^2}{2r}p(r) + C_2\sum_{r=Q+1}^{\infty}\frac{(r-Q-1)^2}{2r}p(r)$$

But $C_1\sum_{r=0}^{Q+1}(Q+1-\frac{r}{2})p(r) = C_1\sum_{r=0}^{Q}(Q+1-\frac{r}{2})p(r)$

$$+C_1(Q+1-\frac{Q+1}{2})p(Q+1)$$

$$= C_1\sum_{r=0}^{Q}(Q-\frac{r}{2})p(r) + C_1\sum_{r=0}^{Q}p(r) + C_1\frac{Q+1}{2}p(Q+1)$$

Again, $C_1\sum_{r=Q+2}^{\infty}\frac{(Q+1)^2}{2r}p(r) = C_1\sum_{r=Q+1}^{\infty}\frac{(Q+1)^2}{2r}p(r) - C_1\frac{(Q+1)^2}{2(Q+1)}p(Q+1)$

$$= C_1\sum_{r=Q+1}^{\infty}\frac{Q^2+2Q+1}{2r}p(r) - C_1\frac{Q+1}{2}p(Q+1)$$

$$= C_1\sum_{r=Q+1}^{\infty}\frac{Q^2}{2r}p(r) + C_1\sum_{r=Q+1}^{\infty}\frac{Q}{r}p(r) + \frac{C_1}{2}\sum_{r=Q+1}^{\infty}\frac{p(r)}{r} - \frac{C_1}{2}(Q+1)p(Q+1)$$

Similarly, $C_2\sum_{r=Q+2}^{\infty}\frac{(r-Q-1)^2}{2r}p(r)$

$$= C_2\sum_{r=Q+1}^{\infty}\frac{(r-Q-1)^2}{2r}p(r) - 0$$

In the first case, $r \leq Q$ as shown in Fig. -5, no shortage occurs. In the second case, $r > Q$ as shown in Fig. -6, both the costs are involved.

Discrete case : when r is a discrete random variable

Let the demand for D units be estimated at a discontinuous rate with probability $p(r)$, $r = 1, 2, \ldots, n, \ldots$ That is, we may expect the demand for one unit with probability $p(1)$, 2 units with probability $p(2)$ and so on. Since all the possibilities are to be taken care of, we must have $\sum_{r=1}^{\infty} p(r) = 1$ and $p(r) \geq 0$.

We also assume that r be only non-negative integers.

Case -1 : In this case, $r \leq Q$. So, there is no shortage and the total inventory is represented by the total area OAMB $= \frac{1}{2}(Q + Q - r)\,T = (Q - r/2)\,T$.

Hence the holding cost for the time period T is $C_1\,(Q - r/2)\,T$. This is the holding cost when r ($\leq Q$) units be demand rate in one period. But, the probability of the demand of r units is $p(r)$. Hence the expected values of this cost is $C_1\,(Q - r/2)\,T\,p\,(r)$.

Now, r can have only values less than Q. Hence the total expected cost where $r < Q$ is equal to

$$\sum_{r=0}^{Q} C_1\left(Q - \frac{r}{2}\right)T p(r)$$

Case -2 : In this case, $r > Q$, both the holding and shortage costs are involved.

Here, the holding cost is $\frac{1}{2}C_1 Q t_1$ and the shortage cost is $\frac{1}{2}C_2(r - Q)t_2$ where t_1 and t_2 represent the no-shortage and shortage cases and $t_1 + t_2 = T$.

Now, from the similar triangles OBC and ACM in Fig. -6, we have

$$\frac{t_1}{Q} = \frac{t_2}{r - Q}$$

or, $\dfrac{t_1}{Q} = \dfrac{t_2}{r - Q} = \dfrac{t_1 + t_2}{r}$ or, $\dfrac{t_1}{Q} = \dfrac{t_2}{r - Q} = \dfrac{T}{r}$

or, $t_1 = \dfrac{QT}{r}$ and $t_2 = \dfrac{(r - Q)T}{r}$

(vi) The holding cost, C_1 per unit quantity per unit time and the shortage cost, C_2 per unit quantity per unit time are known and constant.

Let x be the amount on hand before an order is placed.

Also, let Q be the level of inventory in the beginning of each period and r units is the demand per time period. Depending on the amount D, two cases may arise :

Case - 1 : $r \leq Q$

Case - 2 : $r > Q$

In both cases, the inventory situation is shown in Fig.-5 and Fig.-6 respectively.

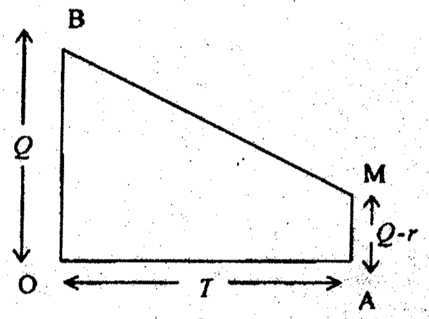


Fig. -5

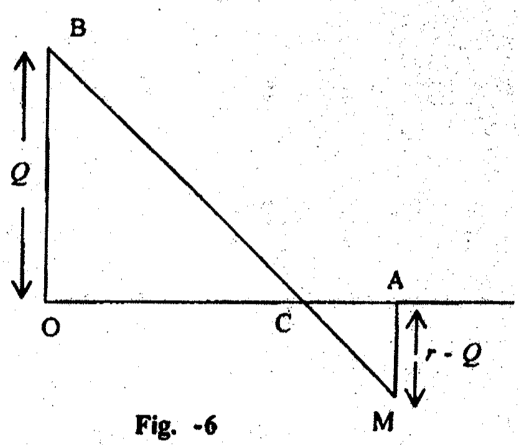


Fig. -6

Since, $0 < Q_1^* < 100$, so we have to compute $C(Q_1^*)$, $C(100)$, $C(200)$.

Now, $C(Q_1^*) = C'(Q_1^*) = p_1 D + \frac{1}{2} C_1 Q_1^* + \frac{C_3 D}{Q_1^*} = 20 \times 400 + \frac{1}{2} \times 0.2 \times 20 \times 71 + \frac{25 \times 400}{71}$

$$= \text{Rs. } 8282.85$$

$$C(100) = C''(100) = p_2 D + \frac{1}{2} C_1 \times 100 + \frac{C_3 D}{100} = 18 \times 400 + \frac{1}{2} \times 0.2 \times 18 \times 100 + \frac{25 \times 400}{100}$$

$$= \text{Rs. } 7480.00$$

and $$C(200) = C'''(200) = p_3 D + \frac{1}{2} C_1 \times 200 + \frac{C_3 D}{200}$$

$$= 16 \times 400 + \frac{1}{2} \times 0.2 \times 16 \times 200 + \frac{25 \times 400}{200}$$

$$= \text{Rs. } 6770.00$$

Since $C(200) < C(100) < C(Q_1^*)$, then the optimal order quantity is 200 units i.e., $Q^* = 200$.

❍ 6.2 Probabilistic Inventory model

Now we consider the situations when the demand is not known exactly but the probability distribution of demand is some how known. The control variable in such cases is assumed to be either the scheduling period or the order level or both. The optimum order levels will thus be derived by minimizing the total expected cost rather than the actual cost involved.

❍ 6.2.1 Single period model with continuous demand (No replenishment cost model)

In this model, we have to find the optimum order quantity so as to minimize the total expected cost with the following assumptions :

(i) Scheduling period T is fixed and known. Hence it is a prescribed constant, so we do not include the set-up cost in our derivation as it is a constant.

(ii) Production is instantaneous.

(iii) Lead time is zero.

(iv) The demand is uniformly distributed over the period.

(v) Shortages are allowed and fully backlogged.

Example - 7 :

Find the optimum order quantity for a product for which the price breaks are as follows :

Range of Quantity to be purchased	Purchase cost per unit
$0 < Q < 100$	Rs. 20.00
$100 \le Q < 200$	Rs. 18.00
$200 \le Q$	Rs. 16.00

The monthly demand for the product is 400 units. The storage cost is 20% of the unit cost of the product and the cost of ordering is Rs 25.00 per order.

Solution :

Here $D = 400$ units/month, $C_3 =$ Rs. 25.00 per order

$C_1 = 20\%$ of purchase cost per unit $= 0.2$ time of purchase cost per unit

Let $Q_3^* = \sqrt{\dfrac{2C_3D}{C_1}}$ for $Q \ge 200$

$$= \sqrt{\frac{2\times 25\times 400}{0.2\times 16}} = 79$$

Since $Q_3^* < 200$, Q_3^* is not the optimum order quantity. Therefore we have to proceed to calculate Q_2^*.

Now $Q_2^* = \sqrt{\dfrac{2C_3D}{C_1}}$ for $100 \le Q < 200$

$$= \sqrt{\frac{2\times 25\times 400}{0.2\times 18}} = 75$$

Again, since $Q_2^* < 100$, therefore Q_2^* is not optimum order quantity.

Now we have to proceed to calculate Q_1^*

$$Q_1^* = \sqrt{\frac{2C_3D}{C_1}} \text{ for } 0 < Q < 100$$

$$= \sqrt{\frac{2\times 25\times 400}{0.2\times 20}} = 71$$

Step - 2 : Evaluate Q^* by the formula $Q^* = \sqrt{\frac{2C_3D}{C_1}}$ for the case $Q < b$ and evaluate $C'(Q^*)$ and $C''(b)$

(i) If $C'(Q^*) < C''(b)$, then Q^* is the optimum lot-size.

(ii) Otherwise, b is the optimal lot-size.

❍ 6.1.6.2 Purchase inventory model with two price breaks

Range of quantity to be purchased	unit purchase cost
$0 < Q < b_1$	p_1
$b_1 \le Q < b_2$	p_2
$b_2 \le Q$	p_3

The procedure used involving one price break is extended to the case with two price breaks.

Working rule :

Step-1 : Compute Q^* for $Q \ge b_2$ (say $Q_3{}^*$). If $Q_3{}^* \ge b_2$ then the optimal lot-size is $Q_3{}^*$, otherwise go to step-2.

Step-2 : Compute Q^* for $b_1 \le Q^* < b_2$ (say $Q_2{}^*$). since $Q_3{}^* < b_2$ then $Q_2{}^*$ is also less than b_2. In this case there are two possibilities i.e., either $Q_2{}^* \ge b_1$ or $Q_2{}^* < b_1$. If $Q_2{}^* \ge b_1$ then compare the cost $C(Q_2{}^*)$ and $C(b_2)$ to obtain the optimum lot-size. The quantity with lower cost will naturally be the optimum one. If $Q_2{}^* < b_1$, then go to Step-3.

Step-3 : If $Q_2{}^* < b_1$, then compute $Q_1{}^*$ for the case $0 < Q < b_1$ and compare the cost $C(Q_1{}^*)$, $C(b_1)$ and $C(b_2)$ to determine the optimal lot-size. The quantity with lower cost will naturally be the optimum one.

$$\left.\frac{dC(Q)}{dq}\right|_{Q=Q^*}=0$$

which implies $Q^* = \sqrt{\dfrac{2C_3D}{C_1}}$ (1)

Now we consider the case in which $Q^* \geq$ b and $Q^* <$ b.

(i) If Q^* [given by (1)] > b then the optimal lot-size Q^* is obtained by (1) and in this case, the minimum total average cost is given by

$$C_{\min}(Q^*) = \frac{C_3D}{\sqrt{\dfrac{2C_3D}{C_1}}} + p_2D + \frac{1}{2}C_1\sqrt{\frac{2C_3D}{C_1}}$$

$$= p_2D + \sqrt{2C_1C_3D}$$

(ii) If $Q^* <$ b then there may arise two cases as follows :

Case -1 : $C''(b) < C'(Q^*)$ *for* $Q^* < b$

Case -2 : $C''(b) > C'(Q^*)$ *for* $Q^* < b$

Now, $C'(Q^*) = p_1D + \sqrt{2C_1C_3D}$

and $C''(b) = p_2D + \dfrac{C_3D}{b} + \dfrac{1}{2}C_1b$

Hence, if $C''(b) > C'(Q^*)$ then Q^* given by (1) is the optimum order quantity. Otherwise, if $C''(b) < C'(Q^*)$, then b is the optimum order quantity.

Working rule :

Step - I : Compute Q^* by the formula $Q^* = \sqrt{\dfrac{2C_3D}{C_1}}$ for the case $Q \geq b$ and then compare this Q^* with the value of b.

(i) If $Q^* \geq b$ then the optimum lot-size is Q^*.

(ii) If $Q^* < b$, then go to step-2.

❍ 6.1.6.1 Purchasing inventory model with single price break

Let D be the demand rate, C_1, be the holding cost per unit quantity per unit time, C_3, the fixed ordering cost per order. Also, let p_1 be the purchasing cost per unit quantity if the ordered quantity is less than b and p_2 $(p_2 < p_1)$ be the purchasing cost per unit quantity if the ordered quantity is greater or equal to b quantities, i.e.,

Range of quantity to be purchased	unit purchase cost
$0 < Q < b$	p_1
$b \le Q$	p_2

Hence the total average cost $C(Q)$ is given by

C(Q) = <ordering cost> + <purchasing cost> + <holding cost>

i.e.,

$$C(Q) = \begin{cases} C'(Q) & \text{for} \quad 0 < Q < b \\ C''(Q) & \text{for} \quad Q > b \end{cases}$$

where $C'(Q) = C_3 \dfrac{D}{Q} + p_1 D + \dfrac{1}{2} C_1 Q$

and $C''(Q) = C_3 \dfrac{D}{Q} + p_2 D + \dfrac{1}{2} C_1 Q$

Thus $C(Q)$ has a discontinuity at $Q = b$ and it may be shown that the minimum value of $C(Q)$ occurs either where $\dfrac{dC(Q)}{dQ} = 0$ or at the point of discontinuity.

We have $\dfrac{dC(Q)}{dQ} = -\dfrac{C_3 D}{Q^2} + \dfrac{1}{2} C_1$ except at $Q = b$ where it is not defined. Thus the optimal value of Q is given by

❍ 6.1.6 Inventory models with price breaks

In the earlier discussion, we have assumed that the unit production cost or unit purchase cost is constant. So, we need not consider this cost in the analysis. However, in the real world, it is not always true that the unit cost of an item is independent of the quantity procured or produced. Again, discounts are offered by the supplier or wholesaler or manufacturer for the purchase of large quantities. Such discounts are referred to as quantity discounts or price breaks.

In this section, we shall consider a class of inventory in which cost is a variable. When items are purchased in bulk, some discount price is usually offered by the supplier.

Let us assume that the unit purchase cost of an item is p_j when the purchased quantity lies between b_j-1 and b_j (j =1, 2, ..., m). Explicitly, we have

Range of quantity to be purchased	unit purchase cost
$b_0 < Q < b_1$	p_1
$b_1 \le Q < b_2$	p_2
$b_2 \le Q < b_3$	p_3
... ...	... ...
$b_{j-1} \le Q < b_j$	p_j
... ...	... ...
$b_{m-1} \le Q < b_m$	p_m

In general, $b_0 = 0$ and $b_m = \infty$ and $p_1 > p_2 > .. > p_j > .. > p_m$. The values $b_1, b_2, b_3, ..., b_{m-1}$ are termed as price breaks as the unit price lies in the intervals between these values.

Our problem is to determine an economic order quantity Q which minimizes the total cost.

In these models, the assumptions are

(i) Demand rate is known and uniform.

(ii) Shortages are not permitted.

(iii) Production for supply of commodities is instantaneous.

(iv) Lead time is zero.

$$Q_2^* = \sqrt{\frac{2\times40\times120}{0.02}} = 693$$

And $$Q_3^* = \sqrt{\frac{2\times60\times75}{0.04}} = 474$$

Therefore, the total average inventory is (447 + 693 + 474)/2 units = 807 units.

But the average inventory is 750 units. Therefore, we have to determine the value of parameter λ^* by trial and error method for computing Q_i^* by using

$$Q_i^* = \sqrt{\frac{2C_{3i}D}{C_{1i}+2\lambda^*}} \text{ and } \frac{1}{2}\sum Q_i^* = 750$$

Now, for $\lambda^* = 0.005$,

$$Q_1^* = \sqrt{\frac{2\times50\times100}{0.05+2\times0.005}} = 408,\ Q_2^* = 566 \text{ and } Q_3^* = 424$$

Therefore, the total average inventory is (408+566+424)/2 = 699 units which is less than the given average inventory of items.

Again, for $\lambda^* = 0.003$,

$$Q_1^* = 423,\ Q_2^* = 608,\ Q_3^* = 442$$

and the average inventory = (423 + 608 + 442) = 737 which is less than 750 units

Again, for $\lambda^* = 0.002$ then $Q_1^* = 430$, $Q_2^* = 632$, $Q_3^* = 452$

and the average inventory =(430+632+452) = 757 which is greater than 750 units.

Therefore, the most suitable value of λ^* lies between 0.002 and 0.003.

Let us assume that for λ^* = x, the average inventory will be 750.

Now, considering the linear relationship between λ^* and the average inventory, we have

$$\frac{x-0.003}{750-737} = \frac{0.003-0.002}{737-757} \quad \text{or, } x-0.003 = \frac{13\times0.001}{-20}$$

or, x = .00235 or, $\lambda^* = 0.00235$

For $\lambda^* = 0.00235$, $Q_1^* = 428$, $Q_2^* = 623$, $Q_3^* = 449$

Hence the optimal production quantities for products 1, 2 and 3 are 428, 623 and 449 units respectively.

$$\frac{x-6}{640-617.4} = \frac{5-6}{663.3-617.4}$$

or, $x-6 = \frac{-22.6}{45.9}$ or, x = 5.5 (approx)

$\therefore \quad \therefore \lambda^* = 5.5$

For this value of λ^*,

$$Q_1^* = \sqrt{\frac{2\times5000\times100}{2+2\times5.5\times0.60}} = 341$$

$$Q_2^* = \sqrt{\frac{2\times2000\times200}{2+2\times5.5\times0.80}} = 272$$

And $Q_3^* = \sqrt{\frac{2\times10000\times75}{1+2\times5.5\times0.45}} = 502$

Hence the optimal lot-size of three machine parts A, B, C are $Q_1^* = 341$ units, $Q_2^* = 272$ units and $Q_3^* = 502$ units.

Example - 6 :

A company producing three items has a limited inventories of averagely 750 items of all types. Determine the optimal production quantities for each item separately, when the following information is given :

Product	1	2	3
Holding cost (Rs.)	0.05	0.02	0.04
Ordering cost (Rs.)	50	40	60
Demand	100	120	75

Solution :

Neglecting the restriction of the total value of inventory level, we get the optimal values Q_i^* for i-th item which is given by $Q_i^* = \sqrt{\frac{2C_{3i}D_i}{C_{1i}}}$, $i = 1,2,3$

Hence, $Q_1^* = \sqrt{\frac{2\times50\times100}{0.05}} = 447$

This storage space is greater then the available storage space 640 sq. meters. Therefore, we shall try to find the suitable value of λ^* by trial and error method for computing Q_i^* by using

$$Q_i^* = \left[\frac{2C_{3i}D_i}{C_{1i} + 2\lambda^* a_i}\right]^{\frac{1}{2}}$$

and $\sum_{i=1}^{3} a_i Q_i^* = 640$

If we take $\lambda^* = 5$

$$Q_1^* = \sqrt{\frac{2C_{3i}D_i}{C_{1i} + 2\times 5\times a_1}} = \sqrt{\frac{2\times 5000\times 100}{2+2\times 5\times 0.60}} = 354$$

$$Q_2^* = \sqrt{\frac{2\times 2000\times 200}{3+2\times 5\times 0.80}} = 270$$

and $Q_3^* = \sqrt{\frac{2\times 10000\times 75}{1+2\times 5\times 0.45}} = 522$

Hence the corresponding storage space is 0.60354+0.80270+0.45552 = 663.30 sq. meters. This storage space is greater than the available storage space 640 sq. meters.

Again, if we take $\lambda^* = 6$, then

$$Q_1^* = \sqrt{\frac{2\times 5000\times 100}{2+2\times 6\times 0.60}} = 330$$

$$Q_2^* = \sqrt{\frac{2\times 2000\times 200}{3+2\times 6\times 0.80}} = 252$$

and $Q_3^* = \sqrt{\frac{2\times 10000\times 75}{1+2\times 6\times 0.45}} = 484$

Hence the corresponding storage space is 0.60330 + 0.80352+0.45484 = 617.40 sq. meters. which is less then the available storage space 640 sq. meters.

Hence it is clear that the most suitable value of λ^* lies between 5 and 6.

Let us assume that the required storage space will be 640 sq. meters for $\lambda^* = x$.

Now considering the linear relationship between the value of λ^* and the required storage space, we have

	Items		
	A	B	C
Cost per unit (Rs.)	10	15	5
Storage space required (sq. meter/unit)	0.60	0.80	0.45
Ordering cost (Rs.) (C_3)	100	200	75
No. of units required/year	5000	2000	10,000

The carrying charge on each item is 20% of unit cost.

Solution : Considering one year as one unit of time, we have,

the carrying charge of A (C_{11}) = Rs. (20% of 10) = Rs. 2.00,

carrying charge of B (C_{12}) = Rs. (20% of 15) =Rs. 3.00

carrying charge of C (C_{13}) = Rs. (20% of 5) = Re. 1.00

Now, without considering the effect of restriction on storage space availability, the optimal value Q_i^* of i-th item is given by $Q_i^* = \sqrt{\frac{2C_{3i}D_i}{C_{1i}}},\ i = 1,2,3$

$$\therefore Q_1^* = \sqrt{\frac{2D_iC_{3i}}{C_{11}}} = \sqrt{\frac{2\times 5000\times 100}{2}} = 707$$

$$Q_2^* = \sqrt{\frac{2\times 2.000\times 200}{3}} = 516$$

$$Q_3^* = \sqrt{2\times 10000\times 75} = 1225$$

Then the total storage space required for the above values of Q_i^* (i = 1, 2, 3) is

$$\sum_{i=1}^{3} a_iQ_i^* = a_1Q_1^* + a_2Q_2^* + a_3Q_3^*$$

= (0.60707+0.80516+0.451225) sq. maters

= 1388.25 sq. maters

$$L = \sum_{i=1}^{n}\left[\frac{1}{2}C_{1i}Q_i + C_{3i}D_i / Q_i\right] + \lambda\left[\sum_{i=1}^{n} C_{4i}Q_i - M\right]$$

where λ(>0) is the Lagrange multiplier.

The necessary conditions for L to be minimum are

$$\frac{\partial L}{\partial Q_i} = 0, i = 1, 2, ..., n$$

and $$\frac{\partial L}{\partial \lambda} = 0$$

Now from $\frac{\partial L}{\partial Q_i} = 0$, we have

$$Q_i = \left[\frac{2C_{3i}D_i}{C_{1i} + 2\lambda C_{4i}}\right]^{\frac{1}{2}}, i = 1,2,...,n$$

Again, from $\frac{\partial L}{\partial \lambda} = 0$, we have

$$\sum_{i=1}^{n} C_{4i}Q_i - M = 0 \quad \text{or,} \quad \sum_{i=1}^{n} C_{4i}Q_i = M$$

Hence the optimum value of Q_i is given by

$$Q_i^* = \left[\frac{2C_{3i}D_i}{C_{1i} + 2\lambda^* C_{4i}}\right]^{\frac{1}{2}} \qquad (10)$$

and $$\sum_{i=1}^{n} C_{4i}Q_i^* = M \qquad (11)$$

Thus the values of Q_i^* are obtained from (10) subject to the condition given by (11) where the optimal value λ^* of λ can be found by either successive trial and error method or linear interpolation method.

Example - 5 :

A workshop produces three machine parts A, B, C and the total storage space available is 640 sq. meters. Obtain the optimal lot-size for each item from the following data:

$$Q_i^* = \left[\frac{2C_{3i}D_i}{C_{1i} + 2\lambda^* a_i}\right]^{\frac{1}{2}}, \quad i = 1, 2, ..., n \tag{7}$$

and $$\sum_{i=1}^{n} a_i Q_i^* = A \tag{8}$$

To obtain the values of Q_i^* from (7) we find the optimal value λ^* from λ by successive trial and error method or linear interpolation method subject to the condition given by (8). The equation (8) implies that Q_i^* must satisfy the inventory constraint in equality sense.

❍ 6.1.5.3. Limitation on investment

In this case, there is an upper limit M on the amount to be invested on inventory. Let C_{4i} be the unit price of the i-th item then

$$\sum_{i=1}^{n} C_{4i} Q_i \le M \tag{9}$$

Now two possibilities may arise :

Case - I : When $\sum_{i=1}^{n} C_{4i} Q_i^* \le M$

In this case, the constraint is satisfied by Q^*_i automatically. Hence the optimal values of Q^*_i ($i = 1, 2, ..., n$) are given by

$$Q_i^* = \sqrt{\frac{2C_{3i}D_i}{C_{1i}}}$$

Case -II : When $\sum_{i=1}^{n} C_{4i} Q_i^* > M$

In this case, our problem is as follows :

Minimize $C = \sum_{i=1}^{n}\left[\frac{1}{2}C_{1i}Q_i + C_{3i}D_i / Q_i\right]$ subject to the constraint (9).

To solve it, we shall use the Lagrange multiplier method and the corresponding Lagrangian function is

In this case, the optimal values Q_i^* (i = 1, 2, ..., n) given by $Q_i^* = \sqrt{\dfrac{2C_{3i}D_i}{C_{1i}}}$ satisfy the constraint (5) directly. Hence these optimal values $Q^*{}_i$ are the required values.

Case -2 : When $\sum_{i=1}^{n} a_i Q_i^* > A$

In this case, we have to solve the problem as follows :

Minimize $C = \sum_{i=1}^{n}\left[\frac{1}{2}C_{1i}Q_i + C_{3i}D_i / Q_i\right]$

subject to the constraint (5).

To solve it, we shall use the Lagrange multiplier method and the corresponding Lagrangian function is

$$L = \sum_{i=1}^{n}\left[\frac{1}{2}C_{1i}Q_i + C_{3i}D_i / Q_i\right] + \lambda\left[\sum_{i=1}^{n} a_i Q_i - A\right]$$

where λ (>0) is the Lagrange multiplier.

The necessary conditions for L to be minimum are

$$\frac{\partial L}{\partial Q_i} = 0, \quad i = 1, 2, ..., n$$

and $\dfrac{\partial L}{\partial \lambda} = 0$

Now from $\dfrac{\partial L}{\partial Q_i} = 0$, we have

$$\frac{1}{2}C_{1i} - \frac{C_{3i}D_i}{Q_i^2} + \lambda a_i = 0, \; i = 1, 2, ..., n \tag{6a}$$

Again, from $\dfrac{\partial L}{\partial Q_i} = 0$, we have

$$\sum_{i=1}^{n} a_i Q_i - A = 0 \tag{6b}$$

Solving (6a) and (6b), we have the optimal values of Q_i as or, $$Q_i = \left[\frac{2C_{3i}D_i}{C_{1i}+2\lambda}\right]^{\frac{1}{2}}$$

Again, from $\frac{\partial L}{\partial \lambda}=0$ we have

$$\sum_{i=1}^{n} Q_i - 2k = 0 \text{ or, } \sum_{i=1}^{n} Q_i = 2k$$

Hence the optimum value of Q_i is

$$Q_i^* = \left[\frac{2C_{3i}D_i}{C_{1i}+2\lambda^*}\right]^{\frac{1}{2}} \quad (3)$$

and $$\sum_{i=1}^{n} Q_i = 2k \quad (4)$$

To obtain the values of Q_i^* from (3) we find the optimal value of λ^* of λ by successive trial and error method or linear interpolation method, subject to the condition given by (4). This equation (4) implies that Q_i^* must satisfy the inventory constraint in equality sense.

○ 6.1.5.2 Limitation of floor space (or Warehouse capacity)

Here we shall discuss the multi-item inventory model with the limitation of warehouse floor space. Let A be the maximum storage area available for n different items, a_i be the storage area required per unit of i-th item, Q_i be the amount ordered for the i-th item.

Thus the storage requirement constraint becomes

$$\sum_{i=1}^{n} a_i Q_i \leq A, \; Q_i > 0$$

or, $$\sum_{i=1}^{n} a_i Q_i - A \leq 0 \quad (5)$$

Now two possibilities may arise :

Case-I : When $\sum_{i=1}^{n} a_i Q_i^* \leq A$

$\frac{1}{2}\sum_{i=1}^{n} Q_i \le k$ [Since the average number of inventory at any time for an item is $\frac{1}{2}Q_i$.]

or, $\sum_{i=1}^{n} Q_i - 2k \le 0$ (2)

Now two cases may arise :

Case - I : When $\frac{1}{2}\sum_{i=1}^{n} Q_i^* \le k$

In this case, the optimal values Q^*_i $(i = 1, 2, ..., n)$ given by

$$Q_i^* = \sqrt{\frac{2C_{3i}D_i}{C_{1i}}}$$

satisfy the constraint directly.

Case - II : When $\frac{1}{2}\sum_{i=1}^{n} Q_i^* > k$

In this case, we have to solve the following problem :

$$\text{Minimize } C = \sum_{i=1}^{n} [\frac{1}{2} C_{1i} Q_i + C_{3i} D_i / Q_i]$$

subject to the constraint (2).

To solve it, we shall use the Lagrange multiplier method and the corresponding Lagrangian function is

$$L = \sum_{i=1}^{n} [\frac{1}{2} C_{1i} Q_i + C_{3i} D_i / Q_i] + \lambda [\sum_{i=1}^{n} Q_i - 2k]$$

where $\lambda (> 0)$ be the Lagrange multiplier.

The necessary conditions for L to be minimum are

$$\frac{\partial L}{\partial Q_i} = 0, \quad i = 1, 2, ..., n \text{ and } \frac{\partial L}{\partial \lambda} = 0$$

Now, from $\frac{\partial L}{\partial Q_i} = 0$ we have

$$\frac{1}{2} C_{1i} - \frac{C_{3i} D_i}{Q_i^2} + \lambda = 0, \; i = 1, 2, ..., n$$

(v) T_i be the cycle length.

Let Q_i be the ordering quantity of i-th item.

Then, $Q_i = D_i T_i \; or, T_i = Q_i / D_i$

Now, the total inventory time units for the i-th item is $\frac{1}{2} Q_i T_i$.

Hence the inventory carrying cost for i-th item over the inventory cycle is $\frac{1}{2} C_{1i} Q_i T_i$.

Therefore, the average cost for the i-th item is

$$C_i = [C_{3i} + \frac{1}{2} C_{1i} Q_i T_i] / T_i$$

$$\text{or, } C_i = C_{3i} D_i / Q_i + \frac{1}{2} C_{1i} Q_i$$

Hence the total average cost for n items is given by

$$C = \sum_{i=1}^{n} C_i$$

$$\text{or, } \quad C = \sum_{i=1}^{n} [C_{3i} D_i / Q_i + \frac{1}{2} C_{1i} Q_i]$$

Here C is a function of $Q_1, Q_2, \ldots, Q_n$.

For optimum values of Q_i ($i = 1, 2, \ldots, n$), we must have

$$\partial C / \partial Q_i = 0$$

$$\text{i.e., } \frac{1}{2} C_{1i} - C_{3i} D_i / Q_i^2 = 0$$

$$\text{or, } Q_i = \sqrt{\frac{2 C_{3i} D_i}{C_{1i}}}$$

Hence the optimum value of Q_i is $Q_i^* = \sqrt{\frac{2 C_{3i} D_i}{C_{1i}}}$ (1)

❍ 6.1.5.1 Limitation on inventories

If there is a limitation on inventories that requires that the average number of all units in inventory should not exceed k units of all types, then the problem is to minimize the cost C subject to the condition that

Solution:

For this problem, it is given that

C_1 = Rs. 0.15 per month, C_2 = Rs. 20 per month, C_3 = Rs. 500.00 per setup, K = 3000 per month, D = 18000 units per year i.e., 1500 units per month

The optimum manufacturing quantity Q^* is given by

$$Q^* = \sqrt{\frac{2C_3(C_1+C_2)}{C_1C_2}}\sqrt{\frac{KD}{K-D}} = \sqrt{\frac{2\times 500\times(0.15+20)}{0.15\times 20}}\sqrt{\frac{3000\times 1500}{3000-1500}} = 4489 \text{ units(approx.)}$$

The optimum shortage quantity is given by

$$S_2^* = C_1\frac{K-D}{K}\frac{Q^*}{(C_1+C_2)} = 17 \text{ units (approx.)}$$

Manufacturing time = $\frac{Q^*}{K} = \frac{4489}{3000} = 1.5$ months and the time between setups $\frac{Q^*}{D} = \frac{4489}{1500} = 3$ months.

❍ 6.1.5 Multi-item Inventory Model

So far we have solved the inventory models for single item or each item separately but if there exist a relationship among the items under some limitations then it is not possible to consider them separately. Thus after constructing the average cost expression in such models, we shall use the method of Lagrange multiplier to minimize the average cost.

In all such problems, first of all we shall solve the problem ignoring the limitations and then consider the effect of limitations.

Now, we shall develop multi-item inventory model under the following assumptions and notations :

(i) There are n items with instantaneous production i.e., the production rate of each item is infinite.

(ii) Shortages are not allowed.

For i-th ($i = 1, 2, ..., n$) item :

(iii) D_i be the uniform demand rate.

(iv) The inventory carrying cost, C_{1i} per unit quantity per unit time and the ordering cost, C_{3i} per order are known and constant.

$$Q^* = \sqrt{\frac{2C_3(C_1+C_2)}{C_1C_2}}\sqrt{\frac{KD}{K-D}} \qquad (32)$$

and $$S_2^* = \sqrt{\frac{2C_1C_3}{C_2(C_1+C_2)}}\sqrt{\frac{D(K-D)}{K}} \qquad (33)$$

$$T^* = \frac{Q^*}{D} = \sqrt{\frac{2C_3(C_1+C_2)}{C_1C_2}}\sqrt{\frac{K}{D(K-D)}} \qquad (34)$$

$$S_1^* = \frac{K-D}{K}Q^* - S_2^* = \sqrt{\frac{2C_2C_3}{C_1(C_1+C_2)}}\sqrt{\frac{D(K-D)}{K}} \qquad (35)$$

Now $$C_{\min} = C(Q^*, S_2^*) = \sqrt{\frac{2C_1C_2C_3}{C_1+C_2}}\sqrt{\frac{D(K-D)}{K}} \qquad (36)$$

Remarks:

(i) In this model, if we assume that the production rate is infinite i.e., $K \to \infty$, then the optimal quantities by taking $K \to \infty$ in (32), (34) and (36) are

$$Q^* = \sqrt{\frac{2C_3(C_1+C_2)}{C_1C_2}D}, \quad T^* = \sqrt{\frac{2C_3(C_1+C_2)}{C_1C_2D}} \text{ and } C_{\min} = \sqrt{\frac{2C_1C_2C_3D}{C_1+C_2}}$$

This means that Model – 4 reduces to Model – 3 if $K \to \infty$.

(ii) If shortages are not allowed in Model – 4, then it reduces to Model – 3. In this case, taking $C_2 \to \infty$ in (32), (34) and (36) we obtain the required expressions of model – 3 which are as follows :

$$Q^* = \sqrt{\frac{2C_3KD}{C_1(K-D)}}, \quad T^* = \sqrt{\frac{2C_3K}{C_1D(K-D)}} \text{ and } C_{\min} = \sqrt{\frac{2C_1C_3D(K-D)}{K}}$$

Example – 4.

The demand for an item in a company is 18000 units per year. The company can produce the item at a rate of 3000 per month. The cost of one set-up is Rs. 500 and the holding cost of one unit per month is Rs. 0.15. The shortage cost of one unit is Rs. 20 per month. Determine the optimum manufacturing quantity and the shortage quantity. Also determine the manufacturing time and the time between setups.

and $S_2 = (K-D)t_4$ or, $t_4 = \dfrac{S_2}{K-D}$ (24)

Since the total quantity produced over the time period T is Q,

Q = DT where D is the demand rate

or, $D(t_1 + t_2 + t_3 + t_4) = Q$

$$\text{or, } D\left(\frac{S_1}{K-D} + \frac{S_1}{D} + \frac{S_2}{D} + \frac{S_2}{K-D}\right) = Q \tag{25}$$

After simplification, we have $S_1 + S_2 = \dfrac{K-D}{K}Q$ (26)

Again, $t_1 + t_2 = \dfrac{K}{D(K-D)}S_1$ and $t_3 + t_4 = \dfrac{K}{D(K-D)}S_2$ (27)

Now substituting the values of $t_1 + t_2$, $t_3 + t_4$ and $T = Q/D$ in (21), we have

$$C(Q, S_1, S_2) = \frac{1}{2Q}\frac{K}{K-D}\left(C_1 S_1 + C_2 S_2^2\right) + \frac{DC_3}{Q} \tag{28}$$

Using (26), the above reduces to

$$C(Q, S_2) = \frac{1}{2Q}\frac{K}{K-D}\left[C_1\left(\frac{K-D}{K}Q - S_2\right)^2 + C_2 S_2^2\right] + \frac{DC_3}{Q} \tag{29}$$

Now, for the extreme values of C(Q, S_2), we have

$$\frac{\partial C}{\partial Q} = 0, \quad \frac{\partial C}{\partial S_2} = 0$$

$$\frac{\partial C}{\partial Q} = 0 \text{ implies } S_2 = C_1\frac{K-D}{K}\frac{Q}{C_1 + C_2} \tag{30}$$

$$\text{Again, } \frac{\partial C}{\partial S_2} = 0 \text{ gives } Q = \sqrt{\frac{2C_3(C_1 + C_2)}{C_1 C_2}} \cdot \sqrt{\frac{KD}{K-D}} \tag{31}$$

For these values of Q and S_2 given in (31) and (30), it can easily be verified that

$$\frac{\partial^2 C}{\partial Q^2} > 0, \quad \frac{\partial^2 C}{\partial S_2^2} > 0 \text{ and } \frac{\partial^2 C}{\partial Q^2}\frac{\partial^2 C}{\partial S_2^2} - \left(\frac{\partial^2 C}{\partial Q \partial S_2}\right)^2 > 0$$

Hence C(Q, S_2) is minimum and the optimal values of Q and S_2 are given by

Let us assume that each production cycle of length T consists of two parts t_{12} and t_{34} which are further subdivided into t_1 and t_2, t_3 and t_4 where (i) inventory is building up at a constant rate K – D units per unit time during the interval $[0, t_1]$, (ii) at time $t = t_1$, the production is stopped and the stock level decreases due to meet up the customer's demand only upto the time $t = t_1 + t_2$, (iii) Shortages are accumulated at a constant rate of D units per unit time during the time t_3 i.e., during the interval $[t_{12}, t_{12} + t_3]$. (iv) Shortages are being filled up immediately at a constant rate K – D units per unit time during the time t_4 i.e., during the interval $[t_{12} + t_3, t_{34}]$. (v) The production cycle then repeats itself after the time $T = t_1 + t_2 + t_3 + t_4$.

Again, let the inventory level is S_1 at $t = t_1$ and at the end of time $t = t_1 + t_2$, the stock level reaches to zero. Now shortages start and suppose that shortages are build up of quantity S_2 at time $t = t_1 + t_2 + t_3$ and then these shortages be filled up upto the time $t = t_1 + t_2 + t_3 + t_4$. The pictorial representation of the inventory situation is given in Fig. – 4.

Now our objectives are to find the optimal value of Q, S_1, S_2, t_1, t_2, t_3, t_4 and T with the minimum average total cost.

Now the inventory carrying cost over the time period T is given by

$$C_h = C_1 \times \Delta OAC = C_1.\frac{1}{2}OC.AB = \frac{1}{2}C_1(t_1 + t_2)S_1$$

and the shortage cost over time T is given by

$$C_S = C_2 \times \Delta CEF = C_2.\frac{1}{2}CF.EH = \frac{1}{2}C_2(t_3 + t_4)S_2.$$

Hence the total average cost of the production system is given by

$$C = [C_3 + C_h + C_s]/T \tag{21}$$

From Fig. – 4, it is clear that $S_1 = (K - D)t_1$ or, $t_1 = \frac{S_1}{K - D}$ (22)

Again, $S_2 = Dt_2$ or, $t_2 = \frac{S_2}{D}$ (23)

Now, in stock-out situation, $S_2 = Dt_3$ or, $t_3 = \frac{S_2}{D}$

❍ 6.1.4 Manufacturing model with shortage or Economic lot-size model with finite rate of replenishment and shortages (Model-4)

In this model, we shall derive the formula for the optimum production quantity, shortage quantity and cycle length of a single product by minimizing the average cost of the production system under the following assumptions and notations:

(i) The production rate or replenishment rate is finite, say K units per unit time ($K > D$).

(ii) The production – inventory planning horizon is infinite and the production system involves only one item and one stocking point.

(iii) Demand of the item is deterministic and uniform at a rate D unit of quantity per unit time.

(iv) Shortages are allowed.

(v) Lead time is zero.

(vi) The inventory carrying cost, C_1 per unit quantity per unit time, the shortage cost, C_2 per unit quantity per unit time and the set up cost, C_3 per set up are known and constant.

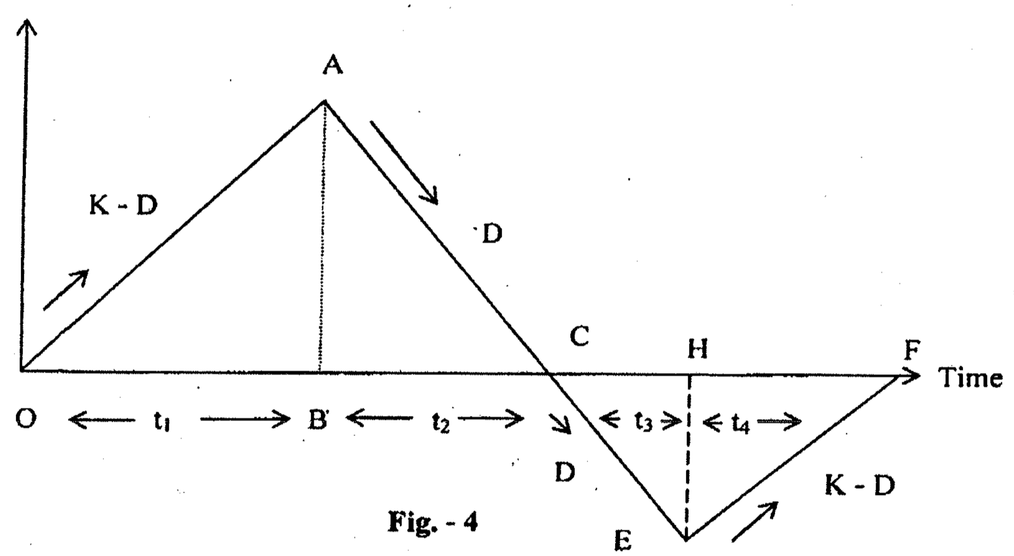


Fig. - 4

(vii) T be the cycle length of the system i.e., T be the interval between production cycle.

(viii) Q be the economic lot-size.

Again, the optimum shortage quantity $Q^* - S_1^* = 3857 - \sqrt{\frac{2C_2C_3D}{C_1(C_1+C_2)}}$

$$= 3857 - \sqrt{\frac{2\times 5\times 400\times 18000}{1.2\times(1.2+5)}} = 746 \text{ units (Approx.)}$$

Optimal cycle length $T^* = \frac{Q^*}{D} = \frac{3857}{18000} = 0.214$ year (Approx.)

Example – 3 :

The demand for an item is deterministic and constant over time and it is equal to 600 units per year. The unit cost of the item is Rs. 50.00 while the cost of placing an order is Rs. 100.00. The inventory carrying cost is 20% of the unit cost of the item and the shortage cost per month is Re. 1. Find the optimal ordering quantity. If shortages are not allowed, what would be the löss of the company?

Solution: It is given that D = 600 units/ year

C_1 = 20% of Rs. 50.00 = Rs. 10.00

C_2 = Re 1.00 per month i.e., Rs. 12.00 per year

C_3 = Rs. 100.00 per order

When shortages are allowed, the optimal ordering quantity Q^* is given by

$$Q^* = \sqrt{\frac{2C_3(C_1+C_2)D}{C_1C_2}} = 148 \text{ units}$$

and the minimum cost per year is $C(Q^*) = \sqrt{2C_1C_2C_3D/(C_1+C_2)}$ = Rs. 809.04

If shortages are not allowed, then the optimal order quantity is

$$Q^* = \sqrt{\frac{2C_3D}{C_1}} = 109.5 \text{ units}$$

and the relevant average cost is given by $C(Q^*) = \text{Rs.}\sqrt{2C_1C_3D}$ = Rs. 1095.44.

Therefore, if shortages are not allowed, the loss of the company will be Rs. (109.44 - 809.04) i.e., Rs. 286.40.

Remark:

(i) If $C_1 \to \infty$ and $C_2 > 0$, inventories are prohibited. In this case $S_1^* = 0$ and each lot-size

$$Q^* = \sqrt{\frac{2C_3D}{C_2}}$$ is used to fill the backorders.

(ii) If $C_2 \to \infty$ and $C_1 > 0$, then shortages are prohibited. In this case, $S_1^* = Q^* = \sqrt{\frac{2C_3D}{C_1}}$ and each batch Q^* is used entirely for inventory.

(iii) If shortage costs are negligible, then $C_1 > 0$ and $C_2 \to 0$.

In this case, $S_1^* \to 0$ and $Q^* \to \infty$.

(iv) If the inventory carrying costs are negligible, then $C_1 \to 0$ and $C_2 > 0$. In this case, $Q^* \to \infty$ and $S_1^* \to \infty$ i.e., $S_1^* \to Q^*$. Thus, due to very small inventory carrying costs, large lot size should be ordered and used to meet up the future demand.

(v) When the inventory carrying costs and shortage costs are equal i.e., when $C_1 = C_2$, $\frac{C_1}{C_1+C_2} = \frac{1}{2}$.

In this case, $Q^* = \sqrt{2}\sqrt{\frac{2C_3D}{C_1}}$ which shows that the lot-size is $\sqrt{2}$ times of the lot-size of ***Model-1.***

Example – 2 :

The demand for an item is 18000 units per year. The inventory carrying cost is Rs. 1.20 per unit time and the cost of shortage is Rs. 5.00. The ordering cost is Rs. 400.00. Assuming that the replenishment rate is instantaneous, determine the optimum order quantity, shortage quantity, cycle length.

Solution : For the problem, it is given that demand (D) = 1800 units per year, carrying cost (C_1) = Rs. 1.20 per unit, shortage cost (C_2) = Rs. 5.00, ordering cost (C_3) = Rs. 400 per order.

The optimum order quantity Q^* is given by

$$Q^* = \sqrt{\frac{2C_3(C_1+C_2)D}{C_1C_2}} = \sqrt{\frac{2\times400\times(1.2+5)\times18000}{1.2\times5}} = 3857\text{units}$$

Case – 2 : In this case, cycle length or scheduling period T is a variable. Like Case - 1, the average cost of the inventory system will be

$$C=\left[C_3+\frac{1}{2}C_1\frac{S_1^2}{D}+\frac{1}{2}C_2\frac{(Q-S_1)^2}{D}\right]/T \tag{15}$$

where $Q = DT$

Here, the average cost C is a function of two independent variables T and S_1.

Now, for optimal value of C, we have

$$\frac{\partial C}{\partial S_1}=0 \quad \text{and} \quad \frac{\partial C}{\partial T}=0$$

Now, $\frac{\partial C}{\partial S_1}=0$ gives $S_1=C_2\frac{DT}{(C_1+C_2)}$ (16)

Again, $\frac{\partial C}{\partial T}=0$ gives $-\frac{C_1}{2D}\frac{S_1^2}{T^2}+C_2\frac{DT-S_1}{T}-\frac{C_2}{2D}\frac{(DT-S_1)^2}{T^2}-\frac{C_3}{T^2}=0$ (17)

Putting $S_1=C_2\frac{DT}{(C_1+C_2)}$ in above and simplifying, we have

$$T=T^*=\sqrt{\frac{2C_3(C_1+C_2)}{C_1C_2D}} \tag{18a}$$

Then, $$S_1=S_1^*=\sqrt{\frac{2C_2C_3D}{C_1(C_1+C_2)}} \tag{18b}$$

Obviously, for the values of T and S_1 given by (18a) and (18b),

$$\frac{\partial^2 C}{\partial S_1^2}>0,\ \frac{\partial^2 S}{\partial T^2}>0 \text{ and } \frac{\partial^2 C}{\partial S_1}\frac{\partial^2 C}{\partial T}-\left(\frac{\partial^2 C}{\partial S_1 \delta T}\right)^2>0$$

Hence C is minimum for the values of T and S_1 given by (18a) and (18b).

Therefore the optimum order quantity for minimum cost is given by

$$Q^*=DT^*=D\sqrt{\frac{2C_3(C_1+C_2)}{C_1C_2D}}=\sqrt{\frac{2C_3(C_1+C_2)D}{C_1C_2}} \tag{19}$$

and $$C_{\min}=C^*=\sqrt{\frac{2C_1C_2C_3D}{(C_1+C_2)}} \tag{20}$$

$$C_1 \text{ (Area of } \Delta OAB) = \frac{1}{2} C_1 S_1 t_1 = \frac{1}{2} C_1 S_1^2/D$$

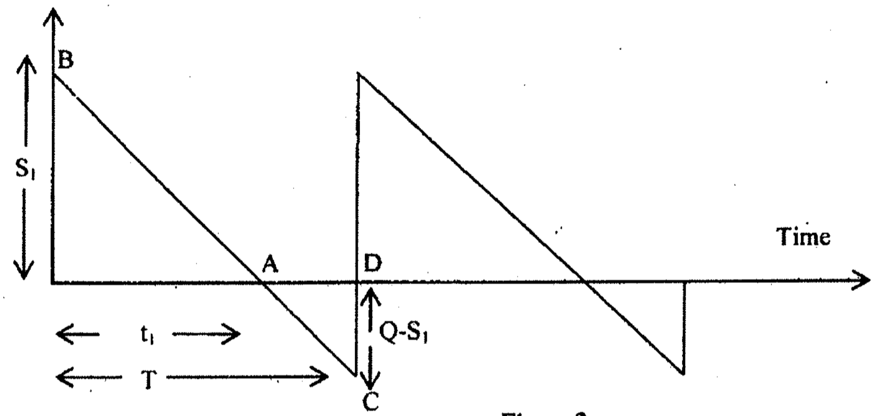


Fig. - 3

Again the shortage cost during the interval (t_1, T) is

$$C_2 \text{ (Area of } \Delta ACD) = \frac{1}{2} C_2 (Q - S_1)(T - t_1)$$

$$= \frac{1}{2} C_2 (Q - S_1)^2/D \quad \left[\because T - t_1 = \frac{Q - S_1}{D}\right]$$

Hence the total average cost of the system is given by

$$C = \left[\frac{1}{2} C_1 \frac{S_1^2}{D} + \frac{1}{2} C_2 \frac{(Q - S_1)^2}{D}\right] / T \tag{13}$$

Since the set-up cost C_3 and time period T are constant, the average set-up cost C_3/T also being constant will not be considered in the cost expression.

Since T is constant, Q = DT is also constant. Hence the above expression i.e., the expression for average cost is a function of single variable S_1. So, we can easily minimize the above expression (13) with respect to S_1 like Model - 1.

In this case, $S_1^* = \dfrac{C_2 Q}{C_1 + C_2} = \dfrac{C_2 DT}{C_1 + C_2}$ and $C_{min} = \dfrac{C_1 C_2 Q}{C_1 + C_2} = \dfrac{C_1 C_2 DT}{C_1 + C_2}$ (14)

If $K \to \infty$ i.e., the production rate is infinite, this model reduces to Model – 1. Therefore, when $K \to \infty$, then Q^*, T^* and C_{min} reduce to the expressions for Q^*, T^* and C_{min} of Model – 1.

❍ 6.1.3 Purchasing inventory model with shortages (Model – 3)

In this model, we shall derive the optimal order level and the minimum average cost under the following assumptions and notations:

(i) Demand is deterministic and uniform at a rate D unit of quantity per unit time.

(ii) Production is instantaneous (i.e., production rate is infinite).

(iii) Shortages are allowed and fully backlogged.

(iv) Lead time is zero.

(v) The inventory planning horizon is infinite and the inventory system involves only one item and one stocking point.

(vi) Only a single order will be placed at the beginning of each cycle and the entire lot is delivered in one batch.

(vii) The inventory carrying cost, C_1 per unit quantity per unit time, the shortage cost, C_2 per unit quantity per unit time, the ordering cost, C_3 per order are known and constant.

(viii) Q be the lot-size per cycle where as S_1 is the initial inventory level after fulfilling the backlogged quantity of previous cycle and $Q - S_1$ be the maximum shortage level.

(ix) To be the cycle length or scheduling period whereas t_1 be the no shortage period.

According to the assumptions of (viii) and (ix), we have $Q = DT$.

Regarding the cycle length or scheduling period of the inventory system, two cases may arise:

Case – 1 : Cycle length or scheduling period T is constant.

Case – 2 : Cycle length or scheduling period T is a variable.

Case – 1 : In this case, T is constant i.e., inventory is to be replenished after every time period T. As t_1 be the no shortage period, $S_1 = Dt_1$, or, $t_1 = S_1 / D$.

Now, the inventory carrying cost during the period 0 to t_1 is

Therefore the total average cost is given by $C(Q) = \frac{X}{T}$

Or, $C(Q) = \frac{C_3}{T} + \frac{1}{2}C_1 S$

Or, $C(Q) = \frac{C_3 D}{Q} + \frac{1}{2}C_1 \frac{K-D}{K}Q \quad \left[\because Q = DT \text{ and } S = \frac{K-D}{K}Q\right]$ (7)

The optimum of Q which minimizes C(Q) is obtained by equating the first derivative of C(Q) with respect to Q to zero

i.e., $\frac{dC}{dQ} = 0$

Or, $-\frac{C_3 D}{Q^2} + \frac{1}{2}C_1 \frac{K-D}{K} = 0$

Or, $Q = \sqrt{\frac{2C_3}{C_1} \cdot \frac{DK}{K-D}}$ (8)

Again, $\frac{d^2C}{dQ^2} = \frac{2C_3 D}{Q^3}$ = + ve quantity for $Q = \sqrt{\frac{2C_3}{C_1} \cdot \frac{DK}{K-D}}$

Hence C(Q) is minimum for which the optimum value of Q is

$$Q^* = \sqrt{\frac{2C_3}{C_1} \cdot \frac{DK}{K-D}} \quad (9)$$

The corresponding time interval is

$$T^* = \frac{Q^*}{D} = \sqrt{\frac{2C_3 K}{C_1 D(K-D)}} \quad (10)$$

and the minimum average cost is given by

$$C_{min} = \frac{1}{2}\frac{K-D}{K}C_1 Q^* + \frac{C_3 D}{Q^*} = \sqrt{2C_1 C_3 D \frac{K-D}{K}} \quad (11)$$

Remark:

(i) For this model, Q^*, T^* and C_{min} can be written in the following form :

$$Q^* = \sqrt{\frac{2C_3 D}{C_1} \frac{1}{1-D/K}}, \quad T^* = \sqrt{\frac{2C_3}{DC_1} \frac{1}{1-D/K}} \quad (12)$$

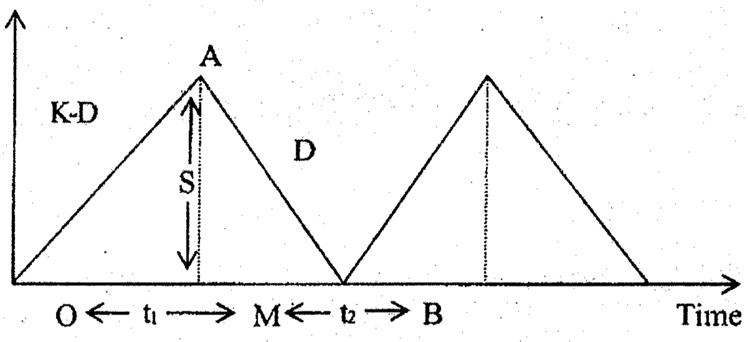


Fig. - 2

In this model, each production cycle time T consists of two parts t_1 and t_2 where

(i) t_1 is the period during which the stock is growing up a constant rate $K - D$ units per unit time.

(ii) t_2 is the period during which there is no replenishment (or production) but inventory is decreasing at the rate of D units per unit time.

Further, it is assumed that S is the stock available at the end of time t_1 which is expected to be consumed during the remaining period t_2 at the consumption rate D.

Therefore, $(K-D)t_1 = S$

or, $$t_1 = \frac{S}{K-D} \tag{5}$$

Since the total quantity produced during the production period t_1 is Q,

$$\therefore Q = Kt_1$$

or, $Q = K\frac{S}{K-D}$ which implies $$S = \frac{K-D}{K}Q \tag{6}$$

Again, $Q = DT$ i.e., $T = \dfrac{Q}{D}$

Now the inventory carrying cost for the entire cycle T is $(\Delta OAB)C_1 = \frac{1}{2}TSC_1$

and the setup cost for time period T is C_3.

Therefore, the total cost for the entire cycle T is given by $X = C_3 + \frac{1}{2}C_1ST$

Inventory carrying cost = interest costs + Deterioration and Obsolence costs + Storage costs

$$=\left(0.06+0.004+\frac{1000}{5000}\right)\text{rupees per unit per year} = \text{Rs. } 0.264 \text{ per unit per year}$$

Hence the economic order quantity is given by

$$Q^{*}=\sqrt{\frac{2C_3D}{C_1}}=\sqrt{\frac{2\times1500\times5000}{0.264}}=753.8(approx.)$$

Also, the minimum average cost is

$$\sqrt{2C_1C_3D}=Rs.\sqrt{2\times0.264\times1500\times5000}=Rs.1989.97(Approx.)$$

❍ 6.1.2 Manufacturing model with no shortages or economic lot-size model with finite rate of replenishment and without shortage (Model – 2)

In this model, we shall derive the formula for the optimum production quantity per cycle of a single product so as to minimize the total average cost under the following assumptions and notations:

(i) Demand is deterministic and uniform at a rate D unit of quantity per unit time.

(ii) Shortages are not allowed.

(iii) Lead time is zero.

(iv) The production rate or replenishment rate is finite, say, K units per unit time $(K > D)$.

(v) The production – inventory planning horizon is infinite and the production system involves only item and one stocking point.

(vi) The inventory carrying cost, C_1 per unit quantity per unit time, the setup cost, C_3 per production cycle are known and constant.

(vii) T be the cycle length and Q, the economic lot size.

(iii) In the above model, if we always maintain an inventory B on hand as buffer stock, then the average inventory at any time is $\frac{1}{2}Q+B$. Therefore, the total cost per unit time is

$$C(Q)=\left(\frac{1}{2}Q+B\right)C_1+C_3\frac{D}{Q}$$

As before, we obtain the optimal values of Q and T as follows:

$$Q=Q^*=\sqrt{\frac{2C_3D}{C_1}} \text{ and } T=T^*=\sqrt{\frac{2C_3}{DC_1}}$$

(iv) In the above model, if the ordering cost is taken as C_3+bQ (where b is the purchase cost per unit quantity) instead of fixed ordering cost then there is no change in the optimum order quantity.

Proof: In this case, the average cost is given by

$$C(Q)=\frac{1}{2}C_1Q+\frac{D}{Q}(C_3+bQ) \qquad (4)$$

The necessary condition for the optimum of $C(Q)$ in (4), we have

$$C'(Q)=0 \text{ implies } Q=\sqrt{\frac{2C_3D}{C_1}} \text{ and } C''(Q)>0.$$

Hence $Q^*=\sqrt{\frac{2C_3D}{C_1}}$

This shows that there is no change in Q^* in spite of change in the ordering cost.

Example 1 :

An engineering factory consumes 5000 units of a component per year. The ordering, receiving and handling costs are Rs. 300 per order while the trucking cost is Rs. 1200 per order, Interest cost Rs. 0.06 per unit per year, Deterioration and obsolence cost Rs. 0.004 per unit per year and storage cost Rs. 1000 per year for 5000 units. Calculate the economic order quantity and minimum average cost.

Solution : In the given problem, we have demand (D) = 5000 units

Ordering cost / Replenishment cost = Ordering, receiving, handling costs and trucing costs = Rs. (300 + 1200) = Rs. 1500 per order.

i.e., $\frac{dC}{dQ}=0$ or, $\frac{1}{2}C_1-\frac{C_3D}{Q^2}=0$

or, $Q = \sqrt{\frac{2C_3D}{C_1}}$

Again, $\frac{d^2C(Q)}{dQ^2}=\frac{2C_3D}{Q}Q^2=\sqrt{\frac{2C_3D}{C_1}}$ which is + ve for $Q=\sqrt{\frac{2C_3D}{C_1}}$

Hence C(Q) is minimum for which the optimum value of Q is

$$Q^* = \sqrt{\frac{2C_3D}{C_1}} \qquad (3)$$

This is known as economic lot size formula or EOQ formula. The corresponding optimum time interval is $T^* = \frac{Q^*}{D} = \sqrt{\frac{2C_3}{C_1D}}$

and the minimum cost per unit time is given by $C_{min} = \frac{C_3D}{Q^*}+\frac{1}{2}C_1Q^* = \sqrt{2C_1C_3D}$.

This model was first developed by Ford Harris of the Westing House Corporation, USA, in the year 1915. He derived the well-known classical lot size formula (3). This formula was also developed independently by R.H. Wilson after few years and it has been named as Harris – Wilson formula.

Remark :

(i) The total inventory time units for the entire cycle T is $\frac{1}{2}QT$, so the average inventory at any time is

$\frac{1}{2}QT/T=\frac{1}{2}Q$

(ii) Since $C_1 > 0$ from $f(Q) = ½ C_1Q$ it is obvious that the inventory carrying cost is a linear function of Q with a + ve slope i.e., for smaller average inventory, the inventory carrying costs are lower. In contrast, $g(Q)=\frac{C_3D}{Q}$ i.e., ordering cost increases as Q decreases.

Let us assume that an enterprise purchases an amount of Q units of item at time t = 0. This amount will be depleted to meet up the customer's demand. Ultimately, the stock level reaches to zero at time t = T. The inventory situation is shown in the Fig. – 1.

Clearly, $Q = DT$ (1)

Now, the inventory carrying cost for the entire cycle T is C_1 ×(area of ΔAOB) = $C_1 . (\frac{1}{2} QT)$ = $\frac{1}{2} C_1 QT$ and the ordering cost for the said cycle T is C_3.

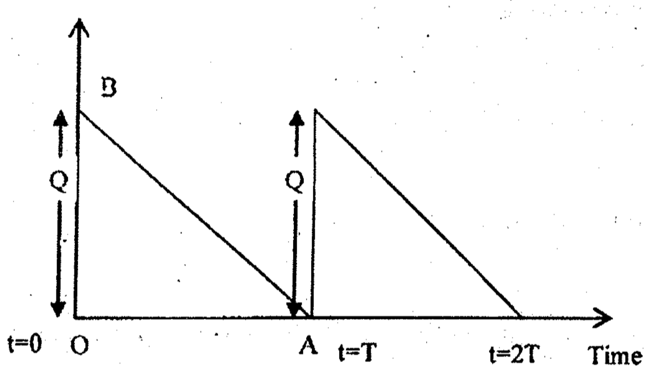


Fig. - 1

Hence the total cost for time T is given by

$$X = C_3 + \frac{1}{2} C_1 QT$$

Therefore, the total average cost is given by $C(Q) = \frac{X}{T}$

or, $C(Q) = \frac{C_3}{T} + \frac{1}{2} QC_1$

or, $$C(Q) = \frac{C_3 D}{Q} + \frac{1}{2} C_1 Q \qquad \left[\because Q = DT \therefore T = \frac{Q}{D} \right] \qquad (2)$$

The optimum value of Q which minimizes $C(Q)$ is obtained by equating the first derivative of $C(Q)$ with respect to Q to zero.

(ii) *Probabilistic inventory models*

These are the inventory models in which the demand is a random variable having a known probability distribution. Here, the future demand is determined by collecting data from the past experience.

❑ 112.6.1. Deterministic inventory models

There are different types of models under this category, namely

(a) Purchasing inventory model with no shortage

(b) Manufacturing inventory model with no shortage

(c) Purchasing inventory model with shortages

(d) Manufacturing model with shortages

(e) Multi-item inventory model

(f) Price break inventory model

❍ 6.1.1 Purchasing inventory model with no shortage (Model-1)

In this model, we want to derive the formula for the optimum order quantity per cycle of a single product so as to minimize the total average cost under the following assumptions and notations:

(i) Demand is deterministic and uniform at a rate D units of quantity per unit time.

(ii) Production is instantaneous (i.e., production rate is infinite).

(iii) Shortages are not allowed.

(iv) Lead time is zero.

(v) The inventory planning horizon is infinite and the inventory system involves only one item and one stocking point.

(vi) Only a single order will be placed at the beginning of each and the entire lot is delivered in one batch.

(vii) The inventory carrying cost, C_1 per unit quantity per unit time, the ordering cost, C_3 per order are known and constant.

(viii) T be the cycle length and Q, the ordering quantity per cycle.

stocks are held. This cost generally includes the costs such as insurance, taxes, obsolescence, deterioration, rent of warehouse, light, heat, maintenance and interest on the money locked up.

Shortage cost or Stock out cost

The shortage cost or stock out cost is the penalty incurred for being unable to meet up a demand when it occurs. This cost arises due to shortage of goods, lost sales for delay in meeting up the demand or total inability to meet up the demand. In the case, where the unfulfilled demand for the goods can be satisfied a latter date (backlogging case), this cost depends on the shortage quantity and delaying time both. On the other hand, if the unfulfilled demand is lost (no backlogging case); shortage cost becomes proportional to the shortage quantity only. In both cases, there is a loss of goodwill which can not be quantified for the development of mathematical model.

Disposal cost

When an amount of some units of an item remains excess at the end of inventory cycle and if this amount is sold at a lower price in the next cycle to derive some advantages like clearing the stock, winding up the business, etc., the revenue earned through such a process is called the disposal cost.

Salvage values

During storage, some units are partially spoiled or damaged i.e., some units loose their utility partially. In a developing country, it is normally observed that some of these are sold at a reduced price (less than the purchase price) to a section of customers and this gives some revenue to the management. This revenue is called the salvage value.

❑ 112.6. CLASSIFICATION OF INVENTORY MODELS

The inventory problems (models) may be classified into two categories.

(i) Deterministic inventory models

These are the inventory models in which demand is assumed to be known constant or variables (dependent on time, stock-level, selling price of the item, etc.). Here, we shall consider deterministic inventory models for known constant demand. Such models are usually referred to as Economic lot-size models or Economic Order Quantity (EOQ) models.

varies with the size of stock and the duration for which the stress is applied. Items made of glass, china-clay, ceramic, mud etc. are examples of such products.

Perishable items are those which have finite life time (fixed or random). Fixed life time product (e.g. human blood, etc.) has a deterministic self life while the random life time scenario is closely related to the case of an inventory which experiences continuous physical depletion due to deterioration or decay.

Various types of inventory costs

Inventory costs are the costs associated with the operation of an inventory system and result from action or lack of action on the part of management in establishing the system. They are basic economic parameters to any inventory decision model.

Purchase or Unit cost

The purchase or unit cost of an item is the unit purchase price to obtain the item from an external source or the unit production cost for the internal production. It may also depend upon the demand. When production is done in large quantities, it results in reduction of production cost per unit. Also, when quantity discounts are allowed for bulk orders, unit price is reduced and dependent on the quantity purchased or ordered.

Ordering / set up cost

The ordering or set up cost originates from the experience of issuing a purchase order to an outside supplier or from an internal production set up costs. The ordering cost includes clerical and administrative costs, telephone charges, telegrams, transportation costs, loading and unloading costs etc. generally, this cost is assumed to be independent of the quantity ordered for or produced. In some cases, it may depend on the quantity of goods purchased because of price break or quantity discounts or transportation cost, etc.

Holding or Carrying cost

The holding or carrying cost is the cost associated with the storage of the inventory until its use or sale. It is directly proportional to the amount / quantity in the inventory and the time for which the

Constraints

Constraints are the limitations imposed on the inventory system. It may be imposed on the amount of investment, available space, the amount of inventory held, average instantaneous expenditure, number of orders, etc.

Fully backlogged / partially backlogged shortages

During stock out period, the sales or goodwill may be either by a delay or complete refusal in meeting the demand. If the unfulfilled demand for the goods is satisfied completely at a later date, then it is a case of fully backlogged shortage i.e., it is assumed that no customer balk away during this period and the demand of all these waiting customers is met up at the beginning of the next period gradually after the commencement of next production.

Again, it is normally observed that during the stock out period, some of the customers wait for the product and others balk away. When this happens, the phenomenon is called partially backlogged shortages.

Lead time

The time gap between the time of placing an order or production start and the time of arrival of goods in stock is called the lead time. It may be a constant or a variable. Again, variable lead time may be probabilistic or imprecise.

Planning / time Horizon

The time period over which the inventory level will be controlled is called the time / planning horizon. It may be finite or infinite depending upon the nature of the inventory system for the commodity.

Deterioration / Damageability / Perishability

Deterioration is defined as decay, evaporation, obsolescence and loss of utility or marginal value of a commodity that results in the decreasing usefulness from the original condition. Vegetables, foodgrains and semi conductor chips, etc. are examples of such products.

Damageability is defined by the damage when the items are broken or loose their utility due to the accumulated stress, bad handling, hostile environment etc. The amount of damage by the stress

❑ 112.5. BASIC TERMINOLOGIES IN INVENTORY

The inventory system depends on several system factors and parameters such as demand, replenishment rate, shortages, constraints, various types of costs etc.

Demand

Demand is defined as the number of units of an item required by the customer in a unit time and has the dimension of a quantity. It may be known exactly or known in terms of probabilities or may be completely unknown.

The demand pattern of items may be either deterministic or probabilistic. Problems in which demand is known and fixed are called deterministic problem. Whereas those problems in which the demand is assumed to be a random variable are called stochastic or probabilistic problems.

In case of deterministic demand it is assumed that the quantities needed over subsequent periods of time are known exactly. Further, the known demand may be fixed or variable with time or stock level or selling price of an item etc.

Probabilistic demand occurs when requirements over a certain period of time are not known with certainty but their pattern can be described by a known probability distribution.

In some cases, demand may also be represented by uncertain data in non-stochastic sense i.e., by vague / imprecise data. This type of demand is termed as fuzzy demand and the system as a fuzzy system.

Replenishment

Replenishment refers to the amount of quantities that are scheduled to be into inventories, at the time when decisions are made about ordering these quantities or to the time when they are actually added to stock. It can be categorized according to size, pattern and lead time. Replenishment size may be constant or variable, depending upon the type of the inventory system. It may depend on time, demand and / or on-hand inventory level. The replenishment patterns are usually instantaneous, uniform or in bateh. The replenishment quantity again may be probabilistic or fuzzy in nature.

❑ 112.4. TYPES OF INVENTORY

There are different types of inventory, namely:
(i) Transportation inventories, (ii) Fluctuation inventories, (iii) Anticipation inventories, (iv) De-coupling inventories, (v) Lot-size inventories.

Transportation inventories

This arises due to transportation of inventory items to various distribution centres and customers from the various production centres. When the transportation time is long, the items under transport can not be served to customers. These inventories exist solely because of transportation time.

Fluctuation inventories

These have to be carried because sales and production times can not be predicted accurately. In real-life problems, there are fluctuations in the demand and lead-times that affect the production of items.

Anticipation inventories

These are building up in advance by anticipating or foreseeing the future demand for the season of large sales due to a promotion programme or a plant shutdown period.

De-coupling inventories

The inventories used to reduce the interdependence of various stages of production system are known as de-coupling inventories.

Lot-size inventories

Generally, the rate of consumption is different from the rate of production or purchasing. Therefore, items are produced in larger quantities which result in lot-size, also called as cycle inventories.

model. We do not know what the real world is. Therefore, it is almost impossible to construct a realistic model with complete accuracy. For this reasons, some approximations and simplifications must be used during the model building process. Again, the solution of the inventory problem is a set of specific values of variables that minimizes the total (or average) cost of the system or maximizes the total (or average) profit of the system.

❑ 112. 2. OBJECTIVES

In this unit, our objectives are to

- discuss the concepts of inventory as well as the various forms of inventory and reasons for maintaining inventories.
- define the different terminologies of inventory like demand, replenishment, constraints, shortages, lead time, planning/time horizon, various types of inventory costs.
- derive the single item purchasing and manufacturing inventory models with and without shortages.
- determine the optimal order quantity for single item price break inventory models and multi-item purchasing model with different constraints like investment, average inventory and space constraints.
- introduce the probabilistic models for discrete (well known Newspaper-boy problem) and continuous cases.
- discuss the fuzzy inventory model with fuzzy non-linear programming.

❑ 112.3. KEYWORDS

Inventory, production, customer, commodity, resource, buffer stock , supplier, raw material, demand, replenishment, constraints, backlogged shortages, lead time, planning/time horizon, deterioration, perishable items, Damageability, ordering/set up cost, holding cost, shortage cost, disposal cost, salvage value, E.O.Q. model, price break, probabilistic demand, fuzzy membership function, fuzzy goal, fuzzy costs.

(vi) to plan overall operating strategy through decoupling of successive stages in the chain of acquiring goods, preparing products, shipping to branch warehouses and finally serving the customers.

(vii) to motivate the customers to purchase more by displaying large number of goods in the showroom / shop.

(viii) to take the advantages in purchasing of some raw materials and some commonly used physical goods (such as paddy, wheat etc.) whose prices seasonally fluctuate. In this connection, it is more profitable to procure a sufficient quantity of these raw materials / commonly used physical goods when their prices are low to be used later during the high price season or when need arises.

Production / Inventory planning and control is essentially concerned with the design-operation and control of an inventory system in any sector of a given economy. The problem of inventory control is primarily concerned with the following fundamental questions:

(i) Which items should be carried in stock? or which items should be produced?

(ii) How much of each of these items should be ordered / produced?

(iii) When should an order be placed? or when to produce?

(iv) What type of inventory control system should be used?

In practice, it is a formidable task to determine a suitable inventory policy. Regarding the above-mentioned questions, an inventory problem is a problem of making optimal decisions. In other words, an inventory problem deals with decisions that optimize either the cost function (total or average cost) or the profit function (total or average profit) of the inventory system. However, there are certain types of problems, such as those relating to the storage of water in a dam in which one has no control over the replenishment of inventory. The supply of inventory of water in a dam depends on rainfall and the organization operating the dam has no control over it.

Our aim is to formulate mathematical models of different inventory control systems and to solve those using different mathematical analysis. For this purpose, our task is to construct a mathematical model of the inventory system. However, this type of model is based on different assumptions and approximations. It is difficult both to devise and operate with an exact / accurate

❑ 112.1. INTRODUCTION

In broad sense, inventory is defined as an idle resource of an enterprise / company / manufacturing firm. It can be defined as a stock of physical goods, commodities or other economic resources which are used to meet up the customer's demand or requirement of production. This means that the inventory acts a buffer stock between a supplier and a customer.

The inventory or stock of goods may be kept in any one of the following forms:

(i) Raw materials

(ii) Semi-finished goods (work – in – process inventory)

(iii) Finished (or produced) goods

(iv) Maintenance, repair and operating supplies (MRO) items

In any sector of an economy, the control and maintenance of inventory is a problem common to all organization. Inventories of physical goods are maintained in government and non-government establishments e.g., Agriculture, Industry, Military, Business, etc.

Some reasons for maintaining the inventories are as follows:

(i) to conduct the smooth and efficient running of business.

(ii) to provide the customer service by meeting their demands from stock without delay.

(iii) to earn price discount for bulk purchasing.

(iv) to maintain more stable operating and / or workforce.

(v) to take the financial advantage of transporting/shipping economics.

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X Group-B

Module No. - 112

Advanced Optimization And Operational Research - II

....

MODULE STRUCTURE :

- 112.1. Introduction
- 112.2. Objectives
- 112.3. Keywords
- 112.4. Types of inventory
- 112.5 Basic terminologies of inventory
- 112.6 Classification of inventory models
  - 112. 6.1. Deterministic inventory models
    - 6.1.1 Purchasing inventory model without shortage
    - 6.1.2 Manufacturing inventory model without shortage
    - 6.1.3 Purchasing inventory model with shortages
    - 6.1.4 Manufacturing inventory model with shortages
    - 6.1.5 Multi-item inventory model
      - 6.1.5.1 Limitation on inventories
      - 6.1.5.2 Limitation of floor space (or warehouse capacity)
      - 6.1.5.3 Limitation on investment
    - 6.1.6 Price break inventory model
      - 6.1.6.1 Purchasing inventory model with single price break

[3] Klaus Ritzberger, Foundations of Non-cooperative Game Theory, 2002

[4] Martin J. Osborne and Ariel Rubinstein, A Course in Game Theory. The MIT Press, August 1994.

[5] McKinsey, J.C.C., Introduction to the Theory of Games, Dover Publications, 2003.

[6] Michael Mesterton Gibbons, An Introduction to Game-Theoretic Modelling, 2001.

[7] Nicolaj, Nikolaevich and Ralph Philip Boas, Foundations of Game Theory: Noncooperative Games, 1994.

---- 0 -----

2. Let M be the payoff function of a continuous game which possesses a solution where for all x and y in the interval $M(x, y) = - M(y, x)$. Prove that the value of the game is zero and every optimal strategy of one player is the optimal strategy of the other player.
3. Let M be the continuous pay-off function of a continuous game and suppose u is a real number and that F_0 and G_0 are distribution functions such that for all x and y in [0, 1]

$$\int_0^1 M(x,y)\,dG_0(y) \le u \le \int_0^1 M(x,y)\,dF_0(x)$$

Prove that u is the game and F_0, G_0 are optimal strategies of first and second players respectively.
4. Define strategically equivalent games. Prove that strategically equivalent games possess the same equilibrium situation. Show that equilibrium situation for an antagonistic game is the saddle point of the payoff function.
5. Let v be the value of the continuous game $\langle X, Y, M \rangle$. Prove that

$$\sup_{x\in X} \inf_{y\in Y} M(x,y) \le v \le \inf_{y\in Y} \sup_{x\in X} M(x,y)$$

6. The pay-off function of a continuous game is sin $2\pi(x - y)$. Find the value of this game. Show that $F_0(x) = x$ and $G_0(y) = y$ are optimal strategies of first and second players respectively.
7. Define a non-cooperative game with n number of players. What is equilibrium situation? When two games are said to be strategically equivalent? Show that strategically equivalent games obey symmetric and transitive properties.

❑ 111.11. REFERENCES

[1] A. J. Jones, Game Theory : Mathematical Model of Conflict, Horwood Publications, 2001

[2] Ferenc Forgó, Jenö Szep and Ferenc Szidarovszky, Introduction to the Theory of Games: Concepts, Methods, Applications, Springer Verlag, 1999.

are satisfied by choosing $F(x)=\begin{cases}\dfrac{7x}{2\pi}; & 0\le x\le \dfrac{2\pi}{7}\\ 1 & \dfrac{2\pi}{7}\le x\le 1\end{cases}$

In fact, $\int_0^1 \cos 7x\, dF(x)=\int_0^{\frac{2\pi}{7}} \cos 7x\, d\left(\dfrac{7x}{2\pi}\right)+\int_{\frac{2\pi}{7}}^1 \cos 7x\, d(1)$

$$=\frac{7}{2\pi}\left[\frac{\sin 7x}{7}\right]_0^{\frac{2\pi}{7}}=0$$

and $\int_0^1 \sin 7x\, dF(x)=\int_0^{\frac{2\pi}{7}} \sin 7x\, d\left(\dfrac{7x}{2\pi}\right)+\int_{\frac{2\pi}{7}}^1 \sin 7x\, d(1)$

$$=\frac{7}{2\pi}\left[-\frac{\cos 7x}{7}\right]_0^{\frac{2\pi}{7}}=-\frac{7}{2\pi}(1-1)=0$$

Similarly it can easily be verified that

$$G(y)=\begin{cases}\dfrac{8y}{2\pi} & 0\le y\le \dfrac{2\pi}{8}\\ 1 & \dfrac{2\pi}{8}\le y\le 1\end{cases}$$

is an optimal strategy for the player-2.

Note: This is example of a game where there exists infinite number of optimal strategies.

❑ 111.10. SELF ASSESSMENT QUESTIONS/EXERCISE

1. Let M be a continuous pay off function of a continuous game then prove that the following conditions are equivalent.

(i) F_0 and G_0 are optimal strategies of players 1 and 2 respectively.

(ii) $E(F, G_0)\le E(F_0, G_0)\le E(F_0, G)$ for all strategies F and G of players 1 and 2 respectively.

(iii) $\int_0^1 M(x',y)\,dG_0(y)\le E(F_0,G_0)\le \int_0^1 M(x,y')\,dF_0(x)$ for any $x', y'\in[0,1]$.

These are satisfied by $F_A(x) = \Theta\left(x - \frac{\pi}{7}\right)$

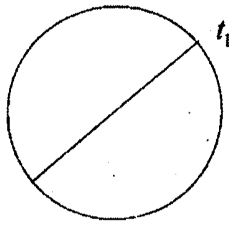
Similarly, $F_B(x)$ satisfies

$$1 = \int_0^1 r_1(t)\, dF_B(t) = \int_0^1 \cos 7t\, dF_B(t)$$

$$0 = \int_0^1 r_2(t)\, dF_B(t) = \int_0^1 \sin 7t\, dF_B(t)$$

These are satisfied by $F_B(x) = \Theta(x)$ or $\Theta\left(x - \frac{2\pi}{7}\right)$

Hence an optimal strategy for the player-1 is



either $F^*(x) = \frac{1}{2}\Theta\left(x - \frac{\pi}{7}\right) + \frac{1}{2}\Theta(x)$

or, $F^*(x) = \frac{1}{2}\Theta\left(x - \frac{\pi}{7}\right) + \frac{1}{2}\Theta\left(x - \frac{2\pi}{7}\right)$

Hence, we can also write

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \cos 7t_1 \\ \sin 7t_1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \cos 7t_2 \\ \sin 7t_2 \end{pmatrix} \text{ where } t_2 - t_1 = \frac{\pi}{7} \quad 0 \le t_1, t_2 \le 1$$

Hence a most general optimal strategy for the player-1 is

$$F^*(x) = \frac{1}{2}\Theta(x - x_1) + \frac{1}{2}\Theta(x - x_2) \text{ where } x_2 - x_1 = \frac{\pi}{7} \quad 0 \le x_1, x_2 \le 1$$

Similarly, a most general optimal strategy for the player-2 is

$$G^*(y) = \frac{1}{2}\Theta(y - y_1) + \frac{1}{2}\Theta(y - y_2),\ y_2 - y_1 = \frac{\pi}{8} \quad 0 \le y_1, x_2 \le 1$$

Also, $0 = \int_0^1 \cos 7x\, dF(x)$

$$0 = \int_0^1 \cos 7x\, dF(x)$$

$$\sum_{j=1}^{q} a_{ij} q_j + c_i = 0, \quad i = 1,2$$

i.e., $2q_1 + 5q_2 = 0$

$2q_1 + q_2 = 0$

$\therefore q_1 = 0, \ q_2 = 0$

$\therefore \chi = (0,0)' \in$ interior of Q.

Hence Π is the only fixed point of P and χ is the only fixed point of Q.

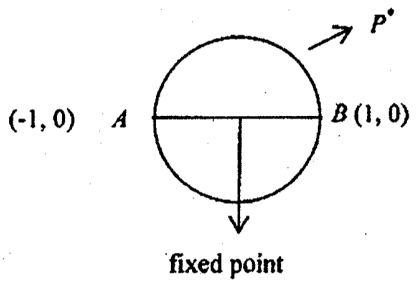
Value of the game

The value of the game is

$$E(\Pi, \chi) = \sum_{i=1}^{2} b_i p_i + d = \sum_{j=1}^{2} c_j q_j + d = 0$$

Optimal strategies

We write the fixed point of P i.e., $\Pi = (0,0)'$ as a convex combination of the points on P^*. We can obviously write



$$\begin{pmatrix}0\\0\end{pmatrix} = \frac{1}{2}\begin{pmatrix}-1\\0\end{pmatrix} + \frac{1}{2}\begin{pmatrix}+1\\0\end{pmatrix}$$

Hence an optimal strategy for the player-1 is

$$F^*(x) = \frac{1}{2} F_A(x) + F_B(x)$$

where $F_A(x)$ is a strategy corresponding to $A \equiv (-1,0)'$ and $F_B(x)$ is a strategy corresponding to $B = (1,0)'$.

Now to find $F_A(x)$, we note that it satisfies

$$-1 = \int_0^1 r_1(t)\, dF_A(t) = \int_0^1 \cos 7t \, dF_A(t)$$

$$0 = \int_0^1 r_2(t)\, dF_A(t) = \int_0^1 \sin 7t \, dF_A(t)$$

Hence $a_{11} = 3,\quad a_{12} = 5,\quad b_1 = b_2 = 0,\quad d = 0$

$a_{21} = 2,\quad a_{22} = 1,\quad c_1 = c_2 = 0$

$$\therefore \det\left[a_{ij}\right] = \begin{vmatrix} 3 & 5 \\ 2 & 1 \end{vmatrix} = 3 - 10 = -7 \neq 0$$

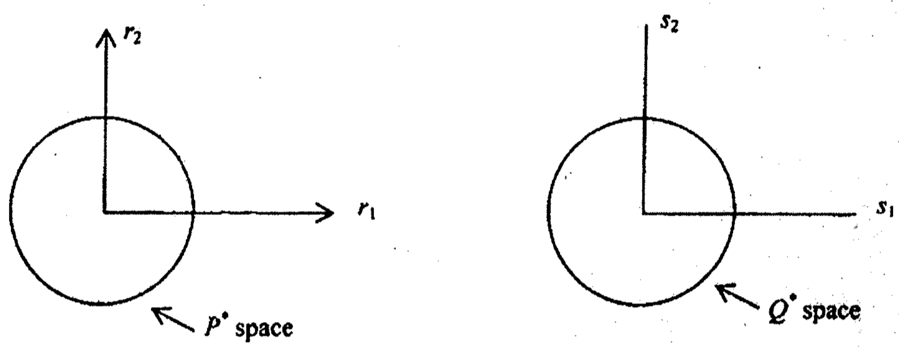
Hence this representation is canonical.

P^* and Q^* spaces

Here $P^* = \left\{\rho = (r_1, r_2)' : r_1 = \cos 7t,\ r_2 = \sin 7t,\ 0 \le t \le 1\right\}$

$Q^* = \left\{\sigma = (s_1, s_2)' : s_1 = \cos 8t,\ s_2 = \sin 8t,\ 0 \le t \le 1\right\}$

P^* and Q^* – spaces: P^* space is the circle with its interior and Q^* space is the circle with its interior.



First and Second critical points

$\Pi = (p_1, p_2)'$ is the unique solution of

$$\sum_{i=1}^{n} a_{ij}\, p_i + c_j = 0, \qquad j = 1,2$$

i.e. $3p_1 + 2p_2 = 0$

$5p_1 + p_2 = 0$

$\Rightarrow p_1 = 0,\ p_2 = 0$

$\therefore \Pi = (0,0)' \in$ interior of P space.

Now $\chi = (q_1, q_2)'$ is the unique solution of

Since Π is the unique solution of $\sum_{i=1}^{n} a_{ij} p_i + c_j = 0,\ j = 1,2,\cdots,n$, we have for some $k \le n$,

$\sum_{i=1}^{n} a_{ik} u_i + c_k \ne 0$ and equal to g

Let h be a real number of opposite sign to g, and is small enough to ensure that the point

$$\bar{\beta} = (v_1, \cdots, v_{k-1}, v_k + h, v_{k+1}, \cdots, v_n) \in Q$$

Now $$E(\alpha,\beta) = \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij} u_i + c_j\right) v_j + \sum_{i=1}^{n} b_i u_i + d$$

$$E(\alpha,\bar{\beta}) = E(\alpha,\beta) + h\left(\sum_{i=1}^{n} a_{ik} v_k + c_k\right)$$

$$= E(\alpha,\beta) + hg$$

$$< E(\alpha,\beta) \text{ as } hg < 0$$

This means that $\alpha \notin P(\beta)$ which contradicts the assumption that α and β are both fixed points of P and Q respectively. Hence α must coincide with Π.

Corollary 1: Let the payoff of a separable game be given in a canonical form and let the first critical point $\notin P$, then every fixed point of Q is in the boundary of Q. Similarly, if the second critical point $\notin Q$, then every fixed point of P is in boundary of P.

Corollary 2: Let the payoff function of a separable game be given in a canonical form and Π and χ be the first and second critical points respectively. Let Π belongs to the interior of P and χ belongs to the interior of Q, then Π is the only fixed point of P and χ is the only fixed point of Q.

Example: Solve the separable game whose payoff function is

$$M(x,y) = 3\cos 7x \cos 8y + 5\cos 7x \sin 8y + 2\sin 7x \cos 8y + \sin 7x \sin 8y$$

Solution: Let $r_1(x) = \cos 7x, \quad r_2(x) = \sin 7x$

$$s_1(y) = \cos 8y, \quad s_2(y) = \sin 8y$$

Then $M(x,y) = 3r_1(x)s_1(y) + 5r_1(x)s_2(y) + 2r_2(x)s_1(y) + r_2(x)s_2(y)$

and also let $\Pi=(p_1,p_2,\cdots,p_n)'$, $\chi=(q_1,q_2,\cdots,q_n)'$ be respectively the first and second critical points and suppose $\Pi\in P$, $\chi\in Q$. Then Π and χ are the fixed points and the value of the game is given by

$$E(\Pi,\chi)=\sum_{i=1}^{n}b_ip_i+d=\sum_{j=1}^{n}c_jq_j+d.$$

***Proof*:** By the definition of critical points

$$\sum_{i=1}^{n}a_{ij}p_i+c_j=0,\quad j=1,2,\cdots,n$$

$$\text{and}\quad \sum_{j=1}^{n}a_{ij}q_j+b_i=0,\quad i=1,2,\cdots,n$$

Now, for any point $\beta\in Q\left[\beta=(v_1,v_2,\cdots,v_n)'\right]$

$$E(\Pi,\beta)=\sum_{j=1}^{n}\left(\sum_{i=1}^{n}a_{ij}p_i+c_j\right)v_j+\sum_{i=1}^{n}b_ip_i+d$$

$$=0+\sum_{i=1}^{n}b_ip_i+d \text{ which is independent of } \beta.$$

Hence $Q(\Pi)=Q$. Similarly $P(\chi)=P$

Hence $\Pi\in P(\chi)$ and $\chi\in Q(\Pi)$, so that Π and χ are fixed points and the value of game is

$$E(\Pi,\chi)=\sum_{i=1}^{n}b_ip_i+d=\sum_{j=1}^{n}c_jq_j+d$$

***Theorem 2*:** Let the interior of Q contains a fixed point, then the first critical point belongs to P and the only fixed point. Similarly, if the interior of P contains a fixed point, then the second critical point belongs to Q and is the only fixed point of Q.

***Proof*:** Let $\beta=(v_1,\cdots,v_n)'$ be a fixed point of Q which lies in the interior of Q. Let $\alpha=(u_1,\cdots,u_n)'$ be a fixed point of P and let $\Pi=(p_1,\cdots,p_n)'$ be the first critical point, we wish to show that $\alpha=\Pi$. Suppose if possible, $\alpha\neq\Pi$.

Then $u_i = \int_0^1 r_i(x)dF(x), \qquad i = 1,2,\cdots,n$

$$v_j = \int_0^1 s_j(y)dG(y), \quad j = 1,2,\cdots,n$$

and $E(F,G) \equiv E(\alpha,\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}u_iv_j + \sum_{i=1}^{n} b_iu_i + \sum_{j=1}^{n} c_jv_j + d$

$$= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}u_i + c_j\right)v_j + \sum_{i=1}^{n} b_iu_i + d$$

Alternatively,

$$E(F,G) = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}v_i + b_i\right)u_i + \sum_{j=1}^{n} c_jv_j + d$$

since $\det\left[a_{ij}\right] \neq 0$, then the system of equations

$$\sum_{i=1}^{n} a_{ij}u_i + c_j = 0, \quad j = 1,2,\cdots,n \tag{1}$$

has unique solution, and so also the system of equation

$$\sum_{j=1}^{n} a_{ij}v_j + b_i = 0, \quad i = 1,2,\cdots,n \tag{2}$$

The unique solution of (1), say $\Pi = (p_1, p_2, \cdots, p_n)'$ is called the first critical point of the game (with respect to the gives representation) and the unique solution of (2), say $\chi = (q_1, q_2, \cdots, q_n)'$ is called the second critical point of the game (with respect to the given representation).

Note: The point Π may not belong to the *P*-space and the point χ may not belong to the *Q*-space. However, if $\Pi \in P$ and $\chi \in Q$, then the corresponding game problem can be solved easily by using the following two theorems.

Theorem 1: Let the payoff function of a separable game has a given canonical form

$$M(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}r_i(x)s_j(y) + \sum_{i=1}^{n} b_ir_i(x) + \sum_{j=1}^{n} c_js_j(y) + d, \ \det\left[a_{ij}\right] \neq 0$$

$$\text{Now, } \det\left[a_{ij}\right]=\begin{vmatrix}1 & -1 & 2\\ 2 & 3 & 1\\ 0 & 5 & -3\end{vmatrix}=-14+14=0$$

Now from $M(x,y)=xy-xe^{y}+2x\cos y+2e^{x}y+3e^{x}e^{y}+e^{x}\cos y+5\cos xe^{y}-3\cos x\cos y$

we have

$$M(x,y)=(x-2\cos x)\left(y+\frac{7}{5}\cos y\right)-(x-2\cos x)\left(e^{y}-\frac{3}{5}\cos y\right)+2\left(e^{x}+\cos x\right)$$

$$\left(y+\frac{3}{5}\cos y\right)+3\left(e^{x}+\cos x\right)\left(e^{y}-\frac{3}{5}\cos y\right)$$

i.e., $M(x,y)=r_1(x)s_1(y)-r_1(x)s_2(y)+2r_2(x)s_1(y)+3r_2(x)s_2(y)$

where $r_1(x)=x-2\cos x \qquad s_1(y)=y+\frac{7}{5}\cos y$

$r_2(x)=e^{x}+\cos x \qquad s_2(y)=e^{y}-\frac{3}{5}\cos y$

Hence

$$a_{11}=1,\ a_{22}=-1,\ b_1=b_2=0\quad a_{21}=2,\ a_{22}=3\quad c_1=c_2=0\quad d=0$$

$$\text{Therefore, } \det\left[a_{ij}\right]=\begin{vmatrix}1 & -1\\ 2 & 3\end{vmatrix}=5\neq 0$$

Thus this representation is canonical.

First critical point and second critical point

Let a separable function $M(x, y)$ has the following canonical representation:

$$M(x,y)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}r_i(x)s_j(y)+\sum_{i=1}^{n}b_i r_i(x)+\sum_{j=1}^{n}c_j s_j(x)+d$$

where $r_i(x),\ s_j(y)\ \ (i,j=1,2,\cdots,n)$ are continuous function on [0, 1] and $\det\left[a_{ij}\right]\neq 0$.

In this case, let F, a mixed strategy for the player-1 corresponds to $\alpha=\left(u_1,\cdots,u_n\right)'\in P$ and let G, a mixed strategy for the player-2 correspond to $\beta=\left(v_1,\cdots,v_n\right)'\in Q$ (Note that both $P,Q\subset E^{n}$).

Solution: We can write $M(x,y)=\sum_{i=1}^{2}\sum_{j=1}^{2}a_{ij}\,r_i(x)s_j(y)$

where $r_1(x)=\cos 4x \qquad s_1(y)=\cos 5y \qquad 0\le x\le 1$

$r_2(x)=\sin 4x \qquad s_2(y)=\sin 5y \qquad 0\le y\le 1$

and $a_{11}=3,\ a_{12}=5,\ a_{21}=1,\ a_{22}=1$

Thus $P^*=\left\{\rho=(r_1,r_2)'\ \ r_1=\cos 4t,\ r_2=\sin 4t,\ 0\le t\le 1\right\}$

and $Q^*=\left\{\sigma=(s_1,s_2)',\ \ s_1=\cos 5t,\ s_2=\sin 5t,\ 0\le t\le 1\right\}$

Canonical Representation of the payoff function in a separable game

Let for a separable game, the payoff function $M(x, y)$ be represented by

$$M(x,y)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\,r_i(x)s_j(y)+\sum_{i=1}^{n}b_i r_i(x)+\sum_{j=1}^{n}c_j s_j(y)+d$$

where

(i). $r_1(x),r_2(x),\cdots,r_n(x)$ and $s_1(y),\cdots,s_n(y)$ are continuous over [0, 1].

(ii). $\det\left[a_{ij}\right]\neq 0$.

Any representation of a separable function in this form is called a canonical representation of the function $M(x, y)$.

Note that every separable function has a canonical form. For if $M(x, y)$ is separable then it can always be written in the form $\sum_{j=1}^{n}r_j(x)s_j(y)$ so that $\det\left[a_{ij}\right]=1\neq 0$.

Example: $M(x,y)=xy-xe^y+2x\cos y+2e^x y+3e^x e^y+e^x\cos y+5\cos x e^y-3\cos x\cos y$

Let $r_1(x)=x \quad s_1(y)=y,\quad r_2(x)=e^x \quad s_2(y)=e^y,\quad r_3(x)=\cos x \quad s_3(y)=\cos y$

Then $a_{11}=1,\ a_{12}=-1,\ a_{13}=2,\quad b_1=b_2=b_3=0$

$a_{21}=2,\ a_{22}=3,\ a_{23}=1,\quad c_1=c_2=c_3=0$

$a_{31}=0,\ a_{32}=5,\ a_{33}=-3,\qquad d=0$

and $E(\alpha_1, \beta) = \min_{\eta \in Q} E(\alpha_1, \eta)$ (4)

Thus we have

$$E(\alpha, \beta) \le \max_{\xi \in P} E(\xi, \beta) \quad \text{[by the definition of maxima]}$$

$$= E(\alpha_1, \beta) \quad \text{[by (3)]}$$

$$= \min_{\eta \in Q} E(\alpha_1, \eta) \quad \text{[by 4]}$$

$$\le E(\alpha_1, \beta_1) \quad \text{[by the definition of minima]}$$

$$\le \max_{\xi} E(\xi, \beta_1) \quad \text{[by the definition of maxima]}$$

$$= E(\alpha, \beta_1) \quad \text{[by (1)]}$$

$$= \min_{\eta \in Q} E(\alpha, \eta) \quad \text{by (2)}$$

$$\le E(\alpha, \beta) \quad \text{[by the definition of minima]}$$

Since the first and last terms of this continued set of inequalities are equal. We therefore conclude that all the quantities involved are equal, and thus in particular, that

$$E(\alpha, \beta) = \max_{\xi \in P} E(\xi, \beta) = \min_{\eta \in Q} E(\alpha, \eta)$$

which means that $\alpha \in P(\beta)$ and $\beta \in Q(\alpha)$ i.e., α is an image point of β and β is an image point of α.

General procedure to solve a separable game

Step-1: Plot the curves P^* and Q^* and determine their convex hulls to give P and Q space.

Step-2: Find $Q(\alpha)$ for every $\alpha \in P$ and $P(\beta)$ for every $\beta \in Q$ i.e., find out all the image point of P and Q space.

Step-3: Using the results of Step-2 find the fixed points of P and Q.

Step-4: Express the fixed points as convex linear combination of points in P^* and Q^* respectively, then find distribution function in which these fixed points correspond to the distribution function will be the optimal strategies.

Example: $M(x, y) = 3\cos 4x \cos 5y + 5\cos 4x \sin 5y + \sin 4x \cos 5y + \sin 4x \sin 5y$

Proof: Let F_1 be a mixed strategy for the player-1 and let it corresponds to α. Now, α is a fixed point of P means that for some point $\beta \in Q$, we have $\beta \in Q(\alpha)$ and $\alpha \in P(\beta)$ i.e., β is an image point of α and α is an image point of β i.e.,

$$E(\alpha,\beta) = \min_{\eta \in Q} E(\alpha,\eta)$$

$$\text{and} \quad E(\alpha,\beta) = \max_{\xi \in P} E(\xi,\beta)$$

Let G_1 be a strategy (mixed) for the player-2 and let it corresponds to $\beta \in Q$. Then

$$E(F_1,G_1) \equiv E(\alpha,\beta) = \min_{\eta \in Q} E(\alpha,\eta) = \min_{G \in D} E(F_1,G)$$

$$\therefore E(F_1,G_1) \le E(F_1,G) \quad \forall G \in D$$

Again $E(F_1,G_1) \equiv E(\alpha,\beta) = \max_{\xi \in P} E(\xi,\beta) = \max_{F \in D} E(F,G_1) \ge E(F,G_1) \ \forall F \in D$

i.e., $E(F,G_1) \le E(F_1,G_1) \quad \forall F \in D$

Hence we obtain

$$E(F,G_1) \le E(F_1,G_1) \le E(F_1,G) \quad \forall F,G \in D.$$

Thus F_1 and G_1 are optimal strategies.

Theorem: If α is any fixed point of P and β is any fixed point of Q, then α is an image point of β, and β is an image point of α.

Proof: Since α is a fixed point of P, then $\exists$ a point β_1 of Q such that β_1 is an image point of α and α is an image point β_1 i.e., $\alpha \in P(\beta_1)$ and $\beta_1 \in Q(\alpha)$ i.e. such that

$$E(\alpha,\beta_1) = \max_{\xi \in P} E(\xi,\beta_1) \tag{1}$$

$$\text{and} \quad E(\alpha,\beta_1) = \min_{\eta \in Q} E(\alpha,\eta) \tag{2}$$

Similarly, since β is a fixed point of Q, then there exists a point $\alpha_1 \in P$ such that α_1 is an image point of β and β is an image point of α_1

i.e., $\alpha_1 \in P(\beta)$ and $\beta \in Q(\alpha_1)$

$$\text{i.e.,} \quad E(\alpha_1,\beta) = \max_{\xi \in P} E(\xi,\beta) \tag{3}$$

Here we can write $M(x,y)=\sum_{i=1}^{3}\sum_{j=1}^{3}a_{ij}\,r_i(x)s_j(y)$

where

$$\begin{aligned}
&r_1(x)=x, && s_1(y)=y, && a_{11}=0,\ a_{12}=0,\ a_{13}=1\\
&r_2(x)=\cos 2\pi x, && s_2(y)=\cos 2\pi y, && a_{21}=0,\ a_{22}=1,\ a_{23}=0\\
&r_3(x)=1, && s_3(y)=1, && a_{31}=1,\ a_{32}=0,\ a_{33}=0
\end{aligned}$$

Thus $P^{*}=\left\{\rho=(r_1,r_2,r_3)':r_1(t)=t,\ r_2(t)=\cos 2\pi t,\ r_3(t)=1,\quad 0\le t\le 1\right\}$

$$Q^{*}=\left\{\sigma=(s_1,s_2)':s_1(t)=t,s_2(t)=\cos 2\pi t,s_3(t)=1,\quad 0\le t\le 1\right\}$$

Image of $\alpha\in P$

Definition: The image of $\alpha\in P$ is defined as the set of points $\beta\in Q$ such that $E(\alpha,\beta)=\min_{\eta\in Q}E(\alpha,\eta)$ and we denote this image of α by $Q(\alpha)$ and $Q(\alpha)\subset Q$. Note that $Q(\alpha)$ is in general a set of points $\in Q$.

Image of $\beta\in Q$

Definition: Similarly, the image of $\beta\in Q$ is defined as the set of points $\alpha\in P$ such that $E(\alpha,\beta)=\max_{\xi\in P}E(\xi,\beta)$ and we denote this image of β by $P(\beta)$ and $P(\beta)\subset P$. Note that $P(\beta)$ in general is a sets of points $\in P$.

Fixed point

Definition: If $\beta\in Q$ is an image point of $\alpha\in P$ and $\alpha\in P$ is an image point of $\beta\in Q$, then α is called a fixed point of P-space and β is a fixed point of Q-space.

Theorem: If F_1 is any distribution function and $\alpha\in P$ corresponds to F_1, then F_1 is an optimal strategy for the player-1 iff α is a fixed point P. Similarly, G is an optimal strategy for the player-2 iff the corresponding point $\beta\in Q$ is a fixed point Q.

Since α_k corresponds to $F_k(x)$, $k=1,\cdots,p$, we have

$$u_i=\int_0^1 r_i(x)\,dF_k(x),\ i=1,\cdots,m$$

Let $\alpha=(u_1,\cdots,u_m)'$

Since $\alpha=\sum_{k=1}^{p}c_k\alpha_k$, then $u_i=\sum_{k=1}^{p}c_k u_i^{(k)},\ i=1,2,\cdots,m$

$$=\sum_{k=1}^{p}c_k\int_0^1 r_i(x)\,dF_k(x)$$

$$=\int_0^1 r_i(x)\,d\left(\sum_{k=1}^{p}c_k F_k(x)\right)=\int_0^1 r_i(x)\,dF(x)$$

Since $F\in D$, α corresponds to F and hence $\alpha\in P$.

The proof for Q-space is similar.

P' and Q' spaces

Let us define $P^*=\left\{\rho=(r_1,\cdots,r_m)':r_i(t),\ 0\le t\le 1\right\}$

$$\therefore P^*\in E^m$$

$$Q^*=\left\{\sigma=(s_1,\cdots,s_n)':s_j(t),\ 0\le t\le 1\right\}$$

$$\therefore Q^*\subset E^n$$

Properties of P' and Q' space

Property-1: $P^*\subset P$ and $Q^*\subset Q$

Property-2: Since $r_i(x)$ and $s_j(y)$ are continuous function defined on a closed set, P^* and Q^* are bounded, closed and connected sets.

Property-3: P space is the convex hull of P^*. Similarly, Q-space is the convex hull of Q^*.

Example:

Let the payoff function be $M(x,y)=\cos(2\pi x)\cos(2\pi y)+x+y,\ 0\le x,y\le 1$.

Thus it is immaterial whether the player-1 uses strategy $F_1(x)$ or $F_2(x)$. In this case we shall call the two strategies F_1 and F_2 to be equivalent (with respect to the game).

P-space and Q-space

Let $\alpha = (u_1, \cdots, u_m)'$, $\beta = (v_1, \cdots, v_n)'$, then $\alpha \in E^m, \beta \in E^n$.

Let $P = \left\{\alpha = (u_1, u_2, \cdots, u_m)'; u_i = \int_0^1 r_i(x)dF(x), i = 1, 2, \cdots, m; F \in D\right\}$.

Then obviously $P \subset E^m$. We also say that α and F correspond. Note that a given point α of P space, in general, corresponds to many different distribution functions.

Similarly, let $Q = \left\{\beta = (v_1, v_2, \cdots, v_n)'; v_j = \int_0^1 s_j(y)dG(y), \; j = 1, 2, \cdots, n; \; G \in D\right\}$

Then obviously $Q \subset E^n$. We also say that β and G correspond. Note that a given point β in the Q-space, in general, corresponds to many different distribution functions.

Note: If u corresponds to both $F(x)$ and $F_1(x)$, and v corresponds to both $G(y)$ and $G_2(y)$, then

$$E(F,G) = E(F_1,G) = E(F,G_1) = E(F_1,G_1)$$

$$= \sum_{i,j} a_{ij}\, u_i v_j \text{ and let us denote it by } E(\alpha,\beta)$$

Geometrical properties of P and Q-spaces

Theorem: Let $F_k(x)\,(k = 1, \cdots, p)$ be distribution function (i.e., strategies for the player-1) and let $F(x) = \sum_{k=1}^{r} c_k\, F_k(x)$ where $c_k \geq 0, k = 1, 2, \cdots, p$ and $\sum_{k=1}^{p} c_k = 1$. Then $F \in D$. Let α_k and F_k correspond $(k = 1, \cdots, p)$. Then $\alpha = \sum_{k=1}^{p} c_k\, \alpha_k$ corresponds to F so that $\alpha \in P$.

A similar result holds for the Q-space.

Proof: We have $M(x,y) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\, r_i(x) s_j(y)$

Let $\alpha^k = \left(u_1^{(k)}, \cdots, u_m^{(k)}\right)'$ $u = 1, \cdots, p$.

Expectation function in a separable game

A separable game is $\langle X,Y,M\rangle$ where $X = Y = [0, 1]$, $M(x, y)$ is separable. Hence a mixed strategy for the player-1 is a probability distribution function $F(x)$ and a mixed strategy for the player-2 is a probability distribution function $G(y)$ where $F \in D, G \in D$, where

$$\left\{F : 0 \le F \le 1,\ \int_0^1 dF(x) = 1\right\}$$

or, $D = \left\{G : 0 \le G \le 1,\ \int_0^1 dG(y) = 1\right\}$

If the player-1 chooses a mixed strategy $F(x)$ and player-2, a mixed strategy $G(y)$ $(F, G \in D)$, then the expectation of the player-1 is

$$E(F,G) = \int_0^1\int_0^1 M(x,y)\,dF(x)\,dG(y)$$

$$= \int_0^1\int_0^1 \left\{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\, r_i(x)\, s_j(y)\right\} dF(x)\,dG(y)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} \left\{\int_0^1 r_i(x)\,dF(x)\right\}\left\{\int_0^1 s_j(y)\,dG(y)\right\}$$

$$= \sum_{a_{ij}}^{m}\sum_{j=1}^{n} a_{ij}\, u_i\, v_j$$

where $u_i = \int_0^1 r_i(x)\,dF(x),\ \ i = 1,2,\cdots,m$

and $\quad v_j = \int_0^1 s_j(y)\,dG(y),\ \ j = 1,2,\cdots,n$

***Note* 1:** There may exist two different strategies $F_1(x), F_2(x)$, say, such that

$$\int_0^1 r_i(x)\,dF_1(x) = \int_0^1 r_i(x)\,dF_2(x) \qquad i = 1,2,\cdots,m$$

But

$$E(F_1,G) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\, u_i v_j = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\left(\int_0^1 r_i(x)\,dF_2(x)\right) v_j$$

$$= E(F_2,G)\ \ \forall G \in D$$

$$\max_F \left\{ \min_G E(F,G) \right\} \text{ and } \min_G \left\{ \max_F E(F,G) \right\}$$

exist and are equal.

The proof of the theorem is left to this chapter.

111.10. SEPARABLE GAME

Definition: A function $M(x, y)$ of two variables x, y is called separable, if there exist m continuous functions $r_i(x)$, $i = 1, 2, \ldots, m$, n continuous functions $s_j(y)$, $j = 1, 2, \ldots, n$ and mn constants a_{ij} $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$ such that

$$M(x,y) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}\, r_i(x)\, s_j(y)$$

Example: $M(x,y) = x \sin y + x \cos y + 2x^2$ is separable.

We can take,

$$\begin{array}{lll} r_1(x) = x & s_1(y) = \sin y & a_{11} = 1,\ a_{12} = 1,\ a_{13} = 0 \\ r_2(x) = x^2 & s_2(y) = \cos y & a_{21} = 0,\ a_{22} = 0, a_{23} = 2 \\ & s_3(y) = 1 & \end{array}$$

We can also take

$$\begin{array}{lll} r_1(x) = x & s_1(y) = \sin y + \cos y & a_{11} = 1,\ a_{12} = 0 \\ r_2(x) = 2x^2 & s_2(y) = 1 & a_{21} = 0,\ a_{22} = 1 \end{array}$$

Note:

(i) The representation $M(x,y) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}\, r_i(x)\, s_j(y)$ of a separable function is not unique.

(ii) If $M(x, y)$ is separable, we may also write

$$M(x,y) = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij}\, r_i(x) \right) s_j(y) = \sum_{j=1}^{n} t_j(x) s_j(y)$$

where $t_i(x)$ and $s_j(y)$ are continuous functions of x and y respectively.

Separable Game: A continuous game on the unit square whose payoff function is separable is called a separable game.

Proof of (i): Let F_0 be an optimal strategy. This implies that

$$\inf_G E(F_0,G) = \sup_F \inf_G E(F,G) = v$$

So that $E(F_0,G) \geq v$ for all $G \in D_Y$.

Let us choose G to be a pure strategy $\Theta(y-y_0)$, then we obtain

$$E(F_0,y_0) \geq v \text{ for all } y_0 \in Y$$

i.e., $E(F_0,y) \geq v$ for all $y \in Y$

***Sufficient*:** Let $E(F_0,y) \geq v$ for all $y \in Y$. (1)

Now, we have to show that F_0 is an optimal strategy for the player-1.

Now integrating both sides of (1) with respect to $G(y)$ over Y, we have

$$\int_Y E(F_0,y)dG(y) \geq v \int_Y dG(y)$$

or, $E(F_0,G) \geq v \quad \forall G \in D_Y$

Hence $\inf_G E(F_0,G) = v = \sup_F \left\{\inf_G E(F,G)\right\}$

i.e., the supremum of the function $\inf_G E(F_0,G)$ is attained at F_0. This implies that F_0 is optimal.

❑ 111.9. CONTINUOUS GAME

***Definition*:** An antagonistic game $\Gamma = \langle X,Y,M \rangle$ is called a game on the unit square if the sets of pure strategies for each of the players are the unit segments i.e., $X = Y = [0, 1]$.

Hence the mixed strategies for the players in a game on the unit square are probability distributions on the segment [0, 1] such that $\int_0^1 dF(x) = 1, \quad \int_0^1 dG(y) = 1$.

Here $D_X = D_Y = D$.

***Definition*:** A game on the unit square is called continuous if the payoff function is continuous.

Fundamental theorem for continuous game

Statement : For any continuous game on the unit square with payoff function $M(x, y)$, both players possess optimal strategies,

$$\sup_x\left\{\inf_y M(x,y)\right\} \le v \le \inf_y \sup_x M(x,y).$$

***Theorem 2*:** If player-1 possess a pure optimal strategy x_0 and player-2 possesses an arbitrary strategy (in general mixed), then

$$v = \max_x\left\{\min_y M(x,y)\right\} = \min_y M(x_0,y)$$

Similarly, if player-2 possesses a pure strategy y_0 and the player-1 possesses an arbitrary strategy, then

$$v = \min_y\left\{\max_x M(x,y)\right\} = \max_x M(x,y_0)$$

Student may easily prove this theorem.

***Theorem 3*:** Let u be a real number.

(i) If F_0 be a mixed strategy for the player-1, then

$$u \le E(F_0,y)\ \forall y \in Y \Rightarrow u \le v$$

(ii) If G_0 be a mixed strategy for the player-2, then

$$u \le E(x,G_0)\ \forall x \in X \Rightarrow u \le v.$$

***Proof*:** We have $u \le E(F_0,y)\ \forall y \in Y$

Now integrating both sides with respect to $G(y)$ over Y, we have

$$u\int_Y dG(y) \le \int_Y E(F_0,y)\,dG(y)$$

or, $u \le E(F_0,G)\quad \forall G \in D_Y$

Hence $u \le \inf_G E(F_0,G) \le \sup_F\left\{\inf_G E(F,G)\right\} = v$

$\therefore u \le v$

Proof of (2) is similar.

***Theorem 4*:**

(i) In order that the strategy F_0 for the player-1 be optimal, it is necessary and sufficient that

$$E(F_0,y) \ge v\ \forall y \in Y$$

(ii) In order that the strategy G_0 for the player-2 be optimal, it is necessary and sufficient that

$$E(x,G_0) \le v\ \forall x \in X.$$

Hence, $\sup\limits_{x\in X} E(x,G) = \sup\limits_{F\in D_X} E(F,G)$

Proof of the second result is similar.

Value of an infinite antagonistic game

Definition: If $\sup\limits_F \inf\limits_G E(F,G) = \inf\limits_G \sup\limits_F E(F,G)$, then the common value of these mixed extrema is called the value of the game $\langle X,Y,M\rangle$.

Some properties of the value of the game and optimal strategies

Let v be the value of the game $\langle X,Y,M\rangle$.

Theorem 1: $\sup\limits_x \inf\limits_y M(x,y) \le v \le \inf\limits_y \sup\limits_x M(x,y)$

$\left[\text{For Matrix Game: } \max\limits_i \min\limits_j a_{ij} \le u \le \min\limits_j \max\limits_i a_{ij}\right]$

Proof: We know that $\inf\limits_{y\in Y} E(F,y) = \inf\limits_{G\in D_Y} E(F,G)$ for any arbitrary mixed strategy F for the player-1.

Let us choose a simple strategy $F = \theta(x-x') = \begin{cases} 0 & \text{for } x < x' \\ 1 & \text{for } x > x' \end{cases}$

Then

$$\inf_{y\in Y} M(x',y) = E(x',G) \quad \forall x' \in X$$

i.e., changing x' to x, $\inf\limits_{y\in Y} M(x,y) = \inf\limits_{G\in D_Y} E(x,G)$

But $\sup\limits_x \left\{\inf\limits_y M(x,y)\right\} \le \sup\limits_F \left\{\inf\limits_y E(F,y)\right\}$ [as the set of mixed strategies D_X contains all pure strategies]

$= \sup\limits_F \left\{\inf\limits_G E(F,G)\right\} = v$

i.e., $\sup\limits_x \left\{\inf\limits_y M(x,y)\right\} \le v$ (1)

Similarly, we can prove that $v \le \inf\limits_y \sup\limits_x M(x,y)$. (2)

Hence from (1) and (2) we have

$$E(x,G^*) \le E(F^*,G^*) \le E(F^*,y)$$

$\left[\text{For Matrix Game: } E(i,\eta^*) \le E(\xi^*,\eta^*) \le E(\xi^*,j)\ i=1,2,\cdots,m,\quad j=1,2,\cdots,n\right]$

This proof is left to the reader.

Theorem 3: For any mixed strategy G for player-2.

$$\sup_{x\in X} E(x,G) = \sup_{F\in D_X} E(F,G)$$

where the supremum on the left side is taken over all the pure strategies for player-1 and on the right side over all this mixed strategies.

Similarly, for fixed strategy F for the player-1,

$$\inf_{y\in Y} E(F,y) = \inf_{G\in D_Y} E(F,G)$$

$$\left[\begin{array}{l}\text{For Matrix Game: } \max_i E(i,\eta) = \max_\xi (\xi,\eta) \text{ for some fixed } \eta \\ \qquad\qquad \text{and } \min_j E(\xi,j) = \min_\eta E(\xi,\eta) \text{ for some fixed } \xi\end{array}\right]$$

Proof: Since the set of mixed strategies contains all the pure strategies

$$\left(\text{e.g., } F(x) = \Theta(x-x_0) = \begin{cases}1 & x > x_0 \\ 0 & x < x_0\end{cases}\right)$$

we have $\sup_{x\in X} E(x,G) \le \sup_{F\in D_X} E(F,G)$

Now, let if possible, $\sup_{x\in X} E(x,G) < \sup_{F\in D_X} E(F,G)$

This implies that there exists $F_1 \in D_X$ such that for some $\varepsilon > 0$

$$\sup_{x\in X} E(x,G) < E(F_1,G) - \varepsilon$$

i.e., $E(x,G) < E(F_1,G) - \varepsilon \qquad \forall x \in X$

Now integrating both sides with respect to $F_1(x)$, we have

$$\int_X E(x,G)\,dF_1(x) < \{E(F_1,G) - \varepsilon\} \int_X d\,F_1(x)$$

i.e., $E(F_1,G) < E(F_1,G) - \varepsilon \qquad \left[\because \int_X dF_1(x) = 1\right]$

which is a contradiction

the strategy G. Similarly, $E(F,y_0)=\int_X M(x,y_0)dG(x)$ is the expected payoff to player-1 when he chooses the strategy $F(x)$ and player-2 chooses the simple strategy $\Theta(y-y_0)$.

(iii) Let $\langle X,Y,M\rangle$ be an infinite antagonistic game. Then this is strategically equivalent to the game $\langle D_X,D_Y,E\rangle$, where $E=E(F,G)$, $F\in D_X$, $G\in D_Y$.

Equilibrium situation

Definition: Let (F^*,G^*) be a situation in mixed strategies for the infinite antagonistic game $\langle X,Y,M\rangle$ or $\langle D_X,D_Y,E\rangle$. This situation is called an equilibrium situation in mixed strategies or a saddle point in mixed strategies if for any fixed strategies F, G for player-1 and player-2 respectively, the following inequality is satisfied.

$$E(F,G^*)\le E(F^*,G^*)\le E(F^*,G)\ \forall F\in D_X, G\in D_Y$$

For mixed strategies in infinite antagonistic games, we shall prove the following theorem.

Theorem 1: If $E(x,G)\le u\ \forall x\in X$ and a fixed G, then $E(F,G)\le u\ \forall F\in D_X$ [For matrix game: $F(i,\eta)\le u,\ i=1,2,\cdots,m$ for a fixed $\eta\in S_n$ implies that $E(\xi,\eta)\le u\ \forall\xi\in S_m$]

Proof: Integrating both sides of $E(x,G)\le u$ with respect to $F(x)$ over X so as to obtain $\int_X E(x,G)dF(x)\le u\int_X dF(x)$

i.e., $E(F,G)\le u \quad \left[\because \int_X dF(x)=1\right]$

Similarly $E(x,G)\ge u\ \forall x\in X$ implies $E(F,G)\ge u\ \forall F\in D_X$

and $E(F,y)\underset{<}{>}u\ \forall y\in Y$ implies $E(F,G)\underset{<}{>}u\ \forall G\in D_Y$

Theorem 2: In order that the situation (F^*,G^*) be an equilibrium situation in an infinite antagonistic game $\Gamma=\langle X,Y,M\rangle$ it is necessary and sufficient that for all $x\in X$ and $y\in Y$, the following inequalities are satisfied.

$$\Delta_Y = \left\{G(y) : 0 \le G(y) \le 1, \quad \int_Y dG(y) = 1\right\}$$

Let for a fixed y, the player 1 chooses $x \in X$ with probability distribution $F(x)$ i.e., the probability of choosing x where $x_i'' < x < x_i'$ is $F(x_i') - F(x_i'')$, then the expected payoff to the player 1 is $F(x,y)\{F(x_i') - F(x_i'')\}$. Hence when $x \in X$, the expected payoff to the player is (for a fixed y) is

$$\sum_{i=1}^{n} M(x,y)\{F(x_i) - F(x_{i-1})\}$$

where X is divided into n intervals with the help of the points $x_i, i = 0, 1, \cdots, n$. If we now make $n \to \infty$ in such a way that $\max_i |x_i - x_{i-1}| \to 0$, then this becomes the stieltjes integral with respect to $F(x)$ over the interval X

i.e., $E(F,y) = J(y) = \int_X M(x,y)\,dF(x)$

In a similar way, when the player-2 chooses y with probability distribution $G(y)$ then the expected payoff to the player-j is $\sum_{j=1}^{n'} J(y)\{G(y_j) - G(y_{j-1})\}$ where Y is divided into n' intervals with the help of the points $y_0, \cdots, y_{n'}$. Now making $n' \to \infty$ such a way that $\max_j |y_j - y_{j-1}| \to 0$, we obtain the expected payoff to the player-1 when his strategy is $F(x)$ and player-2's strategy is $G(y)$, is

$$E(F,G) = \int_Y \left(\int_X M(x,y)\,dF(x) \right) dG(y).$$

Note:

(i) If $M(x, y)$ is continuous function of x, y for $x \in X$, $y \in Y$, then

$$E(F,G) = \int_Y \left(\int_X M(x,y)\,dx \right) dy = \int_X \left(\int_Y M(x,y)\,dy \right) dx = \int_Y \int_X M(x,y)\,dx\,dy$$

(ii) When $F(x)$ is the Heaviside function $\Theta(x - x_0) = \begin{cases} 0 \text{ for } x < x_0 \\ 1 \text{ for } x > x_0 \end{cases}$

then $E(F, G)$ becomes $\int_Y M(x_0, y)\,dG(y)$ and we denote it by $E(x_0, G)$. Thus $E(x_0, G)$ is the expected payoff to player-1 when he chooses a simple strategy $\Theta(x - x_0)$ and player-2 chooses subject to

$$\left.\begin{array}{ll}\sum_{j=1}^{n} a_{ij}\bar{y}_j \le 1, & i=1,2,\cdots,m \\ \bar{y}_j \ge 0, & j=1,2,\cdots,n\end{array}\right\} \quad (8)$$

If $z'_{\max}$ is the optimal value and $\bar{\eta}^*$ is an optimal solution of LPP (8), then $v=\frac{1}{z'_{\max}}$ and

$$\eta^* = v\bar{\eta}^* = \frac{\bar{\eta}^*}{z'_{\max}}$$

Note that (8) is dual of (4) and vice versa.

❑ 111.8. Infinite Antagonistic Game

Definition: An infinite antagonistic game is a game in which at least one of the players possesses an infinite number of strategies.

Let X and Y be the set of strategies for the players 1 and 2 respectively and $M(x, y)$ $(x \in X, y \in Y)$ be the payoff function of the antagonistic game $\langle X, Y, M\rangle$. In the case of matrix game recall that X, Y are discrete sets of strategies (pure) for players 1 and 2 respectively. In that case, we have defined a mixed strategy for player-1 as a vector ξ having m (the number of pure strategies) components, each component denting the probability of choosing a pure strategy. Similarly, a mixed strategy η for player-2 was defined. We can extend this idea of mixed strategy when X and Y are infinite.

In this case, a mixed strategy for player-1 will be denoted by a probability distribution function $F(x)$ where $F(x'_i)-F(x''_i)$ denoted the probability of choosing x when $x''_i < x < x'_i$ similarly the probability function $G(y)$ will denote a mixed strategy for player-2. Note that $F(x)$ is nonnegative, non-decreasing function and $\int_X dF(x)=1$. Similarly $G(y)$ is so and $\left.\int_Y dG(y)=0\right]$. Here the integration is in the Stieltjes sense.

Let $\Delta_X = \left\{F(x): 0 \le F(x) \le 1,\ \int_X dF(x)=1\right\}$

If z_{min} is the optimal value and $\bar{\xi}^* = (x_1^*, \cdots, x_m^*)'$ is an optimal solution of the LPP(4), then the value of the game is $v = \frac{1}{z_{\min}}$ and an optimal strategy for the player-1 is $\xi^* = v\bar{\xi}^* = \frac{\bar{\xi}^*}{z_{\min}}$.

Again, we have, $v \geq E(\xi, \eta^*) \ \forall \xi \in S_m$

Hence $v \geq E(i, \eta^*)$, $i = 1, 2, \cdots, m$

Thus v is the smallest number w such that ($v \leq w$) for $\eta \in S_n$,

$$w \geq E(i, \eta) \quad i = 1, \cdots, m \tag{5}$$

η is an optimal strategy for the player-2 iff η satisfies (5) with w replaced by v. Now (5) can be written as

$$\sum_{j=1}^{n} a_{ij} y_j \leq w \quad i = 1, 2, \cdots, m \tag{6}$$

Also we have $\sum_{j=1}^{n} y_j = 1, \ y_j \geq 0, \ j = 1, 2, \cdots, n$ (7)

Since v is assumed to be positive, we can assume $w > 0$, so that from (6) and (7), we have

$$\sum_{j=1}^{n} a_{ij} \bar{y}_j \leq 1, \quad i = 1, 2, \cdots, m$$

$$\sum_{j=1}^{n} \bar{y}_j = \frac{1}{w}, \quad \bar{y}_j \geq 0, \quad j = 1, 2, \cdots, n$$

where $\bar{y}_j = y_j / w$.

For optimal strategy

$$v = \min_{\eta} w \text{ (} w \text{ can be regarded as function of } \eta \text{)}$$

or, $$\frac{1}{v} = \max \frac{1}{w} = \max \sum_{j=1}^{n} \bar{y}_j$$

Thus finding of v and η^* is equivalent to solving the LPP

$$\text{Maximize } z' = \sum_{j=1}^{n} \bar{y}_j$$

i.e., $v \le E(\xi^*, j) \quad \forall j = 1, 2, \cdots, n$

We can say that v is the largest number $u(v \ge u)$ such that for some $\xi \in S_m$

$$u \le E(\xi, j), \quad j = 1, 2, \cdots, n \tag{1}$$

ξ is an optimal strategy for the player 1 iff ξ satisfies (1) with u replaced by v. Now (1) can also be written as

$$\sum_{i=1}^{m} a_{ij} x_i \ge u \quad j = 1, 2, \cdots, n \tag{2}$$

Also, we have, $\sum_{i=1}^{m} x_i = 1, \quad x_i \ge 0, \quad i = 1, 2, \cdots, m$ (3)

Since v is assumed to be positive, u is also so, so that from (2) and (3), we have

$$\sum_{i=1}^{m} a_{ij} \frac{x_i}{u} \ge 1, \quad j = 1, 2, \cdots, n$$

and $\sum_{i=1}^{m} \frac{x_i}{u} = \frac{1}{u}, \quad \frac{x_i}{u} \ge 0, \quad i = 1, 2, \cdots, m$

Now putting $\bar{x}_i$ for $\frac{x_i}{u}$, we get

$$\sum_{i=1}^{m} a_{ij} \bar{x}_i \ge 1, \quad j = 1, 2, \cdots, n$$

and $\sum_{i=1}^{m} \bar{x}_i = \frac{1}{u}, \quad \bar{x}_i \ge 0, \quad i = 1, 2, \cdots, m$

For optimal strategy, $v = \max_{\xi} u$ (we can regard u as a function of ξ) or $\frac{1}{v} = \max_{\xi} \frac{1}{u} = \min \sum_{i=1}^{m} \bar{x}_i$.

Thus the finding of v and ξ^* is equivalent to solving of the following linear programming problem:

Minimize $z = \sum_{i=1}^{m} \bar{x}_i$

subject to

$$\left.\begin{array}{ll} \sum_{i=1}^{m} a_{ij} \bar{x}_i \ge 1, & j = 1, 2, \cdots, n \\ \bar{x}_i \ge 0, & i = 1, 2, \cdots, m \end{array}\right\} \tag{4}$$

Similarly, the strategy η_1 for the player-2 is said to dominate his strategy η_2 is dominated by η_1 or η_1 is superior to η_2) if for any pure strategy i for player-1,

$$E(i,\eta_1) \le E(i,\eta_2), \quad i = 1,2,\cdots,m$$

In particular, the pure strategy i_1 for player-1 dominates his pure strategy i_2 if for any j,

$$a_{i_1 j} \ge a_{i_2 j} \quad j = 1,\cdots,n$$

and the pure strategy j_1 for the player-2 dominates his pure strategy j_2 if for any i

$$a_{ij_1} \le a_{ij_2} \quad i = 1,2,\cdots,m$$

Note: If i_1 dominates i_2, then the elements in the i_2-th row of the matrix can be discarded. Similarly if j_1 dominates j_2, then the elements in the j_2th column can be discarded. By this process, the order of the pay off matrix can be reduced effectively.

An important result:

If a constant is added to all the elements of a matrix game, the optimal strategies remain unaltered and the value of the game is increased by this constant.

This result can easily be proved.

Symmetric game

A matrix game with a skew symmetric matrix is called a symmetric game, obviously $m = n$, $a_{ij} = - a_{ji}$.

Theorem: In a symmetric game, the set of all optimal strategies of the player-1 and 2 are the same and the value of the game is zero.

Matrix Game and L.P.P.

Theorem: A matrix game is equivalent to a certain LPP.

Proof: Let $A = \left[a_{ij}\right]_{m\times n}$ be the pay off matrix and v be the value of the game. We may assume that $v >$ 0. If not, then a suitable constant may be added to each element of A so as to make positive. This change does not affect the optimal strategies for two players, only the value of the game is increased by this constant.

Let ξ^*, η^* be the optimal strategies and $v(>0)$ be the value of the game. Then we know that

$$v \le E(\xi^*,\eta) \quad \forall \eta \in S_n$$

Theorem-5:

(i). If for a strategy ξ_0 of the player-1, $v \le E(\xi_0, j)$, $\forall j = 1,2,\cdots,n$, then ξ_0 is an optimal strategy for player-1.

(ii). If for a strategy η_0 of the player-2, $v \ge E(i, \eta_0)$, $\forall i = 1,2,\cdots,m$, then η_0 is an optimal strategy for the player-2.

Proof: Let η^* be an optimal strategy for the player-2, then $E(\xi, \eta^*) \le v$, $\forall \xi \in S_m$ putting $\xi = e_i^m$; we have $E(i, \eta^*) \le v$ $\forall i = 1,2,\cdots,m$.

But $v \le E(\xi_0, j)$ (given). Hence $E(i, \eta^*) \le v \le E(\xi_0, j)$ $\forall i = 1,2,\cdots,m,$ $\forall j = 1,2,\cdots,n$.

By Theorem 2, it follows that ξ_0 is an optimal strategy for the player-1.

Proof of the second part is similar.

Theorem-6:

(i). If $v \le \min\limits_j E(\xi_0, j)$ for a fixed ξ_0, ξ_0 is an optimal strategy for the player-1.

(ii). If $v \ge \max\limits_i E(i, \eta_0)$ for a fixed η_0, η_0 is an optimal strategy for the player-2.

Proof: (i) $v \le \min\limits_j E(\xi_0, j) \le E(\xi_0, j)$ $\forall j = 1,2,\cdots,n$.

Then by Theorem-5, this theorem can be proved.

Corollary:

(i). A necessary and sufficient conditions for the optimality of strategy ξ_0 for the player-1 is $v = \min\limits_j E(\xi_0, j)$

(ii). Similar condition for the player-2 is $v = \max\limits_i E(i, \eta_0)$

❑ 111.7. DOMINANCE OF STRATEGIES

Definition: The strategy ξ_1 for player-1 is said to be dominate his strategy ξ_2 (i.e. ξ_2 is dominated by ξ_1 or ξ_1 is superior to ξ_2) if for any pure strategy j for player-2,

$$E(\xi_1, j) \ge E(\xi_2, j), \quad j = 1,2,\cdots,n$$

Theorem-3: If $\max_i E(i,\eta_0) \le \min_j E(\xi_0,j)$ is valid for some fixed ξ_0 and η_0. Then ξ_0,η_0 are optimal strategies and the inequality becomes equality.

Proof: we have $E(i,\eta_0) \le \max E(i,\eta_0) \quad \forall i = 1,2,\cdots,m$ [by the definition of maximum]

$$\le \min_j E(\xi_0,j) \text{ [by the given condition]}$$

$$\le E(\xi_0,j) \quad \forall j = 1,2,\cdots,n \text{ [by the definition of minimum]}$$

$$\forall i = 1,2,\cdots,m \text{ and } j = 1,2,\cdots,n$$

Let $u_1 = \max_i E(i,\eta_0)$.

Then $E(i,\eta_0) \le u_1 \le E(\xi_0,j) \quad \forall i = 1,2,\cdots,m$ and $j = 1,2,\cdots,n$.

Hence by Theorem-2, u_1 is the value of the game and ξ_0,η_0 are the optimal strategies.

Again let $u_2 = \min_j E(\xi_0,j)$

Now by the definition of minima, we have

$$E(\xi_0,j) \ge \min_j E(\xi_0,j) \quad \forall j = 1,2,\cdots,n$$

$$\ge \max_i E(i,\eta_0) \text{ [by the given condition]}$$

$$\ge E(i,\eta_0) \quad \forall i = 1,2,\cdots,m \text{ [by the definition of maxima]}$$

$$\therefore E(\xi_0,j) \ge \min_j E(\xi_0,j) \ge E(i,\eta_0)$$

$$\text{i.e., } E(i,\eta_0) \le \min_j E(\xi_0,j) \le E(\xi_0,j)$$

$$\text{or, } E(i,\eta_0) \le u_2 \le E(\xi_0,j)$$

Hence by Theorem-2, u_2 is the value of the game and ξ_0,η_0 are the optimal strategies. But the value of the game is unique. Hence $u_1 = u_2$ i.e. the inequality becomes an equality.

Theorem-4: If $E(i,\eta_0) \le E(\xi_0,j) \quad \forall i = 1,2,\cdots,m, \; j = 1,2,\cdots,n$ then ξ_0,η_0 are the optimal strategies.

Proof: This theorem can be proved from Theorem-3.

$$\therefore\ u\sum_{j=1}^{n} y_j^* \le \sum_{j=1}^{n} E(\xi, j) y_j^*$$

i.e., $u \le E(\xi, \eta^*) \le E(\xi^*, \eta^*)$ where ξ^* is an optimal strategy for player-1.

i.e., $u \le v$

Proof of the second part is similar.

Theorem-2: Let u be a real number such that $E(i,\eta_0) \le u \le E(\xi_0, j)$ for $i = 1,2,\cdots,m$, $j = 1,2,\cdots,n$ and some particular ξ_0, η_0. Then u is the value of the game and ξ_0, η_0 are the optimal strategies for the players 1 and 2 respectively.

Proof: Given that $E(i,\eta_0) \le u$, $(i = 1,2,\cdots,m)$ and some particular η_0

$$\therefore \sum_i E(i,\eta_0) x_i^0 \le u \sum_i x_i^0$$

$$\text{or, } E(\xi_0,\eta_0) \le u \qquad (1)$$

Again $u \ge E(\xi_0, j)$, $j = 1,2,\cdots,n$ gives similarly

$$u \ge E(\xi_0,\eta_0) \qquad (2)$$

Hence (1) and (2) implies that $u = E(\xi_0,\eta_0)$

Again we have $E(i,\eta_0) \le u = E(\xi_0,\eta_0)$, $i = 1,2,\cdots,m$

$$\text{so that } \sum_i E(i,\eta_0) x_i \le E(\xi_0,\eta_0) \sum_i x_i$$

$$\text{or, } E(\xi,\eta_0) \le E(\xi_0,\eta_0) \quad \forall \xi \in S_m \qquad (3)$$

Again $u = E(\xi_0,\eta_0) \le E(\xi_0, j)$, $j = 1,2,\cdots,n$,

$$\text{so that } E(\xi_0,\eta_0) \sum_i y_i \le \sum_j E(\xi_0, j) y_j$$

$$\text{or, } E(\xi_0,\eta_0) \le E(\xi_0,\eta) \quad \forall \eta \in S_n \qquad (4)$$

Hence from (3) and (4), we have

$$E(\xi,\eta_0) \le E(\xi_0,\eta_0) \le E(\xi_0,\eta) \quad \forall \xi \in S_m, \eta \in S_n$$

Hence by the definition of optimal strategies, ξ_0, η_0 are optimal strategies and $E(\xi_0,\eta_0) = u$ is the value of the game.

This gives $a_{11}y^* + a_{12}(1-y^*) - a_{21}y^* - a_{22}(1-y^*) = 0$

or, $(a_{11} - a_{12} - a_{21} + a_{22})y^* = a_{22} - a_{12}$

or, $y^* = \dfrac{a_{22} - a_{12}}{a_{11} - a_{12} - a_{21} + a_{22}}$ provided that $a_{11} - a_{12} - a_{21} + a_{22} \neq 0$

Similarly, from the second part of the inequality, it is seen that $f(x^*, y)$ regard as a function of *y* has a minimum at y^* i.e.,

$$\left.\frac{\partial f}{\partial y}\right|_{(x^*,y^*)} = 0$$

This gives $a_{11}x^* - a_{12}x^* + a_{21}(1-x^*) - a_{22}(1-x^*) = 0$

or, $x^* = \dfrac{a_{22} - a_{21}}{a_{11} - a_{12} - a_{21} + a_{22}}$ provided that $a_{11} - a_{12} - a_{21} + a_{22} \neq 0$.

Now $v^* = f(x^*, y^*)$

$$= a_{11}x^*y^* + a_{12}x^*(1-y^*) + a_{21}(1-x^*)y^* + a_{22}(1-x^*)(1-y^*)$$

$$= \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} - a_{12} - a_{21} + a_{22}}$$

It can be proved that $a_{11} - a_{12} - a_{21} + a_{22} = 0$ implies that *A* has a saddle point. So this case will not arise.

Some properties of optimal strategies in a matrix game

Let $A = [a_{ij}]_{m \times n}$ be a payoff matrix of a game whose value is *v*.

Theorem-1: Let *u* be an arbitrary real number. Then

(i). $u \le E(\xi, j)$, $j = 1, 2, \cdots, n$ for some fixed ξ implies $u \le v$.

(ii). $u \ge E(i, \eta)$, $i = 1, 2, \cdots, m$ for some fixed η implies $u \ge v$.

Proof: Let η^* be an optimal strategy for the player-2, Now we have $u \le E(\xi, j)$ $j = 1, \cdots, n$ for some fixed ξ.

$$v=\max_i\left(\min_j a_{ij}\right)=\min_j a_{i^*j}$$

(ii). If player-2 possesses a pure optimal strategy j^*, then

$$v=\min_j\left(\max_i a_{ij}\right)=\max_i a_{ij^*}$$

Proof: we know that $v=\max_{\xi}\min_j E(\xi,j)$

$$=\min_j E\left(e_{i^*}^m,j\right) \text{ as } \xi^*=e_{i^*}^m \text{ is optimal.}$$

Proof of the second part is similar.

Solution of Matrix Game

Let us consider a 2 × 2 Matrix game whose payoff matrix A is given by $A=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

If A has a saddle point, solution is obvious.

Let A have no saddle point. Let the player-1 has the strategy $\xi=(x_1,x_2)'\equiv(x,1-x)(0\le x\le 1)$ and the player-2 has the strategy $\eta=(y,1-y)'(0\le y\le 1)$.

$$\text{Then } E(\xi,\eta)=\sum_{i=1}^{2}\sum_{j=1}^{2}a_{ij}x_iy_j$$

$$=a_{11}xy+a_{12}x(1-y)+a_{21}(1-x)y+a_{22}(1-x)(1-y)$$

$$=f(x,y)\text{, say}$$

If $\xi^*=(x^*,1-x^*)'$, $\eta^*=(y^*,1-y^*)$ be optimal strategies, then from

$E(\xi,\eta^*)\le E(\xi^*,\eta^*)\le E(\xi^*,\eta)\ \ \forall\xi\in S_2,\ \eta\in S_2$

we have $f(x,y^*)\le f(x^*,y^*)\le f(x^*,y)\ \forall x\in(0,1),\ y\in(0,1)$.

From the first part of the inequality, we set that $f(x,y^*)$ regarded as a function of x has a maximum at x^* i.e.,

$$\left.\frac{\partial f}{\partial x}\right|_{(x^*,y^*)}=0$$

Hence $E(\xi, j_0) \le \min_{\eta} E(\xi, \eta)$ (1)

Again since $\min_{\eta} E(\xi,\eta) \le E(\xi,\eta)\ \forall \eta \in S_n$, then putting $\eta = e^n_{j_0}$, we have

$$\min_{\eta} E(\xi,\eta) \le E(\xi, e^n_{j_0}) = E(\xi, j_0) \quad (2)$$

By (1) and (2), we have

$$\min_{\eta} E(\xi,\eta) = E(\xi, j_0) = \min_{j_0} E(\xi, j)$$

This is valid for any $\xi \in S_m$ and hence taking the maximum of both sides with respect to ξ, we have

$$\max_{\xi}\left\{\min_{\eta} E(\xi,\eta)\right\} = \max_{\xi}\left\{\min_{j} E(\xi, j)\right\}$$

where the outer maxima on both sides are attained for the same value of ξ. It is known, however, that on the left side the outer extremum is attained at the optimal strategies of the player-1. Therefore the same strategies yield the maximum in the right hand side. Thus the first part of the theorem is proved.

Proof of the second part is similar.

Theorem-2: $\max_{i}\left\{\min_{j} a_{ij}\right\} \le v \le \min_{j}\left\{\max_{i} a_{ij}\right\}$.

Proof: By the above theorem, we have $v = \max_{\xi}\left\{\min_{j} E(\xi, j)\right\}\ \forall \xi \in S_m$

But $\max_{\xi}\left\{\min_{j} E(\xi, j)\right\} \ge \min_{j} E(\xi, j)\ \ \forall \xi \in S_m$

$\therefore v \ge \min_{j} E(\xi, j)\ \ \forall \xi \in S_m$

Putting $\xi = e^m_i$ we have $v \ge \min_{j} E(e^m_i, j) = \min_{j} E(i, j) = \min_{j} a_{ij}$ i.e., $v \ge \min_{j} a_{ij}$

The left side v is independent of i so that taking maximum with respect to i, we obtain

$$v \ge \max_{i}\left\{\min_{j} a_{ij}\right\}$$

Proof of the second part is similar.

Theorem-3:

(i). If player-1 possesses a pure optimal strategy i^*, then

Note-2: $\min_{\xi} E(\xi,\eta) = E(\xi,\eta^*)$ $\therefore \max_{\xi}\left\{\min_{\eta} E(\xi,\eta)\right\} = \max_{\xi} E(\xi,\eta^*) = E(\xi^*,\eta^*)$

and $\max_{\xi} E(\xi,\eta) = E(\xi^*,\eta)$ $\therefore \min_{\eta}\left\{\max_{\xi} E(\xi,\eta)\right\} = \min_{\eta} E(\xi,\eta) = E(\xi^*,\eta^*)$

Notations

Expected payoff to the player-1 under (ξ,η) is $E(\xi,\eta) = \sum_{i,j} a_{ij}\, x_i\, y_j$

Expected payoff to the player-1 under (ξ, e_j^n) or simply (ξ, j) is

$$E(\xi, e_j^n) = E(\xi, j) = \sum_i a_{ij} x_i$$

Expected payoff to the player-1 under (e_i^m, η) or simply (i,η) is

$$E(e_j^m, \eta) = E(i,\eta) = \sum_j a_{ij} y_j$$

Thus $E(\xi,\eta) = \sum_i E(i,\eta) = \sum_j E(\xi, j)$

Also $E(e_i^m, e_j^n) = E(i,j) = a_{ij}$

Some properties of the value of a matrix game

The matrix game is characterized by the payoff matrix $A = \left[a_{ij}\right]_{m\times n}$ and v be its value.

Theorem: $v = \max_{\xi}\left\{\min_{j} E(\xi, j)\right\} = \min_{\eta}\left\{\max_{i} E(i,\eta)\right\}$ and the outer extrema are attained at optimal strategies of players.

Proof: We have by definition, $v = \max_{\xi}\left\{\min_{\eta} E(\xi,\eta)\right\}$.

For a fixed ξ, let $E(\xi, j_0) = \min_j E(\xi, j)$, then

$$E(\xi, j_0) \le E(\xi, j), \quad j = 1,2,\cdots,n$$

or, $\sum_j E(\xi, j_0) y_j \le \sum_j E(\xi, j) y_j$

i.e. $E(\xi, j_0) \le E(\xi,\eta)$ for a fixed ξ and all $\eta \in S_n$.

Definition: A mixed extension of a matrix game is an antagonistic game denoted by $\langle S_m, S_n, E\rangle$ in which the set of strategies for the players is the set of their mixed strategies in the original game and the payoff function for the player-1 is defined by

$$E(\xi,\eta)=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\,x_i\,y_j$$

Recalling the definition of saddle point of an antagonistic game, we observe that the situation (ξ^*,η^*) in a mixed extension of a matrix game will be a saddle point (i.e., an equilibrium situation) provided for any $\xi\in S_m$ and $\eta\in S_n$ the following double inequality is satisfied.

$$E(\xi,\eta^*)\le E(\xi^*,\eta^*)\le E(\xi^*,\eta)\quad \forall\xi\in S_m,\,\eta\in S_n.$$

Fundamental Theorem of matrix game

The quantities $\max\limits_{\xi\in S_m}\left\{\min\limits_{\eta\in S_n}E(\xi,\eta)\right\}$ and $\min\limits_{\eta\in S_n}\left\{\max\limits_{\xi\in S_m}E(\xi,\eta)\right\}$ exist and are equal.

The students have proved this theorem at UG level.

Value of a Matrix Game

The common value of $\max\limits_{\xi}\left\{\min\limits_{\eta}E(\xi,\eta)\right\}$ and $\min\limits_{\eta}\left\{\max\limits_{\xi}E(\xi,\eta)\right\}$ is called the value of the matrix game with payoff matrix $A=\left[a_{ij}\right]$ and denoted by $v(A)$ or simply v.

Definition: Equilibrium strategies of players in a matrix game are called their optimal strategies.

Definition: The solution of a matrix game is the process of determining the value of the game and the pairs of optimal strategies.

Note-1: Thus if (ξ^*,η^*) is an equilibrium situation in mixed strategies of the game $\langle S_m,S_n,E\rangle$, then ξ^*,η^* are the optimal strategies for the players 1 and 2 respectively in the matrix game with payoff matrix $A=\left[a_{ij}\right]_{m\times n}$. Hence ξ^*,η^* are optimal strategies for the players 1 and 2 respectively iff

$$E(\xi,\eta^*)\le E(\xi^*,\eta^*)\le E(\xi^*,\eta)\quad \forall\xi\in S_m,\,\eta\in S_n$$

(or probabilities) with which the player-1 chooses the strategies (pure) $1,2,\cdots,m$. Note that $e_i^m-(0,\cdots,0,1,0,\cdots,0)'$, $i=1,2,\cdots,m$ represent the pure strategy of the player-1 and $\xi=\sum_{i=1}^{m}e_i^m x_i'$.

Similarly, a mixed strategy for the player-2 is denoted by $\eta=(y_1,y_2,\cdots,y_n)'$ where $y_j\ge 0,\ j=1,2,\cdots,n$ and $\sum_{j=1}^{n}y_j=1$. Note that $e_j^n=(0,0,\cdots,0,1,0,\cdots,0)'$ represent the pure strategy of the player-2 and $\eta=\sum_{j=1}^{n}e_j^m y_j$.

Let $S_m=\left\{\xi/x_i\ge 0,\ \sum_{i=1}^{m}x_i=1\right\}$. then $S_m\in E_m$, this S_m is the space of mixed strategies for the player-1. Note that a pure strategy is a particular case of a mixed strategy.

Similarly, let $S_n=\left\{\eta/y_j\ge 0,\ \sum_{j=1}^{n}y_j=1\right\}$, then $S_n\in E_n$ thus S_n is the space of mixed strategies for the player-2.

Mixed extension of a matrix game without a saddle point

Let the players 1 and 2 choose their mixed strategies as $\xi=(x_1,x_2,\cdots,x_m)'$, $\eta=(y_1,y_2,\cdots,y_n)'$ independently of one another in a matrix game with payoff matrix $A=\left[a_{ij}\right]_{m\times n}$.

Definition: A pair (ξ,η) of mixed strategies for the players in a matrix game is called a situation in mixed strategies.

In a situation (ξ,η) of mixed strategies each usual situation (i,j) in pure strategies becomes a random event occurring with probabilities x_iy_i. Since in the situation (i,j), player-1 receives a payoff a_{ij}, the mathematical expectation of his payoff under (ξ,η) is equal to

$$E(\xi,\eta)=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\,x_i\,y_j$$

We thus arrive a new game which can be described as follows:

Strategies of Player B

$$\text{Strategies of Player } A \quad \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m \end{array}\overset{\begin{array}{cccc} 1 & 2 & \cdots & n \end{array}}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}$$

$X=\{1,2,\cdots,m\}$ $Y=\{1,2,\cdots,n\}$
$M(x,y)=M(i,j)=a_{ij}$
where (i,j) *be the situation of the matrix game.*

the intersection of the rows and columns correspond to the situations of the game. If we place in each cell the payoff of the first player in the appropriate situation, we then obtain the description of the game in the form of a certain matrix. This matrix is called the matrix of the game or the payoff matrix.

Thus (i^*,j^*) is an equilibrium situation (or a saddle point) in a matrix game if

$$a_{ij^*} \le a_{i^*j^*} \le a_{i^*j} \quad \forall i=1,2,\cdots,m \text{ and } \forall j=1,2,\cdots,n$$

Theorem: If a matrix game possesses a saddle point, it is necessary and sufficient that

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij}$$

The students have proved this theorem at UG level.

Thus the solution of a matrix game is immediate if the payoff matrix has a saddle point. In this case, i^*, j^* are called an optimal strategy of player-1 and player-2, respectively and $a_{i^*j^*}$ is called the value of the game.

However, not all matrix possess a saddle point and so the determination of the optimal solution is not an easy task.

Mixed strategy

If the payoff matrix does not have a saddle point then there exists no equilibrium situation of the form (i^*,j^*) i.e., an optimal strategy of the players cannot be single (pure) strategy.

Let us consider the matrix game whose payoff matrix is $A=\left[a_{ij}\right]_{m\times n}$. By a mixed strategy for the player-1, we shall mean as ordered m-tuple $(x_1,\cdots,x_m)$ or the vector $\xi=(x_1,\cdots,x_m)'$ where $x_i \ge 0$ $i=1, 2, \ldots, m$ and $\sum_{i=1}^{m} x_i = 1$. The components $x_1, x_2,\cdots,x_m$ may be thought of as the relative frequencies players 1 and 2 respectively, and *M* is the payoff function for players 1 and 2 respectively, and *M* is the pay off function for player-1 which is a real valued function such that $M : X \times Y \to R$.

Note: It is obvious that each non-cooperative two-person constant zero sum game is strategically equivalent to a certain antagonistic game.

Equilibrium situation for an antagonistic game

Let $\langle X, Y, M \rangle$ be an antagonistic game, $H_1 = M$, $H_2 = -M$

Hence $s_1 \equiv x \in X$, $s_2 \equiv y \in Y$, $s \equiv (x, y) \in X \times Y$

If (x, y) is an equilibrium situation, then

$$H_1(x,y) \geq H_1(x',y) \; \forall x' \in X \text{ and } H_2(x,y) \geq H_2(x,y') \; \forall y' \in Y$$

$$\text{i.e., } M(x,y) \geq M(x',y) \; \forall x' \in X \text{ and } -M(x,y) \geq -M(x,y') \; \forall y' \in Y$$

$$\text{i.e., } M(x',y) \leq M(x,y) \leq M(x,y') \; \forall x' \in X,\; y' \in Y$$

Changing the notation (x, y) to (x^*, y^*) and (x', y') to (x, y) we see that (x^*, y^*) will be an equilibrium situation for an antagonistic game $\langle X, Y, M \rangle$ if

$$M(x, y^*) \leq M(x^*, y^*) \leq M(x^*, y) \; \forall x \in X,\; y \in Y$$

The point (x^*, y^*) is called the saddle point of $M(x, y)$,

❑ 111.6. MATRIX GAMES OR RECTANGULAR GAMES

Antagonistic games in which each player possess a finite number of strategies are called matrix games or rectangular games.

The name is due to the fact that such games can be described as follows: Consider a rectangular array in which the rows correspond to the strategies for the first player and the columns correspond to the strategies of the second player and the cells in the array located at i.e., $H_i^{(1)}\left(s \,||\, s_i^0\right) \le k\, H_i^{(2)} + C_i$

or, $k\, H_i^{(2)}\left(s \,||\, s_i^0\right) + C_i \le H_i^{(2)}(s) + C_i$ as $H_i^{(1)}\left(s \,||\, s_i^0\right) = k\, H_i^{(2)}\left(s \,||\, s_i^0\right) + C_i$

or, $H_i^{(2)}\left(s \,||\, s_i^0\right) \le H_i^{(2)}(s)\ \ \forall i \in I$ and $s_i^0 \in S_i$ since $k > 0$

which implies s is an equilibrium situation in Γ_2.

Zero-sum-game

A non-cooperative game Γ is called a zero sum game if for each situation $s, \sum_{i \in I} H_i(s) = 0$

Theorem: Any non-cooperative constant sum is strategically equivalent to a certain zero-sum-game.

Proof: Let $\Gamma = \left\langle I, \{S_i\}_{i \in I}, \{H_i\}_{i \in I} \right\rangle$ be a constant sum game, then $\sum_{i \in I} H_i(s) = C$, C being a constant.

Now let us consider the game $\Gamma' = \left\langle I, \{S_i\}_{i \in I}, \{H_i'\}_{i \in I} \right\rangle$ where $H_i' = H_i - C_i$ and $\sum_{i \in I} C_i = C$. Then obviously $\Gamma \sim \Gamma'$ when $k = 1$.

But $\sum_I H_i'(s) = \sum_I H_i(s) - \sum_I C_i = C - C = 0$

So that Γ' is a zero sum game.

❑ 111.5. ANTAGONISTIC GAME

The game $\Gamma = \left\langle I, \{S_i\}_{i \in I}, \{H_i\}_{i \in I} \right\rangle$ is called antagonistic if there are only two players in the game (i.e., $I = \{1, 2\}$) and the values of the payoff function for these players in each situation in the same in absolute value but are of opposite sign.

Thus for an antagonistic game, $I = \{1, 2\}$ and $H_2(s) = -H_1(s)$.

Hence $H_1(s) + H_2(s) = 0$, $s = (s_1, s_2) \in S_1 \times S_2$. Thus an antagonistic game is also a two-person zero-sum game.

Thus to define an antagonistic game, it is sufficient to stipulate the payoff function of one of the players only. It is usually denoted by the triplet $\langle X, Y, M \rangle$ where X, Y are the sets of strategies for

Definition: The games Γ_1 and Γ_2 are called strategically equivalent if a positive number k and numbers C_i (for each of the players $i \in I$) exist such that for any situations

$$H_i^{(1)}(s) = k\,H_i^{(2)}(s) + C_i$$

We denote this fact by the symbol $\Gamma_1 \sim \Gamma_2$.

Properties of strategic equivalence

(i). Reflexivity: $\Gamma \sim \Gamma$, it can easily be proved choosing $k = 1$ and $C_i = 0$.

(ii). Symmetry: $\Gamma_1 \sim \Gamma_2 \Rightarrow \Gamma_2 \sim \Gamma_1$

Proof: Since $\Gamma_1 \sim \Gamma_2$, we can write $H_i^{(1)} = k\,H_i^{(2)}(s) + C_i \quad k > 0,\ \forall\ i \in I$

$$\text{or, } H_i^{(2)}(s) = \frac{1}{k} H_i^{(1)}(s) - \frac{C_i}{k}$$

$$= k' H_i^{(1)}(s) + C_i' \text{ where } k' = \frac{1}{k},\ C_i' = -\frac{C_i}{k}$$

Hence by definition $\Gamma_2 \sim \Gamma_1$.

(iii). Transitivity: If $\Gamma_1 \sim \Gamma_2$, $\Gamma_2 \sim \Gamma_3$ then $\Gamma_1 \sim \Gamma_3$.

It can easily be proved using the definition.

Note: The basic difference between two strategically equivalent games lies in the amount of initial capital possesses by the players and in the relative units in which the payoffs are measured. This is indicated by the coefficient k. It is therefore natural to suppose that the rational behaviour of players in distinct strategically equivalent games should be the same.

Theorem: Strategically equivalent games possess the same equilibrium situation.

Proof: Let $\Gamma_1 \sim \Gamma_2$ where $\Gamma_1 = \left\langle I, \{S_i\}_{i \in I}, \{H_i^{(1)}\}_{i \in I} \right\rangle$

$$\Gamma_2 = \left\langle I, \{S_i\}_{i \in I}, \{H_i^{(2)}\}_{i \in I} \right\rangle$$

and let s be an equilibrium situation in Γ_1. This means that $\forall i \in I$ and $s_i^0 \in S_i$, the inequality

$H_i^{(1)}(s \,\|\, s_i^0) \le H_i^{(1)}(s)$ where $s = (s_1, \cdots, s_{i-1}, s_i, s_{i+1}, \cdots, s_n)$, $s \,\|\, s_i^0 = (s_1, \cdots, s_{i-1}, s_i^0, s_{i+1}, \cdots, s_n)$

Now since $\Gamma_1 \sim \Gamma_2$, $H_i^{(1)} = k\,H_i^{(2)} + C_i\,(k > 0)\ \ \forall i \in I$ and $s_i \in S_i$

Admissible situation

A situation $s=(s_1,\cdots,s_n)$ in a game $\Gamma=\langle I,\{s_i\},\{H_i\}\rangle$ is called admissible situation for the player i if for any other strategy s_i' of this player, we have

$$H_i(s\|s_i')\leq H_i(s)$$

The term 'admissible' can thus be justified by the fact that if in situation s there exists a strategy s_i' for the player i such that $H_i(s\|s_i')>H_i(s)$, then the player i, knowing that the situation s will materialize, may choose the strategy s_i' at the last moment and as a result of this choice end up with a large payoff. In this sense the situation may be viewed as admissible for the player i.

Equilibrium situation

A situation s which is admissible for all player is called an equilibrium situation, i.e.,

$$H_i(s\|s_i')\leq H_i(s)$$

is satisfied for any player i and for any strategy $s_i'\in S_i$.

Equilibrium strategy

It is a strategy that appears in at least one equilibrium situation of the game.

Solution of the game

The process of determination of an equilibrium situation in a non-cooperative game is often referred to as the solution of the game.

Strategic equivalence of games

Let two non-cooperative games with the same set of players and the same set of strategies be given (i.e., the games differ only in their payoff functions) as

$$\Gamma_1=\left\langle I,\{S_i\}_{i\in I},\{H_i^{(1)}\}_{i\in I}\right\rangle$$

$$\Gamma_2=\left\langle I,\{S_i\}_{i\in I},\{H_i^{(2)}\}_{i\in I}\right\rangle$$

The games in which the action of the player are directed to maximize the gain of 'collectives' without subsequent subdivision of the gain among the players within the condition are called cooperative games.

Here we shall consider only non-cooperative games.

Let I denote the set of all players. We shall assume that I is finite. Let $I = \{1, 2, \cdots, n\}$ where n denotes the number of players in the game. Let each player $i \in I$ have at his disposed a certain set S_i of the available action which are called strategies. Each player has at least two distinct strategies.

The process of the game consists of each one of the players choosing a certain strategy $s_i \in S_i$. Thus as a result of each 'round' of the game, a system of strategies $(s_1, \cdots, s_n) \equiv s$ is put together. This system is called a situation. The set of all strategies is denoted by $S = S_1 \times S_2 \times \cdots \times S_n$. In each situation the players attain certain gains (payoff). The payoff of player i in situation s is denoted by $H_i(s)$. The function H_i defined on the set of all situation is called the payoff function of the player i. Thus

$H_i : S \to R$ where R is a set of real numbers.

Definition : A system denoted by $\Gamma = \langle I, \{S_i\}_{i\in I}, \{H_i\}_{i\in I} \rangle$ where I and $S_i\,(i \in I)$ are sets and H_i's are function defined on the set $S = \prod_{i=1}^{n} S_i$ taking on real values, is called a non-cooperative game.

Definition : A non-cooperative game $\Gamma = \langle I, \{S_i\}_{i\in I}, \{H_i\}_{i\in I} \rangle$ is called a constant sum game if there exists a constant C such that $\sum_{i\in I} H_i(s) = C$ for each and every situation $s \in S$.

Admissible situation and the equilibrium situation

Let $s = (s_1, \cdots, s_{i-1}, s_i, s_{i+1}, \cdots, s_n)$ be an arbitrary situation in the game Γ and let s_i be a strategy of the player i. We form a new situation that differs from the situation s only in that strategy s_i for player i which (i.e., s_i) is now replaced by s_i'. Denote this as $s \| s_i' = (s_1, \cdots, s_{i-1}, s_i', s_{i+1}, \cdots, s_n)$. If s_i and s_i' coincide, then obviously $s \| s_i' = s$.

A competitive situation is called a game if it has the following properties :

(i) The number of competitors (participants), called players is finite.

(ii) There is a conflict in interests between the participants.

(iii) Each of the participants has a finite/infinite set of possible courses of action.

(iv) The rules governing these choices are specified and known to the players, a play of the game results when each of the players chooses a single course of action from the list of courses available to him.

(v) The outcome of the game is affected by the choices made by all the players.

(vi) The outcome for all specific set of choices by all the players is known in advance and numerically defined.

(vii) Every play i.e., combination of courses of action determines an outcome (which may be money or point) which determines a set of payments (+ve, -ve or zero) one to each player.

❑ 111. 2. OBJECTIVES

After studying this module, the reader will be able to

- Understand how the optimal strategies are formulated in conflict and competitive situations.
- Learn different types of finite and infinite games and their solution procedure.

❑ 111.3. KEYWORDS

Finite game, Infinite game, Cooperative game, Non-cooperative game, zero-sum game, Rectangular game, Antagonistic game, continuous game, symmetric game, separable game, pure strategy, mixed strategy, optimal strategy, saddle point, pay-off matrix

❑ 111.4. COOPERATIVE AND NON-COOPERATIVE GAMES

The games in which the objective of each participant (called 'player') is to achieve the largest possible individual gain (called 'payoff') are called non-cooperative games.

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X Group-B

Module No. - 111

Advanced Optimization And Operational Research - II

....

MODULE STRUCTURE :

❑ 111.1. Introduction

In many practical problems, it is required to take the decision in a situation where there are two or more opposite parties with conflicting interests and the action of one depends upon the action which is taken by the opponent. Such a situation is termed as competitive situation. A great variety of competitive situation is commonly seen in everyday life e.g., in military battles, political campaign, elections, advertisement, etc.

(xi) The cost of a chemcial process is described by the functional $J = \int_0^1 \left(\frac{1}{2} y'^2 + y \right)$ where $y(0)=1$.

If $y(1)$ is not specified, using the transversatity conditions at $x=1$, determine the optimum trajectory and the corresponding value of J and compare with the earlier result.

(xii) Describe the Pontryagin's Maximum Principle and illustrate it with the help of an example.

8. References :

Following tents are suggested for further study --

(i) A.S. Gupta, Calculus of Variations with Applications, Prencitce Hall of India, New Delhi, 2005.

(ii) B.D. Craven, Control and optimization, Chapman & Hall, 1995.

(iii) W.Fleming & R. Rishel, Deterministic and Stochastic Optimal Control, Springer, 1975.

(iv) L. Hoacking, Optimal control : An Introduction to the theory with Applicatins, Oxford University Press, 1991.

(v) J. Macki & A. Strauss, Introduction to Optimal Control Theory, Springer, 1982.

--- 0 ---

(vi) Find x and y as functions of t, so that $\int_{t_0}^{t_1}\left[\frac{m}{2}\left(x^2+y^2\right)-mgy\right]dt$

May have stationary value. It may be assumed that x and y are given at to and t_1.

(vii) A system is described by $\dot{x}_1 = -2x_1 + u$ and the control $u(t)$ is to be chosen in order to minimize $\int_0^1 u_2 dt$.

Show that theoptimal control which transfer the system from $x_1(0)=1$ to $x_1(1)=1$ is given by $u^* = -\frac{4e^{2t}}{e^4-1}$

(viii) An electrochemical system is characterised by the ordinary differential equations $\frac{dx_1}{dt} = x_2$ and $\frac{dx_2}{dt} = x_2 = u$

Where u is the control variable chosen in such a way that the cost functional $\frac{1}{2}\int_0^{\alpha}\left(x_1^2+4u^2\right)dt$ is minimized.

Show that, if the boundary conditions satisfied by the state variabels are x_1=(0) =a, $x_2(0)=b$ whether a, b are constants and $x_1(R)0, x_2(R)0$ as to (R) a, the optimal choice for u is $u = -0.5\,x_1(t) - 0.414\,x_2(t)$

Illustrate the feed back control in a black diagram

(ix) An electrochemical system is governed by the equations $\frac{dx_1}{dt} = -x_1 + u$ and $\frac{dx_2}{dt} + = x_1$ where u is a control variable so chosen that the cost functional $J = \int_0^{\alpha}\left(x_2^2 + \frac{16}{3}u^2\right)dt$ is minimized.

Assumig the boundary conditions as in the previous problem, show that the optimal control u is given by $u(t) = -0.366x_1(t) - 0.443x_2(t)$

(x) In an inventory control production scheduting problem, the governing equation is $\frac{dI}{dt} = p$ where $I - I(t)$ is the invenvory level and $p = p(t)$ is the production rate remaining after the demand has been met. It is planned that over a fixed time interval)<t<T, I should be increased from its value I0 at $t = 0$ by decreasing p in such a way that the cost functional

$J = \int_0^T\left[(Q-I)^2 + \alpha^2 p^2\right]$ is minimized.

Here, Q and a_2 are positive constants and $Q >$ I0. Determine the optimal production rate pond the inventory level.

$$u^*(t) = \begin{cases} 1 \ if \ 0 \le t \le t^* \\ 0 \ if \ t^* \le t \le 1 \end{cases}$$

For the optimal switching time $t^* = T - 1$.

The solution of the production and consumption model shows that it is also an example of a Bang Bang control.

6. Conclusions

In this module, we have studied the theory of optimal control. Perforamnce indices for different types of control problems are discussed. We have used two techniques, namely -- the calculus of variations and the Pontryagins Masximum Principle to solve the contrll problems. Using the calculus of variations, the Eubr-Lagrange equations are deduced and those are solved to obtain the optimal trajectory. Pontryanins Maximum Principle asserts the existence of a function called the costate, which together with the optimal trajectory satisfy some equations which are analogus with the classical Hamiltoman dynamics. Bang Bang controls are illustrated through examples.

7. Exercise

(i) obtain the performance index for general optimal control systems. When a control problem is said to be of– a) Mayer type b) Bolza type c) Lagrange type?

(ii) Prove that $J = \int_{x_0}^{x_1} F(y, y', x)\,dx$ will be minimum only when $\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0$

(iii) Show that the functional $J = \int_{x_0}^{x_1} \frac{(1+y^2)}{y'^2}dx$ will be extremum of $y = \sinh(c_1x + c_2)$ where c_1, c_2 are arbitrary constants.

(iv) Find the least value of the integral $\int_A^B \frac{1}{y}\left[1 + \left(\frac{dy}{dx}\right)^2\right]^{\frac{1}{2}} dx$ where A is $(-1, 1)$ and B is $(1, 1)$.

(v) Obtain the necessary condition for the existence of a sationary value of the functional

$$J = \int_{x_0}^{x_1} F(y_1, y_2, ..., yn, y_1', y_2], ..., yn', x)dx)$$

$f(x,a) = xa, g = 0, r(x,a) = (1-a)x$; and therefore

$H(x,p,a) = f(x,a).p + r(x,a) = xap + (1-a)x = x + ax(p-1)$.

The dynamical system is

$$\dot{x}(t) = \frac{\delta H}{\delta p}(x,p,u) = ux$$

and $\dot{p}(t) = -\frac{\delta H}{\delta x}(x,p,u) = -1 - u(p-1)$.

The second equation of the above system is called the adjoint equation.

the term inalconditoinsgives $p(T) = \frac{\delta g}{\delta x}(x(T)) = 0$

Lastly, the maximality principle asserts

$$H(x, p, u) = \max_{ae[,1]} H(x, p, a)$$

$$= \max_{ae[0,1]} \{x(t) + ax(t)(p(t) - 1)\}.$$

Since $x(t) > 0$, at each time t the control value $u(t)$ must be selected to maximize $a(p(t)-1)$ for $0 \le a \le 1$.

Thus $u(t) = \begin{cases} 1 \text{ uf } p(t) > 1 \\ 0 \text{ if } p(t) \le 1 \end{cases}$

Hence if p is known, the optimal control can be designed at once. So next we must solve for the constate p. We have from the adjoint equation and the terminal condition.

$p(t) -1- u(t)(p(t) -1),) \le t \le T$ and $p-(T) = 0$

Since $p(T) = 0$, we deduce by continuity that $p(t)<1$ for t close to T, $t<T$. Thus $u(t)=0$ for such values of t. Therefore $p(t) = -1$ and consequently $p(t)=T-t$ for times t in this interval. So we have that $p(t)= T-t$ so long as $p(t)<1$ and this holds for $T-1<t<T$.

But for times $t<T-1$, with t near $T-1$, we have $u(t)=1$ and so $p(t)=-1-(p(t)-1)$

Since $p(T-1)=1$, we see that $p(t) = e^{T-1-t} > 1$ for all times $0 \le t \le T-1$. In particular there are no switches in the control over this time interval.

Rstoring the superscript * to our notation, we deduce that the optimal control is

$$J\int_0^T r\left(x(t),u(t)\right)dt + g\left(x(T)\right)$$

Where the terminal time $T > 0$, running payoff $r : Rn \times A \to R$ and terminal payoff $G : Rn \to R$ are given.

Then the problem is to find a control v^* such that J in minimized over all $U e A$.

Let us define the ontrol theory Hamiltonian by the function $H(x, p, a) f(xma).p + r(x, a)$, where $x, p, e\ Rn$ and aeA. Here the newly introduced variable vector p is called the costate. It may be considered as a sort of Lagranges multiplier.

Now we assume that U^* be the optimal control for the problem under consideration and x^* be the corresponding troaectory. Then the pontryagin's Maximum principle asserts the existence of a function $p^* : [0,T] \to R^w$ such that

$$\dot{x}^* = \Delta p H(x^*(t), p^*(t), U^*(t),$$

$$\dot{p}^*(t) = -\Delta p H(x^*(t), p^*(t), U^*(t)),$$

and $H(x^*(t), p^*(t), U^*(t)) = \max H(x^*(t), p^*(t), a)$, which

signifies the name of the theorem.

Also we have the terminal conditon

$$p^*(T) = \Delta g(x^*(T).$$

Furthermore, $H(x^*(t), p^*(t), u^*(t))$ is independent of t.

We now illustrate the pontryagin's Maximum principle by the following model for optimal consumption in simple economy.

Let $x(t)$ = output of the economy at time t and $u(t)$ = fraction of output reinvested at time t. We have the constraints $0 \le u(t) \le 1, i.e., A = [0,1]$. The economy evolves according to the dynamics

$$\left.\begin{aligned} x(t) &= kx(t)u(t), 0 \le t \le T \\ x(0) &= x^0 \end{aligned}\right\}$$

Where k is a constant, known as the growth factor. In this case we set the growth factor $k = 1$. We want to maximize the total consumption

$$J = \frac{\delta F}{\delta \dot{x}}$$

The problem is to find an optimal control u^*. We apply pontryagin's Maximum principle, and to simplify notation we will not write the superscripts * for the optimal control, trajectory, etc. Here we have $n = m = 1$, or, $\frac{w}{1-w}dw = d\theta$

Integrating we get,

$-w-\log(1-w) = q + c_2, c_2$ is a constant.

At $\theta = 0$, Therefore $c_2 = 0$ and so, $\theta = -w - \log(1-w)$.

The switch over for $p=-1$ to $p=1$ occurs when the value of q from the above expressions are equal

$$\therefore e., -w + \log(1+w) + a = -w - \log(1-w)$$

or, $\log(1-w^2) + a = 0$

or, $1-w^2 = e-a$

or, $w = \sqrt{1-e^{-\alpha}}$

Substituting this value of w in

$\theta = -w - \log(1-w)$, we have the corresponding value of θ as $-\sqrt{1-e^{-\alpha}} - \log\sqrt{1-e^{-\alpha}}$

Thus the control p is given by

$$p = \begin{cases} -1 \text{ for } \theta c \le \theta \le \alpha \\ 1 \text{ for } 0 \le \theta \le \theta c \end{cases}$$

Where $\theta c = -\sqrt{1-e^{-\alpha}} - \log\left(1-\sqrt{1-e^{-\alpha}}\right)$

5. Pontryagin's Maximum Principle.

In this section, we present the theoretically interesting and practically useful theorem due to pontryagin in connection with the optimal contrd theory.

Let the given control system is governed by the ordinary differential equation.

$$x\cdot(t) = f(x(t)), u(t), t \ge 0$$

Where $x(t) = (x_1, x_2, ..., xn)(eR^n$ and $u(t) = (u_1, u_2, ..., um) erm, x(t)$ being the state variable vector and $u(t)$ being the control variable vector having the range set $A \le R^m$ and A be the set of all possible controls.

Let the initial condition be given by

$$x(0) = x^0 = (x_1^0, x_2^0, ..., x_n^0)$$

and the pay of functional be

For : $P : \frac{\partial F}{\partial p} - \frac{d}{d\theta}\left(\frac{\partial F}{\partial p'}\right) = 0$

or, $\lambda + 2\mu p = 0$

or, $\lambda = -2\mu p$

For $z : \frac{\partial F}{\partial p} - \frac{d}{d\theta}\left(\frac{\partial F}{\partial p'}\right) = 0$

or, $2\mu z = 0$

As in the previous problem, $\mu = 0$ is not possible, so that $z = 0$, i.e. $(p + 1)(-p + 1) = 0$, or, $p = +1$.

Now q at $t = 0$, we have $\theta = \alpha$ and $\frac{d\theta}{dt} = 0$

Therefore, initially the control should be -1 and at the and of the path it should be $p = 1$. Assuming that there is only one switch which changes the control from $p = -1$ to $p = 1$, we determine the time when this swtich over takes place. For $p = -1$, the given differential equation can be written as

$$-1 = w\left(1 + \frac{dw}{d\theta}\right)$$

or, $w\frac{dw}{d\theta} = 1 + w$

or, $\frac{w}{1+w}\frac{dw}{d\theta} = 1$

Integrating we get.

$w - \log(1 + w) = -\theta + c_1, c_1$ is constant.

At $\theta = d, w = 0$, Therefore, $c_1 = a$ and so, $\theta = -w + \log(1 + w) + a$

Similarly, for $p = 1$, we have

$$1 = w\left(1 + \frac{dw}{d\theta}\right)$$

$$= \frac{dw}{dt} + w$$

$$= \frac{dw}{d\theta} + \frac{d\theta}{dt} + w$$

$$= w\left(\frac{dw}{d\theta} + 1\right)$$

Where p is the control variable.

Now the problem is to choose w to minimize T subject to the constraints $p = w\left(\frac{dw}{d\theta} + 1\right)$ and $|p| \leq 1$.

The inequality constraint $|p| \leq 1$ can be replaced by introducing a new control variable z as $x^2 = (p+1)(-p+1)$

Now We construct the augented cost functional as

$$T^* = \int_{\alpha}^{0}\left[\frac{1}{w} + \lambda\left\{p - w\left(1 + \frac{dw}{d\theta}\right)\right\} + \mu\left\{z^2 - (p+1)(-p+1)\right\}\right]d\theta$$

Where λ and μ are Lagranges multipliers.

For the optimal path, the Euler's equations are given below.

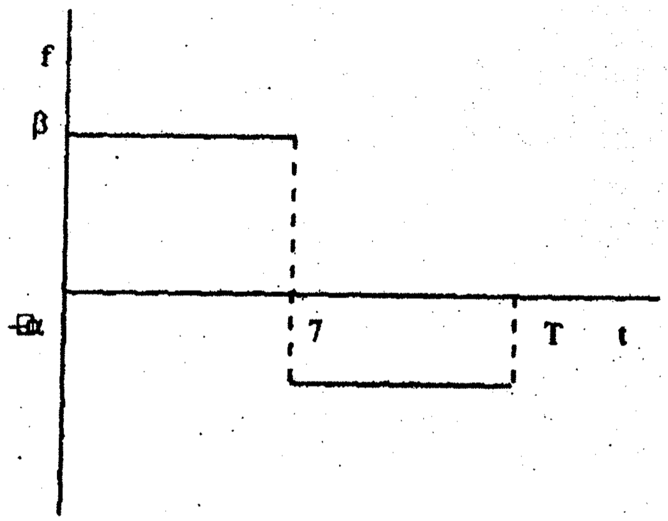
For $W : \frac{dF}{dw} - \frac{d}{d\theta}\left(\frac{dw}{d\theta = 0}\right)$

Where, $F = \frac{1}{w} + \lambda\{p - w(1+w)\} + \mu\{z^2 - (p+1)(-p+1)\}$

or, $-\frac{1}{w^2} - \lambda(1+w') - (-\lambda w') + \lambda' w = 0$

or, $\lambda = \frac{1}{w^2} + \lambda' w$

or, $\lambda' w - \lambda = \frac{1}{w^2}$



The graph reverals the fact that at $t = 7$, there is a discontinuity in f, as we switch it over from its m: ximum positive value to its minimum negative value. For this reason, this type of control is referred to as barg bang :ontrol.

Example 7 : Angular motion of a ship is described by an equation $\frac{d^2\theta}{dt^2} + \frac{d\theta}{dt} = p$

Where p is the rader setting which is subject to the constraint $|p| \le 1$.

To change the angle $\theta = \alpha$ to $\theta = 0$, it is required to make $\frac{d\theta}{dt} = 0$. Find p to minimize the time t iken for correction.

Solution : Let the time taken to change the required direction is T, which is given by

$$T = \int_{t_0}^{T} dt = \int_{\theta=\alpha}^{0} \frac{dt}{d\theta} = \int_{\alpha}^{0} \frac{1}{\frac{d\theta}{dt}} d\theta = \int_{t_0}^{t_1} \frac{1}{w} d\theta$$

Where $w = \frac{d\theta}{dt}$. Now regarding w to be a function of q, we have a state variable w which is the inde)endent variat le. Expressing p in terms of w, we have

$$p = \frac{d^2\theta}{dt^2} + \frac{d\theta}{dt}$$

$$= \frac{d}{dt}\left(\frac{d\theta}{dt}\right) + \frac{d\theta}{dt}$$

$$\frac{d^2x}{dt^2} = f = \begin{cases} \beta, 0 \le t \le \tau \\ -\alpha, \tau \le t \le T \end{cases} \quad (23)$$

We assume that by a switching device the change from β to $-\alpha$ takes place at a time $t = 7$ which is to be determined.

From (23) we get, on integration.

$$\frac{dx}{dt} = \begin{cases} \beta t, 0 \le t \le \tau \\ -\alpha(t-T), \tau \le t \le T \end{cases}$$

Again integrating,

$$x(t) = \begin{cases} \frac{1}{2}\beta t^2 \text{for } 0 \le t \le \tau \\ -\frac{\alpha}{2}(t-T) + a \text{ for } \tau \le t \le T \end{cases}$$

Now, $d(t)$ and $\frac{dx}{dt}$ must be continuous at $t = t$ then we have

$$\beta t = a(t-T)$$

and $\frac{\beta}{2}\tau^2 = \frac{\alpha}{2}(\tau - T)^2 + a$

Slowing, we get $\tau = \sqrt{\frac{2aa}{\beta(\alpha+\beta)}}$ (24)

and $T = \sqrt{\frac{2a(\alpha+\beta)}{\alpha\beta}}$ (25)

Hence the minimum time to bring the car in the stationary position at a distance 'a' is given by (25) and the optimal control to be applied on the car is given by equation (23) where t is given by equation (24).

Graphically control can be represented as follows.

We formulate the angmented cost functional as

$$T^* = \int_0^a \left[\frac{1}{\sqrt{2g}} + \alpha\left(\frac{dg}{dx} - f\right) + \mu\{z^2 - (f+\alpha)(-f+\beta)\} \right] dx$$

Where g is the stae variable, f and z are control variable and λ, μ are Lagranges multipliers. The optimal path will satisfy the following Euler's equalitions for g:

$$\frac{\partial F}{\partial g} - \frac{d}{dx}\left(\frac{\partial F}{\partial g}\right) = 0 \text{ Where } F = \frac{1}{\sqrt{2g}} + \lambda(g^1 - f) + \mu\{z^2 - (f+\alpha)(-f+\beta)\}$$

$$\text{or, } -(2g)^{-3/2} - \frac{d\lambda}{dx} = 0 \qquad (20)$$

$$\text{for } f : \frac{\partial F}{\partial z} - \frac{d}{dx}\left(\frac{\partial F}{\partial z^1}\right) = 0$$

$$\text{or, } -\lambda + 2\mu f + \mu\alpha - \mu\beta = 0 \qquad (21)$$

$$\text{for } f : \frac{\partial F}{\partial z} - \frac{d}{dx}\left(\frac{\partial F}{\partial z^1}\right) = 0$$

$$\text{or, } 2\mu z = 0 \qquad (22)$$

From (22), We have either $z = 0$ or $\mu = 0$

If $m = 0$ from (21) $\lambda = 0$ and from (20).$g \to \alpha$ which is clearly not possible, T will be zero if $g \to a$.

Hence, $m \neq 0$, so, that $z = 0$

$$\therefore (f+\alpha)(-f+\beta) = 0$$

Which gives $f = -\alpha$ or $f = \beta$.

Thus on the optimal path, the acceleration and the deceleration forces take their maximum values.

Hence from the nature of the problem, we may conclude that initially maximum accleration β is applied and then after time $t = 7$ (say), the maximum deceleration $-\alpha$ is applied to bring the car to rest (velocity is zero) at time $t = T$.

$\therefore$ The equation of motions are

$$T = \int_0^T dt = \int_0^a \frac{dt}{dx} dx = \int_0^a \frac{1}{\frac{dx}{dy}} dx = \int_0^a \frac{1}{v} dx$$

Where v is the velocity of the car. Regarding v as a function of x, we can have the end conditions in terms of v as $v(0)$ and $v(a)=0$

$$\text{Now, } f = \frac{d^2x}{dt^2} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d}{dt}(v) = \frac{dv}{dx}\frac{dx}{dt} = \frac{dv}{dx} \cdot v$$

$$= \frac{d}{dx}\left(\frac{1}{2}v^2\right)$$

$$= \frac{dg}{dx}$$

Where, $g = \frac{1}{2}v^2$ or, $v = \sqrt{2g}$

$$\therefore T = \int_0^a \frac{1}{v} dx = \int_0^a \frac{1}{\sqrt{2g}} dx \tag{16}$$

With $g(0) = 0$ and $g(a) = 0$

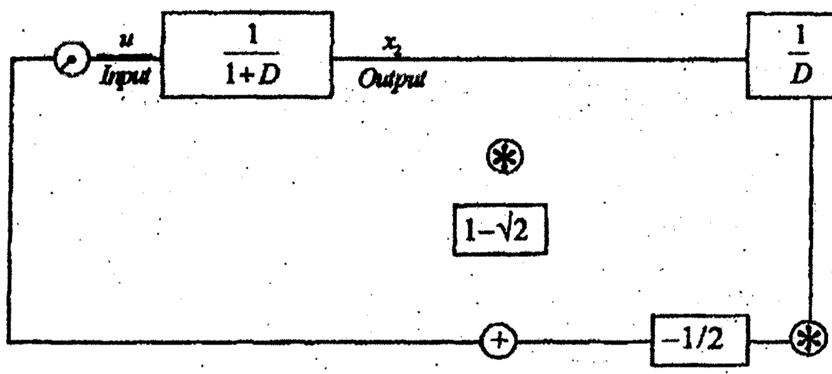
$$\text{Also we have } f = \frac{dg}{dx}, \text{ or } \frac{dg}{dx} - f = 0 \tag{17}$$

The reduced problem is to find the control f which minimizes the functional T given by (16) subject t the constraint (17) and the inequality constraint (14) along with the end conditions $g(0) = 0$ and $g(0)=0$.

Now we change the inequality constraint (14) into the equality constraint by introducing another control variable z where, $z_2 = (f+\alpha)(-f+\beta)$ (18)

z being a real variable and if the inequality constraint (14) is satisfied then $z^2 \geq 0$.

Hence the problem is to minimize T given by (16) subject to the equality constraints (17) and (18) along with the end condition $g(0) = 0$ and $g(a) = 0$.



Block diagram

4.1 Bang Bang control

A concept of Bang Bang control is illustrated by the following simple physical example.

Let a car is drivced from a stationary position on a horizontal way to a stationary position in a garage moving a total distance 'a'. The available control for the driver are the accelarator and the break (for simplicity we consider no grear change). The corresponding equation of motion for the car is

$$\frac{d^2x}{dt^2} = f \tag{13}$$

Where, $f = f(t)$ represents the acceleration or decelertion clearly, f will be subjected to both lower and upper bounds, i.e/, maximum acceleration and maximum deceleration so that $-\alpha \le f(t) \le \beta$ (14)

Where β is the maximum possible acceleration and a is the maximum possible deceleration. Now the problem is to solve the equation (13) subject to the constraint (14) with the initial conditions.

$$\left.\begin{aligned} x(0) = 0, \left(\frac{dx}{dt}\right)_{t=0} = 0 \\ \text{and}\, x(T) = a, \left(\frac{dx}{dt}\right)_{t=T} = 0 \end{aligned}\right\} \tag{15}$$

Where T is the time of travel. Now the problem is to find out the control f which accomplishes the operation in a minimum time.

The time of travel T can be expressed as

$$\therefore c_1 = b + \frac{1}{\sqrt{2}} \text{ since } c_2 = a$$

$$\therefore x_1(t) = \left\{\left(b + \frac{a}{\sqrt{2}}\right)t + a\right\}e^{-\frac{1}{\sqrt{2}}t}$$

$$\text{and } x_2(t) = \left(b + \frac{a}{\sqrt{2}}\right)e^{-\frac{1}{\sqrt{2}}t} - \frac{1}{\sqrt{2}}\left\{\left(b + \frac{a}{\sqrt{2}}\right)t + a\right\}e^{-\frac{1}{\sqrt{2}}t}$$

$$= \left\{b - \left(b + \frac{a}{\sqrt{2}}\right)\frac{t}{\sqrt{2}}\right\}e^{-\frac{1}{\sqrt{2}}t}$$

$$\therefore u = \dot{x}_2 + x_2 = \left[\frac{1}{2}\left(1 - \sqrt{2}\right)\left(b + \frac{a}{\sqrt{2}}\right)t - \left(\sqrt{2} - 1\right)b - \frac{a}{2}\right]e^{-\frac{1}{\sqrt{2}}t}$$

Thus, we have determined the control variable u which minimizes the functional J.

Representation of the solution by block diagram :

The solution of the electrochemical rpocess control problem can be illustrated by a block diagram nothing that u is of the form $u = a_1x_1 + a_2x_2$,

Where a_1, a_2 are constants to be determined.

Comparing the co-effcients of t and constant terms from both sides, we have

$$\frac{1}{2}\left(1 - \sqrt{2}\right)\left(b + \frac{a}{\sqrt{2}}\right) = \left(b + \frac{a}{\sqrt{2}}\right)a_1 - \frac{1}{\sqrt{2}}\left(b + \frac{a}{\sqrt{2}}\right)a_2$$

$$\text{i.e., } 2a_1 = \sqrt{2}a_2 = \left(1 - \sqrt{2}\right)$$

$$\text{and } -\left(\sqrt{2} - 1\right)b - \frac{a}{2} = aa_1 + ba_2$$

$$\text{i.e., } aa_1 + ba_2 = -\frac{a}{2} + \left(1 - \sqrt{2}\right)b$$

Slowing, we get $a_1 = \frac{1}{2}$ and $a_2 = 1 - \sqrt{2}$

$$\therefore u(t) = -\frac{1}{2}x_1(t) + \left(1 - \sqrt{2}\right)x_2(t)$$

$$= x_1 - \frac{1}{\alpha}\frac{d}{dt}\left(\lambda - \dot{\lambda}\right)$$

$$= \ddot{x}_1 - \frac{1}{\alpha}\mu.\left(\therefore \mu = \lambda - \dot{\lambda}\right)$$

$$= \ddot{x}_1 - \frac{1}{\alpha}x_1\left(\therefore x_1 = \dot{\mu}\right)$$

$$\therefore \frac{d^4x_1}{dt^4} - \frac{d^2x_1}{dt^2} + \frac{1}{\alpha}x_1 = 0$$

or, $\left(D^4 - D^2 + \frac{1}{\alpha}\right)x1 = 0$, Where $d \equiv \frac{d}{dt}$

This indicates that the solution of $x1, x2$ and the control variable depends on the choice of the disposable constant α. For simplicity, let us choose $\alpha = \frac{1}{4}$. Then the above equatin reduces to

$$\left(D^4 - D^2 + \frac{1}{4}\right)x_1 = 0$$

or, $$\left\{\left(D + \frac{1}{\sqrt{2}}\right)\left(D - \frac{1}{\sqrt{2}}\right)\right\}^2 x_1 = 0$$

or, $x_1(t) = (c_1t + c_2)e^{\frac{1}{\sqrt{2}}t} + (c_3t + c_4)e^{\frac{1}{\sqrt{2t}}}$

Where c_1, c_2, c_2, c_4 are constants which are to be evaluated with the help of the end conditions.

As $x_1 \to 0$ as $t \to \alpha$, We have $c_3 - c_4 = 0$

$\therefore x_1(t) = (c_1t + c_2)e^{\frac{1}{\sqrt{2}}t}$, So that $x_2 = x_1 = c_1e^{\frac{1}{\sqrt{2t}}} - \frac{1}{\sqrt{2}}(c_3t + c_4)e^{\frac{1}{\sqrt{2t}}}$

Now, $x_1(0)=a$ gives $c_2 = a$, and $x_0(0) = b$ gives $c_1 - \frac{1}{\sqrt{2}}c_2 = b$

for $x_1 : \frac{\partial F}{\partial x_1} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}_1}\right)$

or, $x_1 - \frac{d\lambda}{dt} = 0$

or, $x_1 = \lambda$

for $x_2 : \frac{\partial F}{\partial x_1} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}_1}\right) = 0$

or, $-\mu + \lambda - \frac{d\lambda}{dt} = 0$

or, $\mu = \lambda - 1$.

for $u; \frac{\partial F}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}_1}\right) = 0$

or, $\alpha u - \lambda = 0$ or $\lambda = \alpha u$.

Combining these three relations $x_1 = \lambda, \varpi = \lambda - \dot{\lambda}$ and $\lambda = \alpha u$ along with the constraints, we can eliminate λ and μ and and solve for x_1, x_2 and u as follows.

$$\frac{d^4 x_1}{dt^4} = \frac{d^3}{dt^3}\left(\frac{dx_1}{dt}\right) = \frac{d_3 x_1}{dt_3} = \frac{d^2}{dt^2}\left(\frac{dx_1}{dt}\right) = \frac{d^2}{dt^2}(-x_2 + u)$$

$$= -\frac{d}{dt}\left(\frac{dx_1}{dt}\right)u$$

$$= \frac{d}{dt}(-x_2 + u) + u$$

$$= \dot{x}_2 - \dot{u} + \ddot{u}$$

$$= \ddot{x}_1 \frac{\lambda}{\alpha} + \frac{\lambda}{\alpha} (\because \lambda = \alpha u)$$

Example 6 : An electrochemical system is modelled by the differential equation $\ddot{x} = -\dot{x} + u,$ where x and u are functions of time t.

Minimize the cost functional $J\frac{1}{2}\int_0^{\alpha}\left(x^2 + \alpha \dot{u}^2\right)dt$ when α is a disposable constant by chosing the control variable properly.

Solution : As before, we convert the given second order differential equation to the first order differential equation by introducing the state variables x_1 and x_2

as $x_1 = x$

and $x_2 = \dot{x}_1 = \dot{x}$

Hence, the variabels x_1 and x_2 satisfy the equations

$\dot{x}_2 = -x_2 + u$

and $\dot{x}_1 = x_2$

The end conditions for an electrochemical system are that x and x are given at $t = 0$ and both tend to zero as $t \to a$

In terms of state variables, the above end conditons can be represented as

$x_1(0) = a$ (say)

$x_2(0) = b$ (say)

$x_1 \to 0$ and $x_2 \to 0$ as $t \to a$

Where, a and b are constants.

Hence the problem reduces to minimizing

$$J = \frac{1}{2}\int_0^{\alpha}\left(x_1^2 + \alpha u^2\right)dt$$

subject to the constraints $\dot{x}_2 = -x_1 + u$ and $\dot{x}_1 = x_2$ along with the above end conditions.

We form the angmented cost functons with the help of the lagranges multipliers λ and μ as.

$$J^* = \frac{1}{2}\int_0^{\alpha}\left(x^2 + \alpha u^2\right)dt$$

The Euler's equation for the state variables $x1$ and $x2$ and the control variable u for the above functional are:

With the conditions $x_1(0) = a$ and $x_2(0) = 0$

Now the control problem is to choose u so that the system (x_1, x_2) moves from $(a, 0)$ at $t = 0$ to $(0, 0)$ at some subsequent time.

As an initial guess, let us assume u = constant = c.

Then we have $\dot{x}_2 = -W^2 x_1 + C$

or $\ddot{x}_1 = -w^2 x_1 + C$

or, $x_1 = A\cos wt + B\sin wt + \frac{C}{w^2}$

and $x_2 = \dot{x}_1 = -Aw\sin wt + Bw\cos wt.$

At $t = 0$, we have $x_1 = a$ and $x_2 = 0$

$\therefore B = 0$ and $A = a - \frac{C}{w^2}$

$$\therefore x_1 = \left(a - \frac{C}{w^2}\right)\cos wt + \frac{C}{w^2}$$

and $x_2 = \left(a - \frac{C}{w^2}\right)$

The value of x_1, i.e. velocity will again becomes zero when $t = \frac{\pi}{w}$. At that time

$x_1 = -\left(a - \frac{C}{w^2}\right) + \frac{C}{w^2} = -a + \frac{C}{w^2}$. By the given condition, x_1, i.e., displacement must be zero.

$$\therefore -a + \frac{2C}{w^2} = 0$$

$$\text{or, } C = \frac{aw^2}{2}$$

Hence the control variable, i.e. force $u = \frac{aw^2}{2}$ takes the system from $(a, 0)$ at $t = 0$ to $(0, 0)$ at $t = \frac{\pi}{w}$ and the system is controllable.

Now, using the boundary condition $x1(0)=0$, we have $D=0$ and $x_2(0)=1$ gives $C=1$.

$$\therefore x_2(t)=\frac{At^2}{2}+Bt+C$$

$$\text{or, } x_1=\frac{At^3}{6}+\frac{Bt^2}{2}+t$$

Again $x_1(1)=1$ gives $\frac{A}{6}+\frac{B}{2}=0$

and $x_2(1)=1$ gives $\frac{A}{2}+B=0$, so that $A=B=0$

and thus $x(t)=x_1(t)=t$

Hence, the optimal path is given by $x(t)=t, 0\le t\le 1$

and he corresponding value of the functional is $J^*=\int_0^1\left[1+\left(\frac{d^2x}{dt^2}\right)^2\right]dt=1$

4. Optimal Control.

With the help of the following examples, the problems of optimal control are illustrated in this section.

Example 5 : A particle is attached to the lower end of a vertical spring whose other end is fixed, is oscillating about its equillibrium position. If x denote the particule's displacement from the equilibrium position, the governing differential equation for this motion is $\ddot{x}=-w^2x$.

If the particle is at its maximum displacement $x=a$ at time $t=0$ and at this instant of time, a force is per unit mass is applied to the particle in order to bring the particle to rest when its displacement is zero, find such a force u.

Solution. After the application of the force u at $x=a$ and $t=0$, the governing differential equation becomes $\ddot{x}=-w^2x+u$.

The system can be represented in the form of first order differential equations by using the variables $x_1=x$ and $x_2=\dot{x}_1=\dot{x}$

as $\dot{x}_2=-w_2x_1+u$ and $x_2=\dot{x}_1$

The constraint is incorporated into the functional J, using Lagranges multiplier λ and we formulate the angmented functional as

$$J^* = \int_0^1 \left[1+\dot{x}_2^2 + 1(x_2 - \dot{x}_1)\right] dt$$

To find the optimal path of this functional, we solve the corresponding Euler's equations given below.

For $x_1 : \frac{\delta F}{\delta x_1} - \frac{d}{dt}\left(\frac{\delta F}{\delta \dot{x}_1}\right) = 0$ where $F = 1+\dot{x}_2^2 + \lambda(x_2 - \dot{x}_1)$

or, $\frac{d}{dt}(-1) = 0$

or, $\dot{\lambda} = 0$

or, λ = constant indendent of t

For $x_2 : \frac{\delta F}{\delta x_2} - \frac{d}{dt}\left(\frac{\delta F}{\delta \dot{x}_2}\right) = 0$

or, $\lambda - 2\ddot{x}_2 = 0$

or, $\ddot{x}_2 = \frac{\lambda}{2} = A(say)$

or, $\dot{x}_2 = At + B$

or, $x_2 = \frac{At^2}{2} + Bt + C$

$\therefore \dot{x}_1 = \frac{At^2}{2} + Bt + C$

or, $x_1 = \frac{At^3}{6} + \frac{Bt^3}{2} + Ct + D$

Where A, B, C, D are constants.

$$J = \int_{t_0}^{t_1} F(x_1, x_2, ..., x_n, \dot{x}_1, \dot{x}_2, ..., \dot{x}_n, t)\, dt \qquad (11)$$

Where xi's satisfy the constraints.

$$g(x_1, x_2, ..., x_n; \dot{x}_1, \dot{x}_2, ..., \dot{x}_n; t) = 0 \,\&\, h(x_1, x_2, ..., x_n; \dot{x}_1, \dot{x}_2, ..., \dot{x}_n; t) = 0 \qquad (12)$$

We solve this problem in the same way as we solve on ordinary extremum of a function by introducing two Lagrangian λ and μ and forming the angmented cost functional as

$$J^* = \int_{t_0}^{t_1} (F + \lambda g + \mu h)\, dt$$

Now, we find the extremum of these functional as before and the lagranges multipliers are found from the Euler's equationsn for this new functional and the given constraints.

Example 4 : Find the stationary path $x = x(t)$ for the functional $J = \int_0^1 \left[1 + \left(\frac{d^2x}{dt^2}\right)\right] dt$

Subject to the boundary conditions

$x(0) = 0$

$\dot{x}(0) = 1$

$x(1) = 1$

$\dot{x}(1) = 1$

Solution. To eliminate the second order derivative form the functional J, we introduce two variables, namely $x_1 = x$ and $x_2 = x$

as it is supposed to be a function of t, x and x only. Thus the problem reduces to funding the extremum of the functional $J = \frac{dF}{d\dot{x}}$

with the boundary conditions $x_1(0) = 0$,

$x_2(0) = 1$

$x_1(1) = 1$

$x_1(1) = 1$

$m_2(1) = 1$

and the constraint $x_2 - \dot{x}_1 = 0$

Since hm are perfectly arbitrary and independent of one another, the terms within the square bracket of equation (9) are separately zero. Thus,

$$\frac{\partial F}{\partial x_k}-\frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}_k}\right)=0, k=1,2,3,...,n. \qquad (10)$$

Equation (10) represents a whole set of Euler-Laangrange equations each of which must be satisfied for an extreme value.

Example 3 : Find x and y as functions of t, so that

$$J=\int_{t_0}^{t_1}\left[\frac{m}{2}\left(x^2+y^2\right)-mgy\right]dt$$

May have stationary value. It may be assumed that x and y are given at to and $t1$.

Solution : Here, $F=\frac{m}{2}\left(x^2+y^2\right)-mgy$ contains two dependent variables x and y

Thus, Enler-Lagrange equations are

$$\frac{\partial F}{\partial x}-\frac{d}{dt}\left(\frac{dF}{d\dot{x}}\right)=0 \text{ and } \frac{\partial F}{\partial x}-\frac{d}{dt}\left(\frac{dF}{d\dot{y}}\right)=0$$

These equations reduce to

$$\frac{d}{dt}(\dot{x})=0 \text{ and } \frac{d}{dt}(m\dot{y})+mg=0,$$

Which give $x=c_1t+c^2$ and $y=-\frac{1}{2}gt^2+d_1t+d_2$,

Let $x(t_0)=x_0, d_2$ and the required solution is $x=c_1t+t_2$ and $y=-\frac{1}{2}gt^2+d_1t+d_2$,

3.2 Optimization with constraints :

Let the problem is to find the path $x_i=x_i(t), to \le +t_1, (i=1,2,...,n)$ which maximizes or minimizes the cost functional we have $\frac{dJ}{d\varepsilon} = \int_{t_0}^{t_1} \sum_k \left(\frac{\partial F}{\partial x^* k} \cdot \frac{dxk^*}{d\varepsilon} + \frac{\partial F}{\partial \dot{x}^* k} \cdot \frac{d\dot{x}^* k}{d\varepsilon} \right) dt$

$$= \int_{t_0}^{t_1} \sum_k \left(\frac{\partial F}{\partial x^* k} \eta_k(t) + \frac{\partial F}{\partial \dot{x}^* k} \dot{\eta}_k(t) \right) dt$$

Now, $\left[\frac{dJ(\varepsilon)}{d\varepsilon} \right]_{\varepsilon=0} = \int_{t_0}^{t_1} \sum_k \left(\eta_k \frac{\partial F}{\partial x_k} + \dot{\eta}_k \frac{\partial F}{\partial \dot{x}_k} \right) dt$

Since $\left(x_k^*\right)_{t=0} = x_k$ and $\left(\dot{x}_k^*\right)_{e=0} = \dot{x}_k$.

The condition for the path along which J has stationary value is $\left[\frac{dJ(\varepsilon)}{d\varepsilon} \right]_{\varepsilon=0} = 0$

Hence $= \int_{t_0}^{t_1} \sum_k \left(\eta_k \frac{\partial F}{\partial x_k} + \dot{\eta}_k \frac{\partial F}{\partial \dot{x}_k} \right) dt = 0$ (8)

Integrating the second term by parts,

$$= \int_{t_0}^{t_1} \dot{\eta}_k \frac{\partial F}{\partial \dot{x}_k} dt = \left[\dot{\eta}_k \frac{\partial F}{\partial \dot{x}_k} \right]_{t_0}^{t_1} - \int_{t_0}^{t_1} \eta_k \frac{d}{dt}\left[\frac{\partial F}{\partial \dot{x}_k} \right] dt - \int_{t_0}^{t_1} \eta_k \frac{d}{dt}\left[\frac{\partial F}{\partial \dot{x}_k} \right] dt$$

[since $\eta_k(t_0) = \eta_k(t_1) = 0$]

Using this result, equation (8) becomes

$$\int_{t_0}^{t_1} \sum_k \left[\eta_k \frac{\partial F}{\partial x_k} - \eta_k \frac{d}{dt} \frac{\partial F}{\partial \dot{x}_k} \right] dt = 0.$$

or, $\int_{t_0}^{t_1} \sum_k \eta_k \left[\frac{\partial F}{\partial x_k} - \frac{d}{dt} \frac{\partial F}{\partial \dot{x}_k} \right] dt = 0.$ (9)

Now the euler's equation for this problem, gives $\frac{d}{dt}(2\dot{x})-0=0,$

or $\ddot{x}=0,$ or, $x=At+B,$ Where A, B are constants.

Using the conditoins $x(0)=1$ and $x.(1)=0$, we have $B=1$ and $A=0$

$\therefore x(t)=1$ for all t.

The optimum value of J along $x(t)=1$ is $J=\int_0^1(0+1)dt=1$.

3.1 Cost functional involving several dependent variables :

Let the integrand F be a function of one independent variable t and several dependent variables $x_1(t),x_2(t),...,x_n(t)$. These dependent variables are functions of t only.

Now the problem is to find the functions $x_1(t),x_2(t),...,x_n(t)$ such that the integral

$J=\int_{t_0}^{t_1}F(x_1,x_2,...,x_n,\dot{x}_1,...,\dot{x}_n;t)dt$ (6) may be statonary.

Let us consider two paths out of infinite number of paths between two points P and Q, Such that the difference between then may be described by the functions $\eta_k(t)$ and ε.

The function $\eta_k(t)$ must satisfy the following two conditions :

(i) All the neighbouring paths must pass throught the fixed points P and Q, i.e. $\eta_k(t_0)=\eta_k(t_1)=0$

(ii) $\eta_k(t)$ must be differentiable.

Let P & Q (Figure 1) be the path along which J has stationary value and $PR'Q$ be a neighbouring path. If x^*k and $\dot{x}^*k$ are the values of x_k and $\dot{x}_k$ along the optimum path, then the values of $x_1^*,x_2^*,...,x_n^*$ in terms of $\eta_k(t)$ and ε are $x_i^*=x_i+\delta x_i=x_i+\varepsilon\eta_i(t),\ i=1,2,...,n.$

If the integral has stationary value along PRQ, then

$$J(e)=\int_{t_o}^{t_i}F(x_1+\varepsilon\eta_1,x_2+\varepsilon\eta_2,...,x_n+\varepsilon\eta_n,\dot{x}_2+\varepsilon\dot{\eta}_2...,x.n+\varepsilon\dot{\eta}_n;t_0dt \quad (7)$$

Now, $J(\varepsilon)$ is statonary for $\varepsilon=0$, as in case of single dependent variable. Differentiating (7) with respect to e,

Note-3 : When x is given at one end point only, say at $t = to$, then $\left[\eta \frac{\delta F}{\delta \dot{x}}\right]_{t_0}^{t_1} = 0$, gives $\eta(t_0) = 0$ and $\frac{\delta F}{\delta \dot{x}} = 0$ at $t=t_1$. These end point conditions are called transversality condition.

Example 1 : Find he curve $x = x(t)$ which minimize the functional $J = \int_0^1 (x^2 + 1)dt$ where, $x(0)=1$ and $x(1)=2$.

Solution : The Enler's equation for this problem, is given by

$$\frac{d}{dt}\left(\frac{\delta F}{\delta \dot{x}}\right) = 0 \text{ where } F(x, \dot{x}, t) = \dot{x}^2 + 1$$

Now $\left(\frac{\delta F}{\delta \dot{x}}\right) = 0$ and $\frac{\delta F}{\delta \dot{x}} = 2x.$, so that $\frac{\delta F}{\delta \dot{x}}(2\dot{x}) - 0 = 0, or\ \ddot{x} = 0,$

or, $x = At + B$, Where A, B, are constants.

From the end conditions, we have $x(0) = 1 \Rightarrow A \cdot 0 + B = 1 \Rightarrow B = 1$

and $x(1) = 2 \Rightarrow A + B = 2 \Rightarrow A = 1$

Hence, the optimal path is $x(t) = t + 1$

and the corresponding value of J is $J \int_0^1 (x^2 + 1)dt = \int_0^1 (1+1)dt = 2$

Example 3 : Find the function $x(t)$ which minimizes the functional $J \int_0^1 (\dot{x}^2 + 1)dt$ where $x(0)=1$, but x can take any value at $t = 1$

Solutin. Here also $F(x, \dot{x}, t) = \dot{x}^2 + 1$

Since x is not prescribed at the end point $t = 1$, we have the transversality conditoin $\frac{\delta F}{\delta \dot{x}} = 0$ $t = 1$, i.e. $\dot{x}(1) = 0$

$$\int_{t_0}^{t_1}\left\{F_x\left(x^{*}t\right),x^{*},t)\eta(t)+F\dot{x}\left(x^{*}(t)t\right)\dot{\eta}(t)\right\}dt=0 \qquad (4)$$

Now we consider the term $\int_{t_0}^{t_1} F_{\dot{x}}(x^{*}(t),\dot{x}^{*}(t)\dot{\eta}(t)dt$

$$=\left[F\dot{x}\int\eta(t)dt\right]_{t_0}^{t_1}-\int_{t_0}^{t_1}\left[\left\{\frac{d}{dt}(F\dot{x})\int\dot{\eta}(t)dt\right]dt$$

$$=-\int_{t_0}^{t_1}\left(F_{\dot{x}}\right)\eta(t)dt \quad \left[\because \eta(t_0)=\eta(t_i)=0\right]$$

$$=\text{Thus (4) becomes} = -\int_{t_0}^{t_1}\left(F_x\right)\eta(t)dt-\frac{d}{dt}F_{\dot{x}}\,\eta(t)dt=0$$

$$\text{or, } \int_{t_0}^{t_1}\left\{F_x-\frac{d}{dt}\left(F_{\dot{x}}\right)\right\}\eta(t)dt=0$$

Since $\eta(t)$ is arbitrary deformation of the path, therefore

$$F_x-\frac{d}{dt}\left(F_{\dot{x}}\right)=0$$

$$\text{or, } \frac{d}{dt}\left(\frac{\delta F}{\delta\dot{x}}\right)-\frac{\delta F}{\delta x}=0 \qquad (5)$$

This equation is known as Euler-Lagranges equation or simply Euler's equation and is used frequently in he study of calculus of variations.

Note-1 : When x is given at the end points to and t1, equ. (5) gives the necessary condition for the optimal path.

Note-2 : When x is not given at both the end points to and t_1, then η is completely arbitrary and its value at the end points are also arbitrary. In this case, the necessary Conditions for extremum of J are given by equation (5) together with $\left[\eta\frac{\delta F}{\delta\dot{x}}\right]_{t_0}^{t_1}=0, i.e., \frac{\delta F}{\delta\dot{x}}=0$ at $t=t_0$ & $t=t=t_1$.

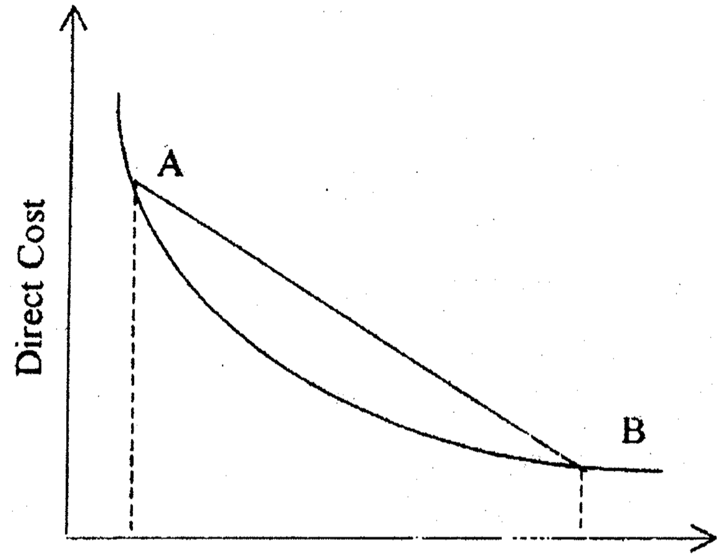


Fig. 14

The direct cost curve (from the relationship of direct cost and time) are shown in the figure. The point B denotes the normal time for completion of an activity whereas the point. A denotes the crash time which indicates the least duration in which activity can be completed. The cost curve is non-linear and asymptotic nature. But, for the sake of simplicity, it can be approximated by a straight line whose slope (in magnitude) is given by

$$\text{Cost slope} = \frac{\text{Crash cost - Normal cost}}{\text{Normal time - Crash time}} = \frac{C_c - C_n}{T_n - T_c}$$

It is also called as crash cost slope or crash rate of increase in the cost of performing the activity per unit reduction in time and is called cost / time trade off. It varies from activity to activity. After assessing the direct and indirect project costs, the total project cost which is the sum of direct and indirect cost can be found out.

Time-cost optimization algorithm/Time-cost trade off procedure.

The following are the steps involved in the project crashing.

Step-1. Considering normal times of all activities, identify the critical activities and find the critical path.

Step-2. Calculate the cost slope for different activities and rank the activities in the ascending order of cost slope.

Step-3. Crash the activities on the critical path as per ranking i.e. activity having lower cost slope would be crashed first to the maximum extent possible (For the crashing of lower cost slope i.e., for the reduction of activity duration time, the direct cost of the project would be increased very slowly).

Step-4. Due to the reduction of critical path duration by crashing in Step-3, other path also become critical i.e., we get parallel critical paths. In such cases, the project duration can be reduced by crashing of activities simultaneously in the parallel critical paths.

Step-5. Repeat the process until all the critical activities are fully crashed or no further crashing is possible.

In the case of indirect cost, the process of crashing is repeated until the total cost is minimum beyond which it may increases. The minimum cost called the optimum project cost and the corresponding time, the optimum project time.

Example 4

The following table shows activities, their normal time and cost and crash time and cost for a project.

Activity	Normal time (days)	Cost (Rs.)	Crash time (days)	Cost (Rs.)
1 – 2	6	1400	4	1900
1 – 3	8	2000	5	2800
2 – 3	4	1100	2	1500
2 – 4	3	800	2	1400
3 – 4	Dummy	–	–	–
3 – 5	6	900	3	1600
4 – 6	10	2500	6	3500
5 – 6	3	500	2	800

Indirect cost for the project is Rs. 300 per day

(i) Draw the network of the project.

(ii) What are the normal duration and associate cost with project?

(iii) What will be the least project duration and the minimum project cost?

(iv) Find the optimum duration and minimum project cost?

Solution :(i)

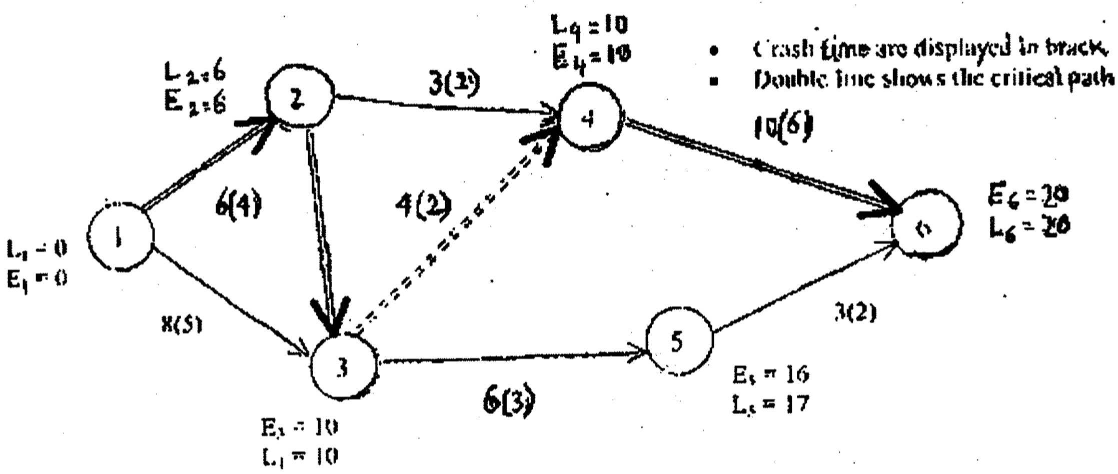


Fig. 15

(ii) Using the normal time duration of each activity, the earliest and latest occurrence time at various nodes are computed and displayed in **Fig. 15 of the network.**

From the network, it is seen that L-values and E-values at nodes 1, 2, 3, 4, 6 are same. This means that the critical path is 1 – 2 – 3 – 4 – 6 and the normal duration of the project is 20 days. The associated cost of the project.

= Direct normal cost + indirect cost for 20 days

= Rs. [(1400 + 2000 + 1100 + 800 + 900 + 2500 + 500) + 20 x 300]

= Rs. [9200 + 6000] = Rs. 15200

(iii) The cost slopes of different activities are computed by using the formula

$$\text{Cost slope} = \frac{\text{Crash cost - Normal cost}}{\text{Normal time - Crash time}}$$

and these are shown in the following table.

Activity	1 – 2	1 – 3	2 – 3	2 – 4	3 – 5	4 – 6	5 – 6
Slope	250	267	200	600	233	250	300

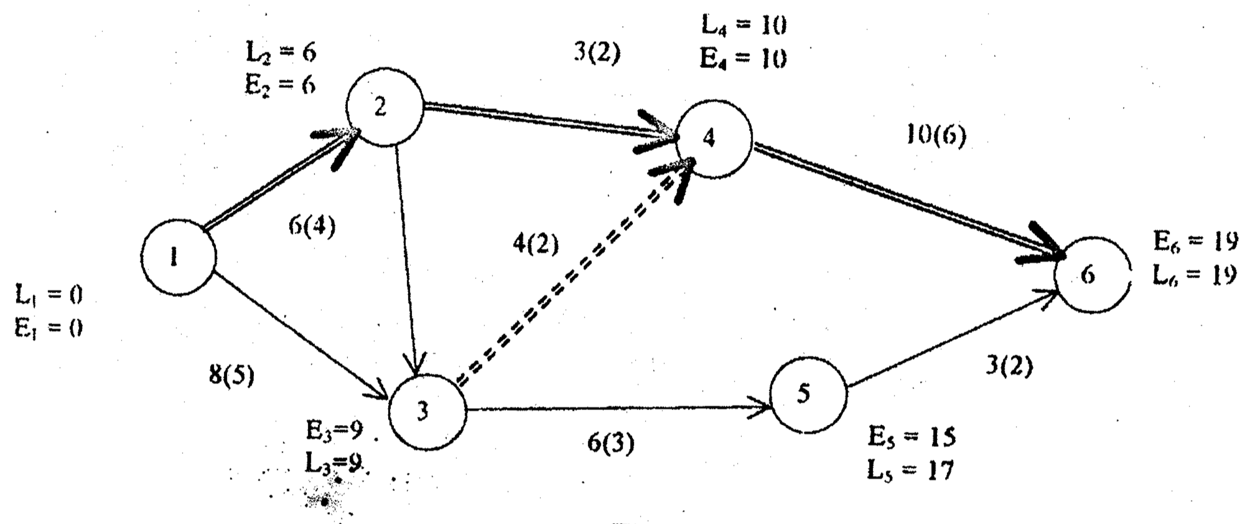
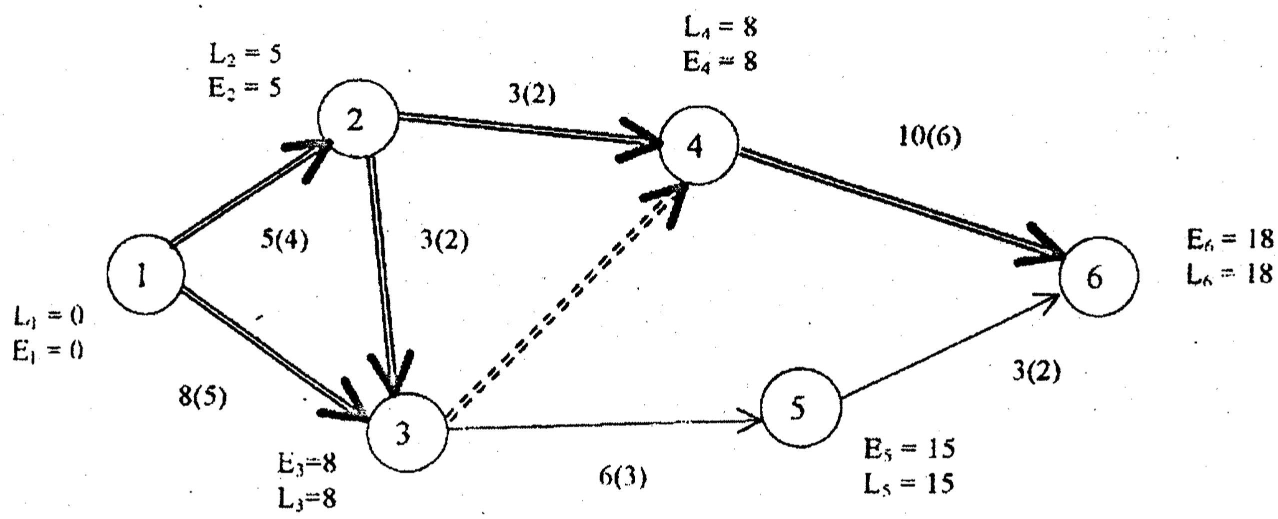


Fig. 16

Now we constant the following table for finding the optimal cost with duration and minimum project duration with cost.

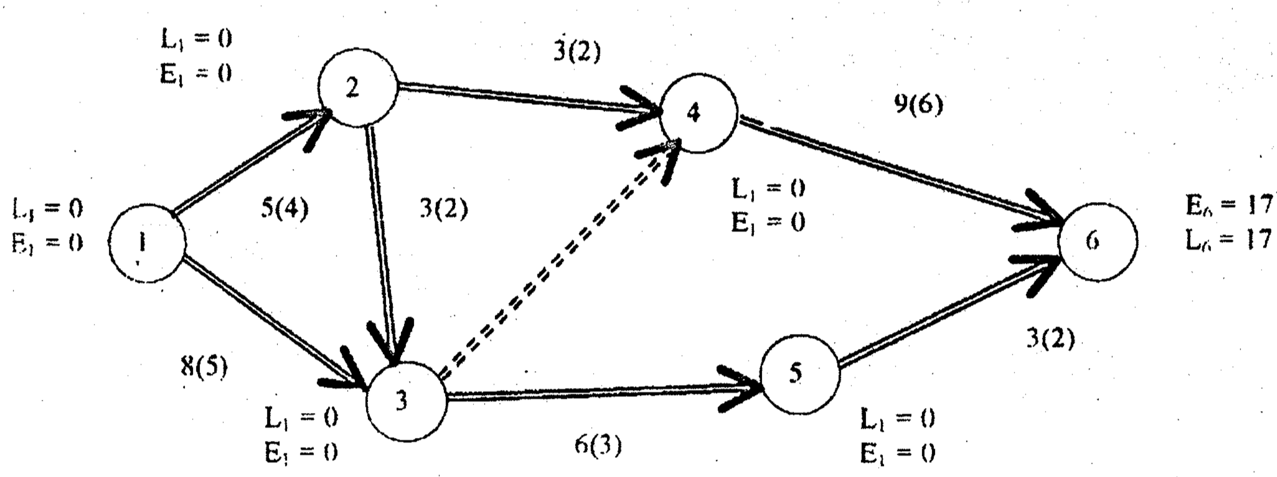
Critical Path(s)	See Figure	Activities crashed (time)	Project length (days)	Normal direct cost (Rs.) (A)	Crashing cost (Rs.) (B)	Indirect cost (Rs. 300/day) (C)	Total cost (Rs.) (A + B + C)
1-2-3-4-6	Fig. 15	–	20	9200	–	300 × 20	15200
1-2-3-4-6	Fig. 15	2-3(1)	19	9200	200×1 = 200	300 × 19	15100
1-2-3-4-6, 1-2-4-6	Fig. 16	1-2(1)	18	9200	200+250×1 = 450	300 × 18	15050
1-2-3-4-6, 1-2-4-6 1-3-4-6	Fig. 17	4-6(1)	17	9200	450+250×1 = 700	300 × 17	15000
1-2-3-4-6,	Fig. 18	3-5(1)	16	9200	700+233×1+250	300 × 16	15183

1-2-4-6, 1-3-4-6, 1-3-5-6, 1-2-3-5-6		4-6(1)			×1= 1183		
1-2-3-4-6, 1-2-4-6, 1-3-4-6, 1-3-5-6, 1-2-3-5-6	Fig. 19	3-5(1) 4-6(1)	15	9200	1183+233×1+250 ×1 = 1666	300 × 15	15366
1-2-3-4-6, 1-2-4-6, 1-3-4-6, 1-3-5-6, 1-2-3-5-6,	Fig. 20	3-5(1) 4-6(1)	14	9200	1666+233×1+250 ×1= 2149	300 × 14	15549
1-2-3-4-6, 1-2-4-6, 1-3-4-6, 1-3-5-6, 1-2-3-5-6	Fig. 21	1-2(1) 1-3(1)	13	9200	2149+250×1+267 ×1 = 2666	300 × 13	15766
1-2-3-4-6, 1-2-4-6, 1-3-4-6, 1-3-5-6, 1-2-3-5-6	Fig. 22	2-3(1) 2-4(1) 1-3(1)	12	9200	2666+200×1+600 ×1+267×1 = 3733	300 × 12	16533



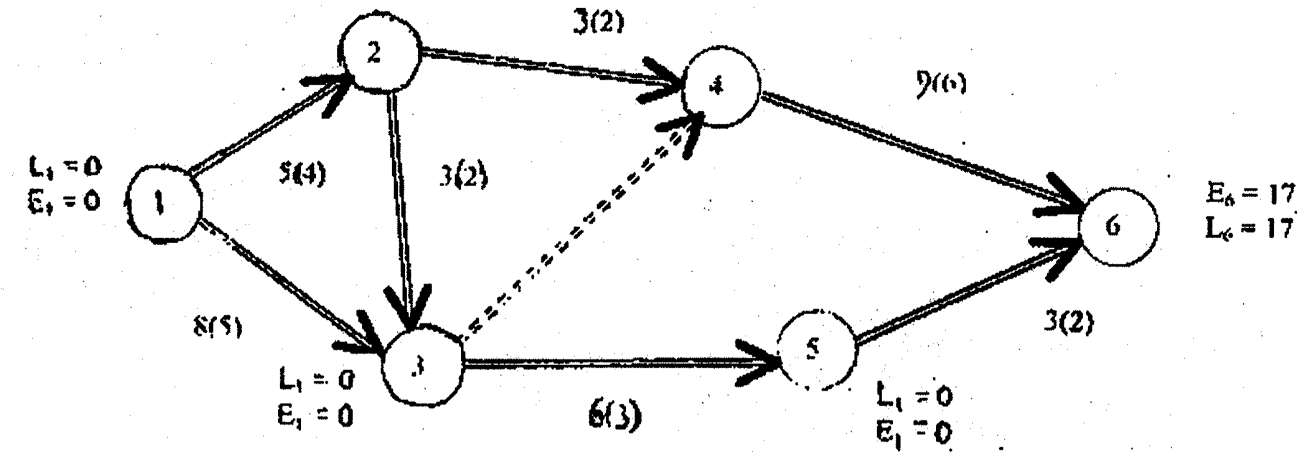
L₂ = 5
E₂ = 5
L₄ = 8
E₄ = 8
3(2)
10(6)
5(4)
3(2)
E₆ = 18
L₆ = 18
L₁ = 0
E₁ = 0
8(5)
3(2)
E₃=8
L₃=8
6(3)
E₅ = 15
L₅ = 15
1
2
3
4
5
6

Fig. 17



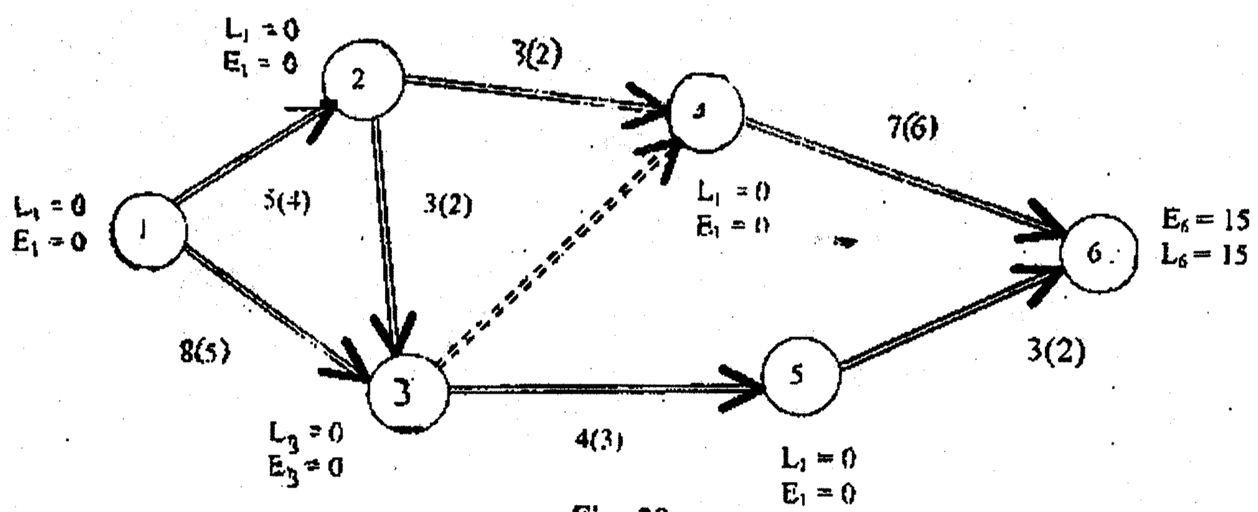
L₁ = 0
E₁ = 0
3(2)
9(6)
L₁ = 0
E₁ = 0
5(4)
3(2)
L₁ = 0
E₁ = 0
E₆ = 17
L₆ = 17
3(2)
8(5)
L₁ = 0
E₁ = 0
6(3)
L₁ = 0
E₁ = 0
1
2
3
4
5
6

Fig. 18



3(2)
9(6)
$L_1 = 0$
$E_1 = 0$
5(4)
3(2)
$E_6 = 17$
$L_6 = 17$
8(5)
3(2)
$L_1 = 0$
$E_1 = 0$
6(3)
$L_1 = 0$
$E_1 = 0$

Fig. 19



$L_1 = 0$
$E_1 = 0$
3(2)
7(6)
$L_1 = 0$
$E_1 = 0$
5(4)
3(2)
$L_1 = 0$
$E_1 = 0$
$E_6 = 15$
$L_6 = 15$
8(5)
3(2)
$L_3 = 0$
$E_3 = 0$
4(3)
$L_1 = 0$
$E_1 = 0$

Fig. 20

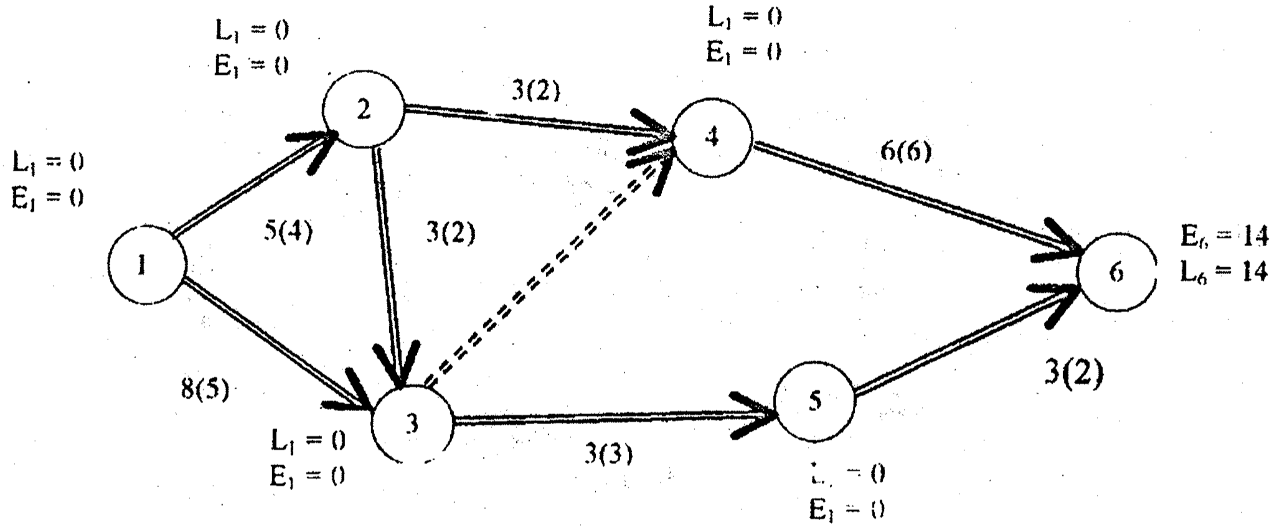

L₁ = 0
E₁ = 0
L₁ = 0
E₁ = 0
L₁ = 0
E₁ = 0
3(2)
6(6)
5(4)
3(2)
E₆ = 14
L₆ = 14
8(5)
3(2)
L₁ = 0
E₁ = 0
3(3)
E₁ = 0
1
2
3
4
5
6

Fig. 21

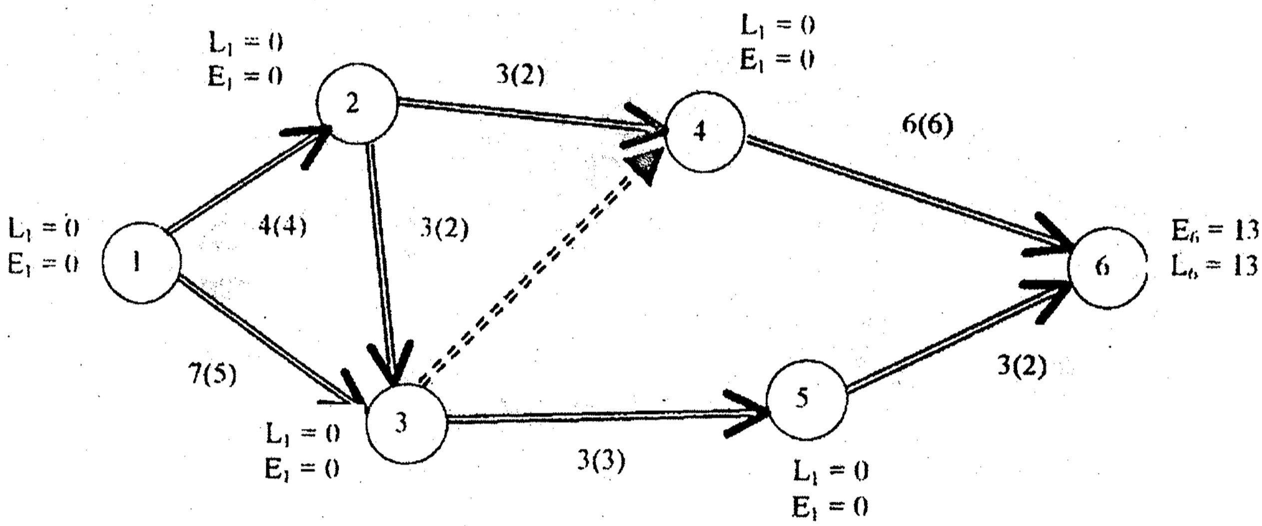

L₁ = 0
E₁ = 0
L₁ = 0
E₁ = 0
3(2)
6(6)
L₁ = 0
E₁ = 0
4(4)
3(2)
E₆ = 13
L₆ = 13
7(5)
3(2)
L₁ = 0
E₁ = 0
3(3)
L₁ = 0
E₁ = 0
1
2
3
4
5
6

Fig. 22

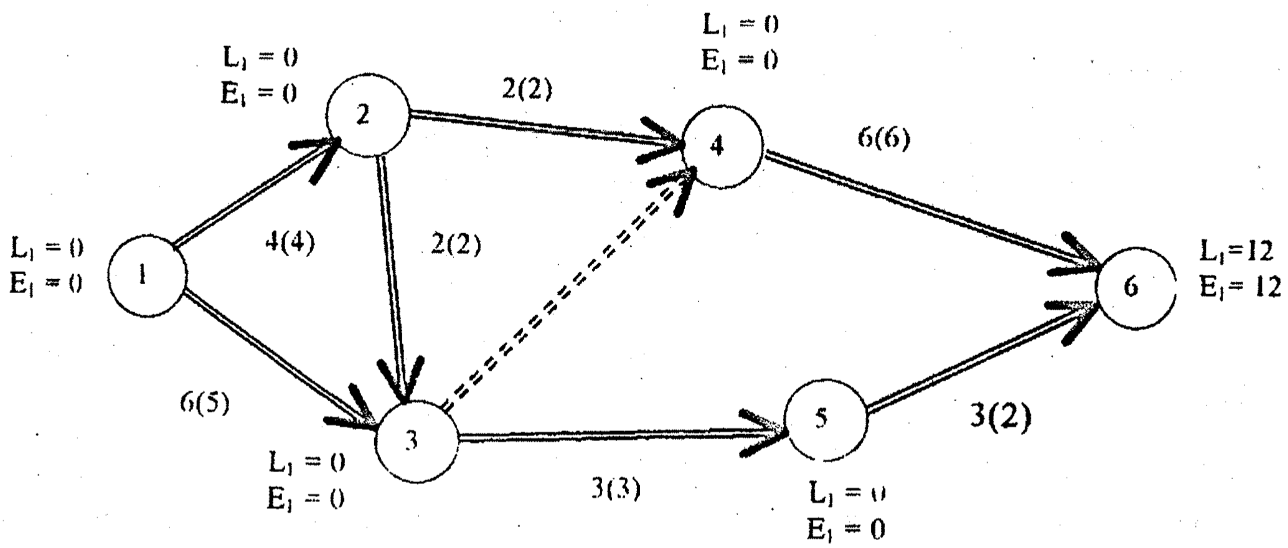


Fig. 23

From **Fig. 23**, it is seen that no further crashing is possible beyond 12 days. Hence the least project duration is 12 days and the corresponding cost of the project will be Rs. 16,493.00.

As the minimum cost occurs for 17 days schedule, optimum duration of the project is 17 days and the minimum project cost is Rs. 15,000.00.

15. SELF ASSESSMENT QUESTIONS / EXERCISE

1. Discuss the different phases of a project.
2. What is network analysis? What are the advantages of it?
3. Discuss the common errors in a network construction.
4. What is critical path? What are the main features of it?
5. Explain the terms : total float, independent float, free float, slack, crashing.
6. What is PERT? What information is revealed by PERT analysis?
7. Explain the following time estimates used in PERT :

   (i) Optimistic, (ii) Pessimistic, (iii) Most likely
8. Distinguish between PERT and CPM.
9. Define expected time and variance in terms of optimistic, pessimistic and most likely time.
10. What do you mean by time-cost trade off? Define cost slope.

11. How do you calculate the earliest starting time and the earliest finish time?

12. Define float and different types of floats.

13. A project consists of a series of tasks labeled A, B, ..., H, I with the following relationships (W < X, Y means X and Y can not start until W is completed; X, Y < W means W can not start until both x and y are completed). With this notation, construct the network diagram having the following constraints :

$$A < D, E; B, D < F; C < G; G < H, F, G < I$$

14. Find also the minimum time of completion of the project, when the time (in days) of completion of each task is as follows:

Task	A	B	C	D	E	F	G	H	I
Time	23	8	20	16	24	18	19	4	10

15. The following are the details of estimated times of activities of a certain project.

Activity	A	B	C	D	E	F
Immediate Predecessor :	–	A	A	B, C	–	E
Estimated Time (weeks)	2	3	4	6	2	8

(a) Find the critical path and the expected time of the project.

(b) Calculate the earliest start time and earliest finish time for each activity.

(c) Calculate the float for each activity.

16. A project consists of eight activities with the following relevant information :

Activity	**Immediate Predecessor**	**Estimated duration (days)**		
		Optimistic	**Most likely**	**Pessimistic**
A	–	1	1	7
B	–	1	4	7
C	–	2	2	8
D	A	1	1	1
E	B	2	5	14
F	C	2	5	8
G	D, E	3	6	15
H	F, G	1	2	3

(ii) Draw the network and find out the expected project completion time.

(iii) What duration will have 95% confidence for project completion?

(iv) If the average duration for activity F increases to 14 days, what will be its effect on the expected project completion time which will have 95% confidence?

17. Draw the network for the following project and compute the earliest and latest times for each event and also find the critical path :

Activity	1–2	1–3	2–4	3–4	4–5	4–6	5–7	6–7	7–8
Immediate	–	–	1–2	1–3	2–4	2–4, 3–4	4–5	4–6	6–7, 5–7
Timde (days)	5	4	6	2	1	7	8	4	3

18. A small project consist of seven activities, the details of which are given below :

Activity	Time estimates			Predecessor
	t_0	t_m	t_p	
A	3	6	9	None
B	2	5	8	None
C	2	4	6	A
D	2	3	10	B
E	1	3	11	B
F	4	6	8	C, D
G	1	5	15	E

Find the critical path. What is probability that the project will be completed by 18 weeks?

19. The following table gives data on normal time-cost and crash time-cost for a project :

Activity	Normal		Crash	
	Time (days)	Cost (Rs.)	Time (days)	Cost (Rs.)
1 – 2	6	650	4	1000
1 – 3	4	600	2	2000
2 – 4	5	500	3	1500
2 – 5	3	450	1	650
3 – 4	6	900	4	2000
4 – 6	8	800	4	3000
5 – 6	4	400	2	1000
6 – 7	3	450	2	800

The indirect cost per day is Rs. 100.

(i) Draw the network and identify the critical path.

(ii) What are the normal project duration and associated cost?

(iii) Crash the relevant activities systematically and determine the optimum project completion time and cost.

20. The following table gives the activities in a construction project and other relevant information :

Activity	Immediate predecessor	Time (days)		Direct cost (Rs.)	
		Normal	Crash	Normal	Crash
A	–	4	3	60	90
B	–	6	4	150	250
C	–	2	1	38	60
D	A	5	3	150	250
E	C	2	2	100	100
F	A	7	5	115	175
G	B, D, E	4	2	100	240

Indirect costs vary as follows:

	15	14	13	12	11	10	9	8	7	6
Cost (Rs.)	600	500	400	250	175	100	75	50	35	25

(i) Draw the network of the project.

(ii) Determine the project duration which will result in minimum total project cost.

16. SUGGESTED FURTHER READINGS

* Taha H.A., Operations Research – an Introduction, PHI.
* Bronson, R. and Naadimuthu. G., Theory and Problems of Coperations Research, Schuam's Outline Series, MGH.
* Swarup, K., Gupta, P.K. and Man Mohan, Operations Research, Sultan Chand & Sons.
* Sharma, J.K. Operations Research – Theory and Applications, Macmillan.
* Gupta, P.K. and Hira, D.S., Operations Research, S. Chand & Co. Ltd.

--- 0 ---

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper - X | Special Paper - OR

Module No. - 114

QUEUING THEORY

1.0 Introduction

A group of customers waiting to receive service including those receiving the service is known as a waiting line or queue. Queuing theory involves the mathematical study of queues or waiting lines. The formation of waiting lines (or queues) is a common phenomenon which occurs whenever the current demand for a service exceeds the current capacity to provide that service. The queues may be seen at a cinema ticket window, bank counter, Doctor's clinic, bus stop, reservation office, etc. for getting service. Some examples of queuing services are given in **Table-1**.

Table 1: Example of queuing service

Situation	Costumers or arrivals	Service facility or servers	Service process
Telephone exchange	Telephone calls	Switching board operators or electric switching operators	Computer
Departmental store	Customer	Checking clerks and bag packers	Bill payment
Bank	Customer	Bank clerk, check drawn. deposits (cash)	Deposit money withdrawn
Job interviewing	Applicant	Interviewee	Selection of the applicant

The person waiting in a queue or receiving the service is called the customer and the person by whom the customer is serviced is called a server. Queuing problem arises because the cost of service may not be so cheep as to ensure sufficient service facilities so that no customer has to wait. Customers arrive at a counter according to a certain probabilistic law (Poisson's input, Erlang input, etc.). On the other hand, the customers will be served by one or more server following a certain principle (FCFS i.e., first come first serve, random service etc.). By providing additional service facility, customers waiting time or queue can be reduced, conversely by decreasing the number of service facilities a long queue will be formed. This may result in loss of sales or customer. The service times are random variables govern by a given probabilistic law. After being served a customer leaves the queue. The objective of queuing model is to determine how to provide the service to the customers so as to minimize teh total cost of service and waiting time of customers by manipulating certain factors (variables) such as number of servers, rate of service and order of service.

Structure:

2. OBJECTIVES

After studying this module, the reader will be able to

* identify and examine the different situations which generate the queuing problems.
* learn the various factors/components of a queuing system and description of each of them.
* analyze a variety of performance measures (operating characteristics) of a queuing system.
* understand the different queuing models and derive the performance measures for each of them.

3. KEY WORDS

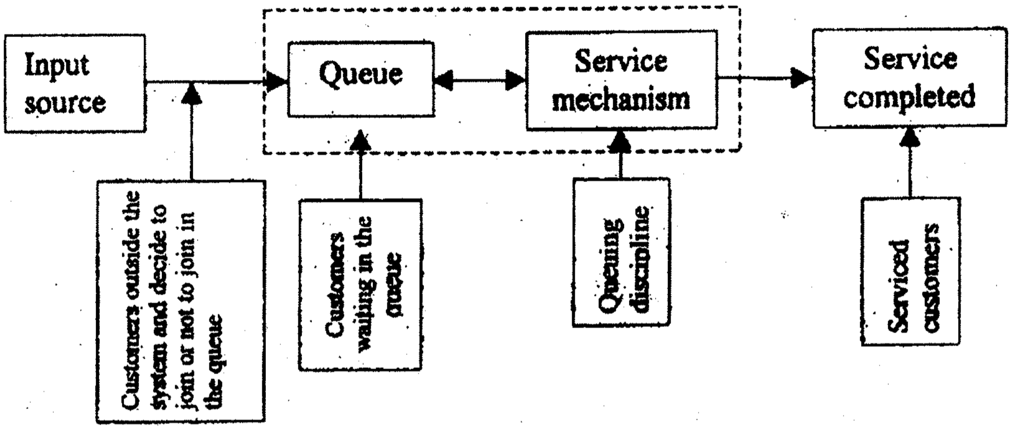
Queue, waiting time, customer, arrival pattern, service pattern, service mechanism, service facility, queue discipline, customer behaviour, balking, reneging, priorities, operating characteristics, transient state, steady state, probability distribution, pure birth process, inter-arrival times, Markovian property, pure death process, traffic intensity, Poisson queue, non-Poisson queue, queue length, busy period, Machine repair problem, break down, flow of arrival, effective arrival rate, idle period

4. CHARACTERISTICS OF QUEUING SYSTEM

Any queuing system can be characterized by the following set of characteristics:

* Input source
* Service mechanism
* Service discipline
* Capacity of the system.

The general frame work of a queuing system is shown below:



4.1 Input source

The input source is a device or group of devices that provides customers at the service facility for teh service. An input source is characterized by the following factors:

* Input Size
* Arrival pattern
* Behaviour of the arrivals

Input Size

If the total number of customers requiring service are only few then the input size is said to be finite. But if potential customers requiring service are sufficient by large in number, then the input source is considered to be infinite. Also if the customers arrived at the service facility in batches of fixed size or variable size instead one at a time, then the input is said to occur in bulk or batches. If the service is not available, they may be required to join the queue.

Arrival Pattern

If the pattern of which customers arrive at the service system or time between successive arrivals (inter-arrival time) is uncertain, then the arrival pattern is measured by either mean arrival rate or inter-arrival time. They are characterized in the form of the probability distribution associated with these random processes. The most common stochastic queuing model assumes that the arrival rate follows a Poisson distribution or equivalently inter-arrival time exponential distribution.

Customer's behaviour

The customers generally behave in different ways:

(a) Balking: A customer may leave the queue because the queue is too long and he/she has no time to wait or there is not sufficient waiting space.

(b) Reneging: A customer who becomes impatient after waiting in the queue for some time and leaevs the queue.

(c) Priorities: In certain applications some customers are served before others regardless of their order of arrival. These customers have priority over others.

(d) Customers may jockey from one queue to another hoping to receive service more quickly.

4.2 Service mechanism

The service mechanism is concerned with the manner in which customers are serviced. It can be characterized by observing the different factors:

(i) Availability of service facility: Service may be available only at certain time but not always.

(ii) Capacity or service facility: It is measured in terms of customers that can be served simultaneously. A queuing system may have one or more parallel service channels or sequence of channels in series as shown in the figures.

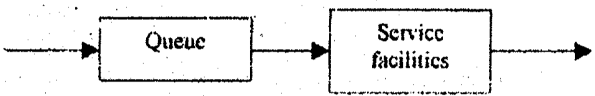


Fig. 1 : Single channel single phase

Fig. 2 :Single channel multiple phase

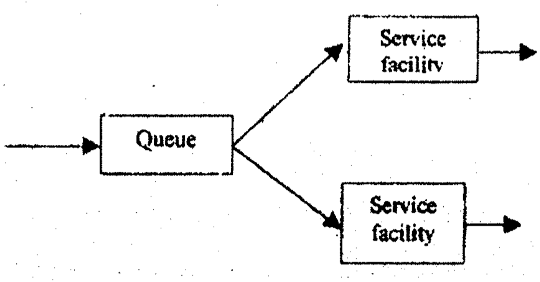


Fig. 3 : Multiple channel single phase

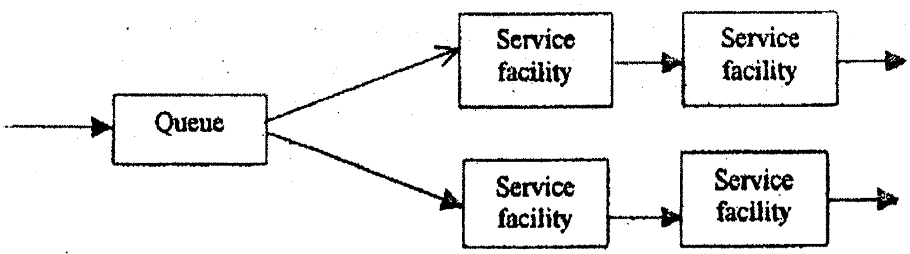


Fig. 4 :Multiple channel multiple phase

(iii) Service time or duration:

Service time or duration can be constant or a random variable. In general, service time is not constant, since it is a common practice to use overtime or extra effect when an excessive queue condition is present.

A variety of probability distribution can be used to characterize the service pattern. The service pattern is assumed to be independent of how customers arrive. However, it may not be true for sometimes of service facilities. The most common stochastic queuing model assumed that service time follows the exponential distribution or equivalently service rate follows a Poisson distribution.

4.3 Service Discipline

The service discipline refers to the order or manner in which customers in the queue will be served. The most common discipline are

(i) **First Come First Served (FCFS);** According to this discipline the customers are served in the order of their arrival. It is also known as FIFO (first input first output). This service discipline may be seen at a cinema ticket window, at a railway ticket window, etc.

(ii) **Last Come First Served (LCFS):** This discipline may be seen in big godown where the units (items) which come last are taken out (served) first.

(iii) **Serving in Random Order (SIRO):** In this case, the arrivals are serviced randomly in respective of their arrival in the system.

(iv) Service on some priority-procedure: Some customers are served before the others without considering their order of arrival i.e., some customers are served on priority basis. There are two types of priority services.

(a) Preemptive priority

In this case customers with the highest priority is allowed to enter service immediately and after entering into the system even of a customer with lower priority is already in service i.e., lower priority customer service is interrupted to start service for a special customer. This interrupted service is resumed again after the highest priority customer is served.

(b) Non-preemptive priority

In this case the highest priority customer goes ahead of the queue but service is started immediately on completion of the current customer service.

4.4 Capacity of the system

In certain cases a queuing system is unable to accommodate more than the required number of customers at a time. No further customers are allowed to enter until the space becomes available to accommodate new customers. Such types of situation are referred to as finite source queue.

5. IMPORTANT DEFINITIONS IN QUEUING PROBLEM

Queue length: Queue length is defined by the number of persons (customers) waiting in the line at any time.

Average length of queue: Average length of queue is defined by the number of customers in the queue per unit time.

Waiting time: It is the time upto which a unit has to wait in the queue before it is taken into service.

Servicing time: The time taken for servicing of a unit is called its servicing time.

Busy period: Busy period of a server is the time during which the server remains busy in servicing. Thus, it is the time between the starting of service of the first unit to the end of service of the last unit in the queue.

Idle period: When all the customers in the queue are served, the idle period of the server begins and it continues upto the time of arrival of the customer. Thus the idle period of a server is the time during which he remains free because there is no customer present in the system.

Mean arrival rate: The mean arrival rate in a queue is defined as the expected number of arrivals occurring in a time interval of unit length.

Mean servicing rate: The mean servicing rate for a particular servicing station is defined as the expected number of services completed in a time interval of unit length, given that the servicing is going on throughout the entire time unit.

Traffic intensity: In case of a simple queue the traffic intensity is the ratio of mean arrival and the mean servicing rate.

i.e. Traffic intensity $= \dfrac{\text{Mean arrival rate}}{\text{Mean servicing rate}}$

6. THE STATE OF THE SYSTEM

The state of the system involves the study of a system's behaviour over time. The states of a system may be classified as follows:

(i) **Transient State:** A system is said to be in transient state when its operating characteristics are dependent on time. Thus a queuing system is in transient state when the probability distributions of arrivals, waiting time and servicing time of the customers are dependent on time. This state occurs at the beginning of the operation of the system.

(ii) **Steady State:** A system is said to be in steady state when its operating characteristics become independent of time. Thus a queuing system is in steady state when the probability distributions of arrivals, waiting time and servicing time of the customers are independent on time.

Let $p_n(t)$ denotes the probability that there are n units in the system at time t, then if the probability $p_n(t)$ remains the same as time passes, the system acquires steady state. Mathematically, in steady state,

$$\lim_{t \to \infty} p_n(t) = p_n \text{ (independent of time)}$$

which implies $\lim_{t\to\infty} \frac{p_n(t)}{dt} = \frac{dp_n}{dt} = 0$

(iii) Explosive state: If the servicing rate is less than the arrival rate, the length of the queue will go on increasing with time and will tend to infinity as $t \to \infty$.

7. PROBABILITY DISTRIBUTION IN QUEUING SYSTEM:

In the queuing system, it is assumed that the customer's arrival is random and follows a Poisson distribution or equivalently the inter-arrival times obey exponential distribution. In most of the cases, service times are also assumed to be exponentially distributed. The basic assumptions (axioms) are as follows:

Basic assumption:

Axiom 1. The number of arrivals in non-over lapping intervals is statistically independent.

Axiom 2. The probability of more than one customers arrived in the time interval $(t, t+\Delta t)$ is negligible and is denoted by $O(\Delta t)$

i.e., $p_0(\Delta t)+p_1(\Delta t)+O(\Delta t)=1$

Axiom 3. The probability that a customer arrives in the time interval $(t,t+\Delta t)$ is equal to $\lambda\Delta t+\Delta(t)$ i.e., $p_1(\Delta t)=\lambda\Delta t+O(\Delta t)$

where λ is a constant and independent of the total number of arrivals up to the time t. Δt is small time interval and $O(\Delta t)$ denotes a quantity which is of smaller order of magnitude that Δt such that

$$\lim_{\Delta t\to 0}\left\{\frac{O(\Delta t)}{\Delta t}\right\}=0$$

Distribution of arrival (pure birth process)

The process in which only arrivals are counted and no departure takes place are called pure birth process. Stated in terms of queuing, birth-death process usually arises by birth or arrival in the system and decrease by death or departure of serviced customers from the system.

Property

If the arrivals are completely random then the probability distribution of number of arrivals in a fixed time interval follows Poisson distribution.

Distribution of inter-arrival times

Inter-arrival time is defined as the time interval between two successive time intervals.

Property of inter-arrival times

If the arrival process follows the Poisson distribution then $p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$ and associated random variable defined as inter-arrival time T follows the exponential distribution $f(t) = \lambda e^{-\lambda t}$ and conversely.

Markovian property of inter-arrival time

The Markovian property of inter-arrival times statess that the probability that a customer currently in service is compected at some time t is independent of how long he has already be in service i.e., prob $\{T \geq t_1 / T \geq t_0\}$ = prob $\{0 \leq T \leq t_1 - t_0\}$

where T is the time between the two successive arrival s.

Distribution of departure (Pure death process)

Sometimes a situation may arise when no additional customer joins the system while service is continued for those who are in line. Let at time t there be $N(\geq 1)$ customer in the system. It is clear that service will be provided at the rate of μ. Number of customers in the system at time $t \geq 0$ is equal to N minus total departure upto the time t. The distribution of departure can be obtained with the help of following basic axioms:

Basic axioms

(i) The probability of departure during the time Δt is $\mu \Delta t$.

(ii) The probability of more than one departure between the time t and $t + \Delta t$ is negligible.

(iii) The number of departure in non overlapping intervals are statistically independent i.e., the process has independent arrival.

8. CLASSIFICATION OF QUEUING MODELS

Generally queuing model may be specified in the symbolic form as $(a/b/c):(d/e/f)$

where a = arrival distribution i.e., type of arrival process

b = service time distribution i.e., type of service process

c = number of server

d = capacity of the system

e = service discipline

f = number of calling source capacity of input service

Generally, the last symbol f is not used. The standard notation for small a and b are taken as M which is called Poisson (or Markovian) arrival or departure distribution or equivalently exponential inter-arrival service time distribution.

E_k = Erlangian or Gamma inter-arrival for service time distribution with parameter k

G = General service time distribution or departure distribution

Thus $(M/E_k/1):(\infty/\text{FIFO}/\infty)$ defines a queuing system in which arrival follows the Poisson distribution, service time are Earlangian, single server, the capacity of the system is infinity i.e., system (queue + service) can hold infinite number of customers and finally the source generating arriving customers has an infinite capacity.

Generally queues with arrivals and departure start under transient condition and gradually reach steady state after a sufficient large time elapsed, provided that the parameters of the system permit reaching steady state.

Now we define some parameters for the steady state condition.

p_n = (steady state) probability of n customers in the system

L_s = expected number of customers in the system

L_q = expected number of customers in the queue

W_s = expected waiting time in the system

W_q = expected waiting time in the queue

By definiting, $L_s = \sum_{n=0}^{\infty} np_n$

$L_q = \sum_{n=c}^{\infty} (n-c) p_n$, c being the number of servers or channels.

8.1 Model I : $(M/M/1):(\infty/\text{FCFS}/\infty)$

This is a queuing model with Poisson arrival, Poisson service or departure, single servicing channel, the capacity of the system is infinite and the service discipline is first come, first serve.

Let λ and μ be the mean arrival rate and mean service rate of units respectively and n be the number of customers in the system.

Hence the probability of one arrival during time Δt is $\lambda\Delta t + O(\Delta t)$ and the probability of one departure or service during Δt is $\mu\Delta t + O(\Delta t)$

To have $n(>0)$ customers in the system in time $t+\Delta t$, there may be the following three cases

(i) n customers in the system at time t, no arrival in time Δt, no service in time Δt.

(ii) $(n-1)$ customers in the system at time t, one arrival in time Δt, no service in time Δt.

(iii) $(n+1)$ customers in the system at time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of n customers in the system at time $(t+\Delta t)$ is given by

$$p_n(t+\Delta t) = p_n(t)\{1-\lambda\Delta t + O(\Delta t)\}\{1-\mu\Delta t + O(\Delta t)\} + p_{n-1}(t)[\lambda\Delta t + O(\Delta t)][1-\mu\Delta t + O(\Delta t)]$$

$$+p_{n+1}(t)[1-\lambda\Delta t + O(\Delta t)][\mu\Delta t + O(\Delta t)] \text{for } n>0$$

$$= p_n(t)[1-\lambda\Delta t - \mu\Delta t] + p_{n-1}(t)\lambda\Delta t + p_{n+1}\mu\Delta t + O(\Delta t)$$

$$= p_n(t) + [-(\lambda+\mu)p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t)]\Delta t + O(\Delta t)$$

$$\therefore p_n(t+\Delta t) - p_n(t) = [\lambda p_{n-1}(t) - (\lambda+\mu)p_n(t) + \mu p_{n+1}(t)]\Delta t + O(\Delta t)$$

or $$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t} = \lambda p_{n-1}(t) - (\lambda+\mu)p_n(t) + \mu p_{n+1}(t) + \frac{O(\Delta t)}{\Delta t} \text{ for } n>0$$

Now taking limit as $\Delta t \to 0$ of both sides in the above we have

$$p_n'(t) = \lambda p_{n-1}(t) - (\lambda+\mu)p_n(t) + \mu p_{n+1}(t) \quad (1)$$

If there is no customer in the system in time $t+\Delta t$, there may be following two cases:

(i) no customer in the system in time t, no arrival in time Δt.

(ii) one customer in the system in time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of no customers in the system at time $t+\Delta t$ is given by

$$p_0(t+\Delta t) = p_0(t)\left[1-\lambda\Delta t + O(\Delta t)\right] + p_1(t)\left[1-\lambda_1\Delta t + O(\Delta t)\right]\left[\mu\Delta t + O(\Delta t)\right]$$

$$= p_0(t) - \lambda p_0(t)\Delta t + \mu p_1(t)\Delta t + O(\Delta t)$$

or, $$\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t} = -\lambda p_0(t) + \mu p_1(t)\frac{O(\Delta t)}{\Delta t}$$

Taking limit as $\Delta t \to 0$ on both sides in the above, we have

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t) \qquad (2)$$

Under the steady state condition of the system i.e., $\lim_{t\to\infty} p_n(t) = p_n$ and $\lim_{t\to\infty} p_n'(t) = 0$ equation (2) and (1) reduce to

$$-\lambda p_0 + \mu p_1 = 0 \quad (3)$$

$$\lambda p_{n-1} - (\lambda+\mu)p_n + \mu p_{n+1} = 0 \text{ for } n > 0 \qquad (4)$$

which are the steady state difference equation of the system.

Solution of the steady state difference equations

From (3), $-\lambda p_0 + \mu p_1 = 0$

which implies $p_1 = \frac{\lambda}{\mu}p_0 = \rho p_0 \quad \left[\because \frac{\lambda}{\mu} = \rho\right]$

Now putting $n = 1$ in (4) we have

$$\lambda p_0 - (\lambda+\mu)p_1 + \mu p_2 = 0$$

or, $\mu p_2 = -\lambda p_0 + (\lambda+\mu)p_1$

or, $$p_2 = -\frac{\lambda}{\mu}p_0 + \frac{\lambda+\mu}{\mu}p_1$$

or, $$p_2 = -\frac{\lambda}{\mu}p_0 + \left(\frac{\lambda}{\mu}+1\right)p_1$$

or, $p_2 = -\rho p_0 + (\rho+1)\rho p_0 = \rho^2 p_0$

Again putting $n=2$ in we have

$$\lambda p_1 - (\lambda+\mu)p_2 + \mu p_3 = 0$$

or, $\mu p_3 = -\lambda p_1 + (\lambda+\mu)p_2 = 0$

or, $p_3 = -\frac{\lambda}{\mu}p_1 + \left(\frac{\lambda}{\mu}+1\right)p_2$

$$= -\rho^2 p_0 + \rho^3 p_0 + \rho p_0 = \rho^3 p_0$$

Hence, $p_3 = \rho^3 p_0$

Proceeding in this way, we will have

$$p_n = \rho^n p_0$$

But, we have $\sum_{n=0}^{\infty} p_n = 1$

or, $p_0 + p_1 + p_2 + \ldots = 1$

or, $p_0 + \rho p_0 + \rho^2 p_0 + \rho^3 p_0 + \ldots = 1$

or, $p_0 \frac{1}{1-\rho} = 1$

or, $p_0 = 1-\rho$ (5)

Hence $p_n = \rho^n p_0 = \rho^n(1-\rho)$ (6)

This gives the probability that there are n units in the system at any time. Equations (5) and (6) rather give the required probability distribution of the queue length.

Analysis of Steady State results:

(i) Probability of queue size greater than or equal to N i.e. Prob (queue size $\geq N$)

Prob (queue size $\geq N$) = $p_N + p_{N+1} + p_{N+2} + \ldots..$

$$= \sum_{n=N}^{\infty} p_n = \sum_{n=0}^{\infty} p_n - \sum_{n=0}^{N-1} p_n$$

$$= 1 - (p_0 + p_1 + \ldots\ldots.. + p_{N-1})$$

$$= 1 - p_0(1 + \rho + \rho^2 + \ldots\ldots\ldots.. + \rho^{N-1})$$

$$= 1 - p_0 \frac{1-\rho^N}{1-\rho}$$

$$= 1-(1-\rho)\frac{1-\rho^N}{1-\rho} \qquad [\because p_0 = 1-\rho]$$

$$= \rho^N = \left(\frac{\lambda}{\mu}\right)^N$$

$\therefore$ Prob (queue size $\geq N$) $= \rho^N = \left(\frac{\lambda}{\mu}\right)^N$

(ii) Expected line length L_s

L_s = expected number of customers in the system i.e., expected line length or number of customers in the system

$$= \sum_{n=0}^{\infty} np_n = \sum_{n=0}^{\infty} np^n p_0 = p_0 \sum_{n=0}^{\infty} np^n$$

$$= p_0\left[\rho + 2\rho^2 + 3\rho^3 + 4\rho^4 + \ldots\right]$$

$$= p_0\rho\left[1 + 2\rho + 3\rho^2 + 4\rho^3 + \ldots\right]$$

$$= p_0\rho(1-\rho)^{-2} = (1-\rho)\rho\frac{1}{(1-\rho)^2} \qquad [\because p_0 = 1-\rho]$$

$$= \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda} \qquad \left[\because \rho = \frac{\lambda}{\mu}\right]$$

(iii) Expected queue length L_q

L_q = expected number of customers in queue (i.e., expected queue length)

$$= \sum_{n=1}^{\infty}(n-1)\, p_n$$ [since there are $(n-1)$ customers in the queue excluding one in service]

$$= \sum_{n=1}^{\infty} np_n - \sum_{n=1}^{\infty} p_n = L_s - \left(\sum_{n=0}^{\infty} p_n - p_0\right)$$

$$= L_s - (1-p_0) = L_s - \{1-(1-\rho)\}$$

$$= L_s - \rho = \frac{\rho}{1-\rho} - \rho\left[\because L_s = \frac{\rho}{1-\rho}\right]$$

$$= \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)} \quad \left[\because \rho = \frac{\lambda}{\mu}\right]$$

It is noted that $L_s = L_q + \rho = L_q + \frac{\lambda}{\mu}$ as $L_q = L_s - \rho$

(iv) Variance of queue length

From the definition of variance, $Var(n) = E\{n^2\} - \{E(n)\}^2$

$$= \sum_{n=1}^{\infty} n^2 p_n - \left\{\sum_{n=1}^{\infty} np_n\right\}^2$$

$$= \sum_{n=1}^{\infty} n^2 p^n (1-\rho) - \left(\frac{\rho}{1-\rho}\right)^2 \quad \left[\because L_s = \sum_{n=1}^{\infty} np_n = \frac{\rho}{1-\rho} \text{ and } p_n = \rho^n(1-\rho)\right]$$

$$= \rho(1-\rho) + 2^2\rho^2(1-\rho) + 3^2\rho^3(1-\rho) + 4^2\rho^4(1-\rho) + \ldots\ldots - \frac{\rho^2}{(1-\rho)^2}$$

$$= \rho(1-\rho)\left[1 + 2^2\rho + 3^2\rho^2 + 4^2\rho^3 \ldots\ldots\right] - \frac{\rho^2}{(1-\rho)^2}$$

$$= \rho(1-\rho)S - \frac{\rho^2}{(1-\rho)^2} \quad \text{where } S = 1 + 2^2\rho + 3^2\rho^2 + 4^2\rho^3 + \ldots\ldots$$

Now, we have to calculate the simplified expression of S.

$S = 1 + 2^2\rho + 3^2\rho^2 + 4^2\rho^3 + \ldots\ldots$

Now integrating both sides with respect to ρ from 0 to ρ, we have

$$\int_{\rho=0}^{\rho} S\,d\rho = \int_{\rho=0}^{\rho} \left(1 + 2^2\rho + 3^2\rho^2 + 4^2\rho^3 + \ldots\ldots\right) d\rho$$

$$= \rho + 2\rho^2 + 3\rho^3 + 4\rho^4 + 5\rho^5 + \ldots$$

$$= \rho(1-\rho)^{-2}$$

$$\therefore \int_{\rho=0}^{\rho} S\,d\rho = \rho(1-\rho)^{-2} = \frac{\rho}{(1-\rho)^2}$$

Differentiating both sides with respect to ρ, we have

$$S=\frac{1}{(1-\rho)^2}+\rho\frac{-2}{(1-\rho)^3}(-1)=\frac{1}{(1-\rho)^2}+\frac{2\rho}{(1-\rho)^3}=\frac{1-\rho+2\rho}{(1-\rho)^3}=\frac{1+\rho}{(1-\rho)^3}$$

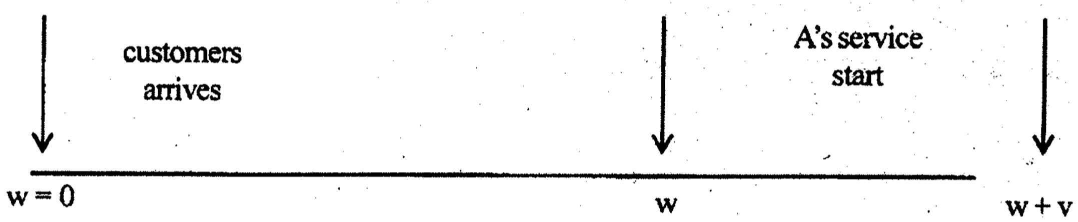
$$\therefore Var(n)=\rho(1-\rho)S-\frac{\rho^2}{(1-\rho)^2}$$

$$=\rho(1-\rho)\frac{1+\rho}{(1-\rho)^3}-\frac{\rho^2}{(1-\rho)^2}$$

$$=\frac{\rho(1+\rho)}{(1-\rho)^2}-\frac{\rho^2}{(1-\rho)^2}=\frac{\rho}{(1-\rho)^2}$$

(v) Probability density function of waiting time excluding the service time distribution.

In a steady state, each customer has the same waiting time distribution. Let this be a continuous function with probability density function(p.d.f.) $\psi(w)$ and we denote by $\psi(w)\,dw$. The probability that customers begins to be served in the interval $(w,\ w+dw)$, where w is measured from the time of his arrival. We suppose that a customer arrives at time $w=0$ and his service begins in the interval $(w,\ w+dw)$.



There may be two possibilities:

(i) There is a finite probability p_0 (the probability that the system is empty) that the waiting time is zero.

(ii) Let there be n customers in the system, $(n-1)$ waiting one in the service, when customer A arrive, therefore before the service of A begins $(n-1)$ customers must leave in the time interval $(0,\ w)$ and n-th customer in $(w, w+dw)$.

As the server's mean rate of service is μ in unit time, or μw in time w and the service distribution begin Poisson, we have

prob [$(n-1)$ customers waiting are served in time w]

$$=\frac{(\mu w)^{n-1}e^{-\mu w}}{(n-1)!}$$

and *prob* [one customer is served in time dw]

$\therefore \psi_n(w)\,dw$ = probability that a new arrival is taken into service after a time lying between w and $w+dw$.

= prob. [$(n-1)$ customers waiting are served in time w] $\times$ prob. [one customer is served in time dw]

$$=\frac{(\mu w)^{n-1}e^{\mu w}}{(n-1)!}\mu dw$$

Let W be the waiting time of the customer who has to wait such that

$$w\le W\le w+dw$$

Since the queue length can vary between 1 and ∞, therefore the probability density of the waiting time is given by

$$\psi(w)\,dw=p(w\le W\le w+dw)$$

$$=\sum_{n=1}^{\infty}[\text{probability that there are n customers in the system}]\times\psi_n(w)\,dw$$

$$=\sum_{n=1}^{\infty}\rho^n(1-\rho)\frac{(\mu w)^{n-1}e^{-\mu w}}{(n-1)!}\mu dw\left[\because p_n=\rho^n(1-\rho)\right]$$

$$=(1-\rho)e^{-\mu w}\mu dw\sum_{n=1}^{\infty}\rho^n\frac{(\mu w)^{n-1}}{(n-1)!}\quad=\rho(1-\rho)e^{-\mu w}\mu dw\sum_{n=1}^{\infty}\frac{(\rho\mu w)^{n-1}}{(n-1)!}$$

$$=\frac{\lambda}{\mu}\left(1-\frac{\lambda}{\mu}\right)e^{-\mu w}\mu dw\sum_{n=1}^{\infty}\frac{\left(\frac{\lambda}{\mu}\mu w\right)^{n-1}}{(n-1)!}\quad=\lambda\left(1-\frac{\lambda}{\mu}\right)e^{-\mu w}dw\sum_{n=1}^{\infty}\frac{(\lambda w)^{n-1}}{(n-1)!}$$

$$= \lambda\left(1-\frac{\lambda}{\mu}\right)e^{-\mu w}dw\left[1+\frac{\lambda w}{1!}+\frac{(\lambda w)^2}{2!}+........\right] = \lambda\left(1-\frac{\lambda}{\mu}\right)e^{-\mu w}.e^{\lambda w}dw$$

$$= \lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)w}dw \text{ where } w>0$$

Hence $\int_0^\infty \psi(w)\,dw = \int_0^\infty \lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)}dw = \lim_{B\to\infty}\int_0^B \lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)w}dw$

$$= \lambda\left(1-\frac{\lambda}{\mu}\right)\lim_{B\to\infty}\left[\frac{e^{-(\mu-\lambda)w}}{-(\mu-\lambda)}\right]_0^B$$

$$= \lambda\left(1-\frac{\lambda}{\mu}\right)\left[0+\frac{1}{(\mu-\lambda)}\right] = \lambda\frac{\mu-\lambda}{\mu}\frac{1}{\mu-\lambda} = \frac{\lambda}{\mu} \neq 1$$

This is because the case for which $w=0$ is included.

Thus $prob[w>0] = \int_0^\infty \psi(w)\,dw = \rho$

and $prob[w=0]$ = *prob* [no customer in the system] = $p_0 = 1-\rho$

Since the sum of these probabilities of waiting time is 1.

Therefore the complete distribution of the waiting time is

(a) continuous for $w \le W \le w+dw$ with probability density function $\psi(w)$ given by

$$\psi(w)\,dw = \lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)w}dw$$

(b) discrete for $w=0$ with Prob $(w=0) = 1-\frac{\lambda}{\mu}$

It is to note that the probability that the waiting time exceeds w_1 is

$$\int_0^\infty \psi(w)\,dw = \left[-\frac{\lambda}{\mu}e^{-(\mu-\lambda)w}\right]_{w_1}^\infty = \frac{\lambda}{\mu}e^{-(\mu-\lambda)w_1}$$

which does not include the service time.

(vi) Busy period distribution i.e., the conditional density function for waiting time, given that a person has to wait

Here we find out the probability density function for the distribution of total time (waiting + service) that an arrival spends in the system.

Let $\psi(w/w>0)$ = Probability density function for waiting time such that a person has to wait

$$=\frac{\psi(w)}{\text{prob }(w>0)}$$

$$=\frac{\psi(w)}{\rho} \qquad \left[\begin{array}{l}\because \text{ prob }(w>0)+\text{ prob }(w=0)=1\\ \therefore \text{ prob }(w>0)=1-\text{ prob }(w=0)=1-(1-\rho)=\rho\end{array}\right.$$

$$=\frac{\lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)w}}{\frac{\lambda}{\mu}}=(\mu-\lambda)e^{-(\mu-\lambda)w} \qquad (10)$$

Now $\int_0^\infty \psi(w/w>0)=\int_0^\infty(\mu-\lambda)e^{-(\mu-\lambda)w}dw=1$

Thus (10) gives the required probability density function for the busy period.

(vii) Mean or expecting waiting time in the queue i.e. w_q (average waiting time) of an arrival in the queue.

We have, $w_q=\int_0^\infty w\psi(w)\,dw$

$$\int_0^\infty w\lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)w}dw$$

$$=w\lambda\left(1-\frac{\lambda}{\mu}\right)\frac{e^{-(\mu-\lambda)w}}{-(\mu-\lambda)}\Bigg]_0^\infty+\lambda\int_0^\infty\left(1-\frac{\mu}{\lambda}\right)\frac{e^{-(\mu-\lambda)w}}{\mu-\lambda}dw \text{ [Integrating by parts taking } w \text{ as first function]}$$

$$=0-0+\lambda\left(1-\frac{\lambda}{\mu}\right)\frac{e^{-(\mu-\lambda)w}}{-(\mu-\lambda)^2}\Bigg]_0^\infty$$

$$= -\lambda\left(\frac{\mu-\lambda}{\mu}\right)\frac{1}{(\mu-\lambda)^2}e^{-(\mu-\lambda)w}\Bigg]_0^\infty$$

$$= 0+\frac{\lambda(\mu-\lambda)}{\mu(\mu-\lambda)^2}1 = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{1}{\lambda}L_q, \text{ where } L_q = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

(vii) Expected waiting time in the system W_s [i.e., average time that an arrival spends in the system

W_s = expected waiting time in the system

= expected waiting time in queue + expected service time

$= w_q + \frac{1}{\mu}$ [Since the expected service time = mean service time = $\frac{1}{\mu}$]

$$= \frac{\lambda}{\mu(\mu-\lambda)} + \frac{1}{\mu}\quad\left[\because w_q = \frac{\lambda}{\mu(\mu-\lambda)}\right]$$

$$= \frac{\lambda+\mu-\lambda}{\mu(\mu-\lambda)} = \frac{\mu}{\mu(\mu-\lambda)} = \frac{1}{\mu-\lambda}$$

As $W_s = \frac{1}{\mu-\lambda}$ and $L_s = \frac{\lambda}{\mu-\lambda}, L_s = \lambda W_s$

(ix) Expected length of non-empty queue i.e., average length of non empty queue i.e., $L / L > 0$

We have $(L/L>0) = \frac{L_q}{\text{prob [an arrival has to wait, } L>0]} = \frac{L_q}{\text{prob }[(n-1)>0]} = \frac{L_q}{\sum_{n=2}^{\infty} p_n}$

$$= \frac{L_q}{\sum_{n=0}^{\infty} p_n - (p_0+p_1)} = \frac{L_q}{1-p_0-p_1} = \frac{L_q}{1-p_0-\rho p_0} = \frac{L_q}{1-p_0(1+\rho)}$$

$$= \frac{\frac{\rho^2}{1-\rho}}{1-(1-\rho)(1+\rho)}$$

$$= \frac{\frac{\rho^2}{1-\rho}}{1-\left(1-\rho^2\right)} = \frac{1}{1-\rho} = \frac{1}{1-\frac{\lambda}{\mu}} = \frac{\mu}{\mu-\lambda}$$

Example 1

Arrivals at a telephone booth are considered to be Poisson with an average time of 10 minutes between one arrival and the next. The length of a phone call is assumed to be distributed exponentially with mean 3 minutes.

(a) What is the probability that a person arriving at the booth will have to wait?

(b) What is the average length of queues that form from time to time?

(c) The telephone company will install a second booth when convinced that an arrival would expect to have to wait at least 3 minute for the phone. By how much must the flow of arrivals be increased to justify a second booth?

(d) Find the average number of units in the system.

(e) What is the probability that an arrival has to wait more than 10 minutes before the phone is free?

(f) Estimate the fraction of a day that the phone will be in use (or busy).

Solution:

This is a $(M/M/1):(\infty/FCFS/\infty)$ problem.

In this problem, the mean arrival rate, $\lambda = \frac{1}{10}$ and mean service time, $\mu = \frac{1}{3}$

(a) Now the probability that a person arriving at the telephone booth will have to wait

$= \text{Prob}[w>0] = 1-\text{Prob}(w=0) = 1-p_0 = 1-(1-\rho) = \rho = \frac{\lambda}{\mu} = 0.3$

(b) Average length of the queues that form from time to time

$$= L/L>0 = \frac{\mu}{\mu-\lambda} = 1.43 \text{ persons}$$

(c) The installation of second booth will be justified if the arrival rate is greater than the waiting time. Then the length of queue will go on increasing.

In that case, let λ' be the mean arrival rate.

We know that the average waiting time of an interval in the queue is $w_q = \frac{\lambda'}{\mu(\mu-\lambda')}$

Here, $w_q = 3, \mu = \frac{1}{3}$ $\therefore \frac{\lambda'}{\frac{1}{3}\left(\frac{1}{3}-\lambda'\right)}$ or, $\lambda' = \frac{1}{3} - \lambda'$ or, $\lambda' = \frac{1}{6}$

Hence, the increase in mean arrival rate $= \lambda' - \lambda = \frac{1}{6} - \frac{1}{10} = \frac{1}{15} = 0.067$ arrivals per minute

Again, the increase in the flow of arrivals $= \frac{1}{15} \times 60 = 4$ per hour.

Therefore, the second booth is justified in the increase in arrival rate is 0.067 persons per minute (or, 4 persons per hour).

(d) Average number of units in the system is $L_s = \frac{\lambda}{\mu-\lambda} = 0.43$ persons.

(e) prob (waiting time of an arrival in queue > 10]

$$= \int_{10}^{\infty} \psi(w)\,dw = \int_{10}^{\infty} \frac{\lambda}{\mu}(\mu-\lambda)e^{-(\mu-\lambda)w}dw = \lim_{B\to\infty} \frac{\lambda}{\mu}(\mu-\lambda)\left.\frac{e^{-(\mu-\lambda)w}}{-(\mu-\lambda)}\right]_{10}^{B}$$

$$= -\frac{\lambda}{\mu}\left[0 - e^{-(\mu-\lambda)10}\right] = 0.03(\text{approx})$$

(f) The fraction of a day that the phone will be busy = traffic intensity $= \rho = \frac{\lambda}{\mu} = 0.3$.

8.2 Model II (M / M / c) : (∞ / FCFS / ∞)

This is the queuing model with Poission arrival, Poisson departure. There are $c(>1)$ channels, service rate at each channel is the same. The service discipline is first come, first service. In this model the length of the waiting line will depend on the member of occupied channels.

Let n be the number of customers in the system. If $n < c$, there will be no customer waiting in the queue as all of them will be served simultaneously and in that case $(c-n)$ service channels will remain idle. In this case the mean rate of service is $\mu_n = n\mu$ where μ is the mean service rate of unit.

If $n = c$, then all the service channels will be busy, each putting out a mean service rate μ and thus the mean rate of service is $\mu_n = c\mu$.

If $n = c$, then all the service channels will be busy while $n-c$ customers will be waiting in queue and the mean rate of service is $\mu_n = c\mu$.

Hence, for this model,

$$\lambda_n = \lambda,\ n = 0,1,2,....$$

$$\mu_n = \begin{cases} n\mu & 0 \le n < c \\ c\mu & n \ge c \end{cases}$$

i.e., the probability of one arrival during Δt is $\lambda\,.\,\Delta t + 0(\Delta t)$ and the probability of one departure during Δt is

$$n\mu\Delta t + O(\Delta t) \text{ if } 0 < n < c \text{ and } c\mu\Delta t + O(\Delta t) \text{ if } n \ge c$$

To have n customers in the system at time $t + \Delta t$, there may be following three cases:

(i) n customers in the system at time t, no arrival in time Δt, no service in time Δt.

(ii) $(n-1)$ customers in the system at time t, one arrival in time Δt, no service in time Δt.

(iii) $(n+1)$ customers in the system at time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of n customers in the system at time $(t + \Delta t)$ is given by

$$p_n(t+\Delta t) = p_n(t)(1-\lambda\Delta t)(1-\mu_n\Delta t) + p_{n-1}(t)\,\lambda\Delta t(1-\mu_{n-1}\,\Delta t) + p_{n+1}(t)(1-\lambda\Delta t)\,\mu_{n+1}\Delta t + O(\Delta t)$$

or, $$p_n(t+\Delta t) = p_n(t)\left[1-(\lambda+\mu_n)\Delta t\right] + p_{n-1}(t)\,\lambda\Delta t + p_{n+1}(t)\,\mu_{n+1}p_{n+1}\Delta t + O(\Delta t)$$

or, $$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t}=\lambda p_{n-1}(t)-(\lambda+\mu_n)p_n(t)+\mu_{n+1}p_{n+1}(t)+\frac{O(\Delta t)}{\Delta t} \quad (1)$$

(a) When $0<n<c$, then $\mu_n = n\mu,\quad \mu_{n+1}=(n+1)\mu$

Hence from (1) we have

$$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t}=\lambda p_{n-1}(t)-(\lambda+n\mu)p_n(t)+(n+1)\mu p_{n+1}(t)+\frac{O(\Delta t)}{\Delta t} \quad (2)$$

(b) When $n\geq c$, then $\mu_n = c\mu,\quad \mu_{n+1}=c\mu$

Hence from (1), we have

$$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t}=\lambda p_{n-1}(t)-(\lambda+c\mu)p_n(t)+c\mu p_{n+1}(t)+\frac{O(\Delta t)}{\Delta t} \quad (3)$$

(c) When $n=0$

To have zero customer in the system in time $t+\Delta t$, there may be following two cases

(i) no customer in the system in time t, no arrival in time Δt.

(ii) one customer in the system in time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of no customer in the system at time $(t+\Delta t)$ is given by

$$p_0(t+\Delta t)=p_0(t)(1-\lambda\Delta t)+p_1(t)(1-\lambda\Delta t)\mu\Delta t+O(\Delta t)$$

or, $$p_0(t+\Delta t)-p_0(t)=-\lambda p_0(t)\Delta t+p_1(t)\mu\Delta t+O(\Delta t)$$

or, $$\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t}=-\lambda p_0(t)+\mu p_1(t)+\frac{O(\Delta t)}{\Delta t} \quad (4)$$

Now taking limit $\Delta t\to 0$, the differential difference equation for this system obtained from (2), (3) and (4) are as follows:

$p_0'(t)=-\lambda p_0(t)+\mu p_1(t)$ for $n=0$

$p_n'(t)=\lambda p_{n-1}(t)+(\lambda+n\mu)p_n(t)+(n+1)\mu p_{n+1}(t)$, for $0<n<c$

$p_n'(t)=\lambda' p_{n-1}(t)-(\lambda+c\mu)p_n(t)+c\mu p_{n+1}(t)$, for $n\geq c$

Under the steady state condition of the system i.e. $\lim_{t\to\infty} p_n(t)=p_n$ and $\lim_{t\to\infty} p_n'(t)=0$, the above three equation reduce to

$$-\lambda p_0 + \mu p_1 = 0 \text{ for } n = 0 \qquad (5)$$

$$\lambda p_{n-1} - (\lambda + n\mu) p_n + (n+1)\mu p_{n+1} = 0 \text{ for } 0 < n < c \qquad (6)$$

$$\lambda p_{n-1} - (\lambda + c\mu) p_n + c\mu p_{n+1} = 0 \text{ for } n \geq c \qquad (7)$$

which are the steady state differential difference equation of the system.

Solution of the steady state equation:

From (5), $\qquad -\lambda p_0 + \mu p_1 = 0$ or, $\mu p_1 = \lambda p_0$ or, $p_1 = \frac{\lambda}{\mu} p_0$

Now putting $n = 1$ in (6), we have

$$\lambda p_0 - (\lambda + \mu) p_1 + 2\mu p_2 = 0$$

$$\text{or, } 2\mu p_2 = (\lambda + \mu) p_1 - \lambda p_0$$

$$\text{or, } p_2 = \frac{\lambda + \mu}{2\mu} p_1 - \frac{\lambda}{2\mu} p_0$$

$$= \frac{\lambda}{2\mu} p_1 + \frac{p_1}{2} - \frac{1}{2} p_1$$

$$= \frac{\lambda}{2\mu} p_1 = \frac{1}{(2)!}\left(\frac{\lambda}{\mu}\right)^2 p_0 \left[\because p_1 = \frac{\lambda}{\mu} p_0\right]$$

$$\therefore p_2 = \frac{1}{(2)!}\left(\frac{\lambda}{\mu}\right)^2 p_0$$

Again putting $n = 2$ in (6), we have

$$\lambda p_1 - (\lambda + 2\mu) p_2 + 3\mu p_3 = 0$$

$$\text{or, } p_3 = \frac{\lambda + 2\mu}{3\mu} p_2 - \frac{\lambda}{3\mu} p_1$$

$$= \frac{\lambda}{3\mu} p_2 + \frac{2}{3} p_2 - \frac{2}{3} \frac{\lambda}{2\mu} p_1$$

$$= \frac{\lambda}{3\mu} p_2 + \frac{2}{3} p_2 - \frac{2}{3} p_2 \quad \left[\because p_2 = \frac{\lambda}{2\mu} p_1\right]$$

$$= \frac{\lambda}{3\mu} p_2 = \frac{1}{(3)!}\left(\frac{\lambda}{\mu}\right)^2 p_0$$

$$p_3 = \frac{1}{3}\frac{\lambda}{\mu} p_2 = \frac{1}{(3)!}\left(\frac{\lambda}{\mu}\right)^3 p_0$$

Hence we have $p_n = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n p_0$ for $1 \le n < c$

$$\therefore p_{c-1} = \frac{1}{(c)!}\left(\frac{\lambda}{\mu}\right)^{c-1} p_0 = \frac{1}{c-1}\frac{\lambda}{\mu} p_{c-2} \qquad (8)$$

Now putting $n = c - 1$ in (6), we have

$$\lambda p_{c-2} - \left[\lambda + (c-1)\mu\right] p_{c-1} + c\mu p_c = 0$$

$$\text{or, } p_c = \frac{\lambda + (c-1)\mu}{c\mu} p_{c-1} - \frac{\lambda}{c\mu} p_{c-2}$$

$$= \frac{\lambda}{c\mu} p_{c-1} + \frac{c-1}{c} p_{c-1} - \frac{c-1}{c}\frac{\lambda}{(c-1)\mu} p_{c-2}$$

$$= \frac{\lambda}{c\mu} p_{c-1} + \frac{c-1}{c} p_{c-1} - \frac{c-1}{c} p_{c-1} \quad [\text{by } (8)]$$

$$= \frac{1}{(c)!}\left(\frac{\lambda}{\mu}\right)^c p_0 = \frac{(c\rho)^c}{(c)!} p_0 \text{ where } \rho = \frac{\lambda}{c\mu}$$

Putting $n = c$ in (7) we have

$$\lambda p_{c-1} - (\lambda + c\mu) p_c + c\mu p_{c+1} = 0$$

$$\text{or, } p_{c-1} = \frac{\lambda + c\mu}{\mu} p_c - \frac{\lambda}{c\mu} p_{c-1}$$

$$= \frac{\lambda}{c\mu} p_c + p_c - p_c \qquad \left[\because p_c = \frac{\lambda}{c\mu} p_{c-1} \right]$$

$$\frac{\lambda}{c\mu} p_c = \frac{1}{c}\left(\frac{\lambda}{\mu}\right) p_c = \frac{1}{c(c)!}\left(\frac{\lambda}{\mu}\right)^{c+1} p_0$$

Similarly $p_{c+2} = \frac{1}{c^2(c)!}\left(\frac{\lambda}{\mu}\right)^{c+2} p_0$

In general, $p_n = \frac{1}{c^{n-c}(c)!}\left(\frac{\lambda}{\mu}\right)^n p_0$ for $n \geq c$

$$p_n = \frac{c^c}{c^n(c)!}\left(\frac{\lambda}{\mu}\right)^n p_0 \text{ for } n \geq c$$

Therefore $p_n = \frac{1}{(n)!}\left(\frac{\lambda}{\mu}\right)^n p_0$ for $1 \leq n < c$

$$= \frac{c^c}{c^n(c)!}\left(\frac{\lambda}{\mu}\right)^n p_0 \text{ for } n \geq c$$

Determination of p_0:

We have $\sum_{n=0}^{\infty} p_n = 1$

or, $p_0 + \sum_{n=1}^{c-1} p_n + \sum_{n=c}^{\infty} p_n = 1$

or, $p_0 + \sum_{n=1}^{c-1} \frac{1}{(n)!}\left(\frac{\lambda}{\mu}\right)^n p_0 + \sum_{n=c}^{\infty} \frac{c^c}{c^{n-a}(c)!}\left(\frac{\lambda}{\mu}\right)^n p_0 = 1$

or, $p_0\left[\sum_{n=1}^{c-1} \frac{1}{(n)!}\left(\frac{\lambda}{\mu}\right)^n + \frac{c^c}{(c)!}\sum_{n=c}^{\infty}\left(\frac{\lambda}{c\mu}\right)^n\right] = 1$

or, $p_0\left[\sum_{n=1}^{c-1} \frac{1}{(n)!}(c\rho)^n + \frac{c^c}{(c)!}\sum_{n=c}^{\infty}\rho^n\right] = 1$ where $\rho = \frac{\lambda}{c\mu}$

or, $$p_0\left[\sum_{n=0}^{c-1}\frac{(c\rho)^n}{(n)!}+\frac{c^c}{(c)!}\rho^c\left(1+\rho+\rho^2+......\infty\right)\right]=1$$

or, $$p_0=\left[\sum_{n=0}^{c-1}\frac{(c\rho)^n}{(n)!}+\frac{(\rho c)^c}{(c)!}\frac{1}{1-\rho}\right]=1$$

or, $$p_0=\frac{1}{\left[\sum_{n=0}^{c-1}\frac{(c\rho)^n}{(n)!}+\frac{(c\rho)^c}{(c)!}\frac{1}{1-\rho}\right]} \qquad (9)$$

Hence

$$p_n=\begin{cases}\frac{1}{(n)!}\left(\frac{\lambda}{\mu}\right)^n=\frac{(c\rho)^n}{(n)!} & \text{for } 1\le n<c \text{ where } \rho=\frac{\lambda}{c\mu}\\ \frac{c^c}{c^n(c)!}\left(\frac{\lambda}{\mu}\right)^n p_0=\frac{c^c}{(c)!}(\rho)^n p_0 & \text{for } n\ge c \text{ where } \rho=\frac{\lambda}{c\mu}\end{cases}$$

where p_0 is given by (9)

(i) Expected queue length (average number of customers in the queue).

If $n>c$, a queue of n customers would consist of c customers being served together with a genuine queue of $n-c$ waiting customers, hence

$$L_q=\sum_{n=c+1}^{\infty}(n-c)p_n=\sum_{n=c}^{\infty}(n-c)p_n$$

$$=\sum_{n=c}^{\infty}(n-c)\frac{c^c\rho^n}{(c)!}p_0 \qquad \left[\because p_n=\frac{c^c}{(c)!}\rho^n p_0 \text{ where } \rho=\frac{\lambda}{c\mu}\right]$$

$$=\frac{c^c\rho^c}{(c)!}p_0\sum_{n=c}^{\infty}(n-c)\rho^{n-c}$$

$$=\frac{(c\rho)^c}{(c)!}p_0\left[\rho+2\rho^2+3\rho^3+......\right]$$

$$=p_c\frac{\rho}{(1-\rho)^2}$$

$$\therefore L_q = p_c \frac{\rho}{(1-\rho)^2}$$

Particular Case: When $c = 1$ i.e. system with one channel only.

$$L_q = \frac{\rho}{(1-\rho)^2} p_1 = \frac{\rho}{(1-\rho)^2} \rho(1-\rho) = \frac{\rho^2}{1-\rho}$$

(ii) Average number of customers in the system

We know that $L_s = L_q + \dfrac{\lambda}{\mu}$

$$\therefore L_s = \frac{\rho p_c}{(1-\rho)^2} + \frac{\lambda}{\mu} = \frac{\rho p_c}{(1-\rho)^2} + c\rho \text{ where } \rho = \frac{\lambda}{c\mu}$$

(iii) Average waiting time in the system

Average waiting time in the system is

$$W_s = \frac{\text{Average number of customers in the system}}{\text{rate of arrival}}$$

$$= \frac{L_s}{\lambda} = \frac{\rho p_c/(1-\rho)^2 + c\rho}{\lambda} = \frac{\rho p_c}{\lambda(1-\rho)^2} + \frac{c\rho}{\lambda}$$

Average waiting time in the queue

i.e., $W_q = \dfrac{\text{Average number of customers in the queue}}{\text{rate of arrival}}$

$$= \frac{L_q}{\lambda} = \frac{\rho p_c}{\lambda(1-\rho)^2}$$

(iv) Expected length of non empty queue i.e., to find $L / L > 0$

$$(L / L > 0) = \frac{L_q}{\text{Prob[an arrival has to wait]}}$$

$$= \frac{L_q}{\text{prob}(n > c)} = \frac{\rho p_c/(1-\rho)^2}{\sum_{n=c+1}^{\infty} p_n} = \frac{\rho p_c/(1-\rho)^2}{\sum_{n=c+1}^{\infty} \frac{c^c}{\lfloor c} p_n p_0}$$

$$= \frac{\rho p_c/(1-\rho)^2}{\frac{(c\rho)^2}{\lfloor c} p_0 \sum_{n=c+1}^{\infty} \rho^{n-c}} = \frac{\rho p_c/(1-\rho)^2}{p_c(\rho+\rho^2+\rho^3+\ldots)} = \frac{\rho/(1-\rho)^2}{\rho(1+\rho+\rho^2+\ldots)} = \frac{\rho/(1-\rho)^2}{\frac{\rho}{1-\rho}} = \frac{1}{1-\rho}$$

(v) Probability that a customer has to wait

Probability that a customer has to wait

$$= \text{prob}\,(n \geq c) = \sum_{n=c}^{\infty} p_n = \sum_{n=c}^{\infty} \frac{c^c}{\lfloor c} \rho^n p_0 = \frac{c^c}{\lfloor c} p_0 \sum_{n=c}^{\infty} \rho^n = \frac{c^c}{\lfloor c} p_0 \left(\rho^c + \rho^{c+1} + \ldots\right)$$

$$= \frac{c^c}{\lfloor c} p_0 \rho^c \left(1+\rho+\rho^2+\ldots\right) = \frac{(\rho c)^c}{\lfloor c} p_0 \frac{1}{1-\rho} = p_c \frac{1}{1-\rho} = \frac{p_c}{1-\rho} \text{ since } p_c = \frac{(\rho c)^c}{\lfloor c}$$

(vi) Probability that an arrival a customer will not have to wait

$$= 1 - \text{prob}(n \geq c) = 1 - \frac{p_c}{1-\rho}$$

Probability that all channels will be occupied

$$= \text{prob}(n \geq c) = \frac{p_c}{1-\rho}$$

Example 2

A telephone exchange has two long distance operators. The telephone company finds that, during the peak load, long distance all arrive in a Poisson fashion at an average rate of 15 per hour. The length of service on these calls is approximately exponentially distributed with mean length 5 minutes.

(a) What is the probability that a subscriber will have to wait for this long distance call during the peak hours of the day?

(b) If the subscriber waits and are serviced in turn, what is the expected waiting time.

Solution:-

This problem is of the model $(M/M/C):(\infty/FCFS/\infty)$

Here $c = 2$ and $\lambda = 15$ cells / hour $= \frac{15}{60} = \frac{1}{4}$ calls/min and $\mu = \frac{\lambda}{c\mu} = \frac{\frac{1}{4}}{2\cdot\frac{1}{5}} = \frac{5}{8}$ and $c\rho = 2\cdot\frac{5}{8} = \frac{5}{4}$

Now, $$p_0 = \frac{1}{\sum_{n=0}^{c-1}\frac{(c\rho)^n}{\lfloor n} + \frac{(c\rho)^c}{\lfloor c(1-\rho)}} = \frac{1}{\sum_{n=0}^{1}\frac{\left(\frac{5}{4}\right)^n}{\lfloor n} + \frac{\left(\frac{5}{4}\right)^n}{\lfloor 2\left(1-\frac{5}{8}\right)}} = \frac{1}{1+\frac{\frac{5}{4}}{\lfloor 1} + \frac{\frac{25}{16}}{2\cdot\frac{3}{8}}} = \frac{3}{13}$$

(a) Probability that a subscriber will have to wait for his long distance call

$$= \text{prob } (n \geq c) = \frac{p_c}{1-\rho} = \frac{\frac{(c\rho)^c}{\lfloor c}p_0}{1-\rho} = \frac{(c\rho)^c p_0}{\lfloor c(1-\rho)} = \frac{\left(\frac{5}{4}\right)^2\cdot\frac{3}{13}}{\lfloor 2\left(-\frac{5}{8}\right)} = \frac{25}{52} = 0.48$$

(b) Now expected waiting time

$$= W_q = \frac{L_q}{\lambda} = \frac{\rho p_c}{\lambda(1-\rho)^2} = \frac{\rho}{\lambda(1-\rho)^2}\cdot\frac{(c\rho)^2}{\lfloor c}p_0 = \frac{\frac{5}{8}}{\frac{1}{4}\left(1-\frac{5}{8}\right)^2}\cdot\frac{\left(\frac{5}{4}\right)^2}{\lfloor 2}\cdot\frac{3}{13}$$

$$= \frac{\frac{5}{8}}{\frac{1}{4}\cdot\left(\frac{3}{8}\right)^2}\cdot\frac{\frac{25}{16}}{2}\cdot\frac{3}{13} = \frac{125}{39} = 3.2 \text{ minutes}$$

Example 3

A super market has two girls ringing up sells at the counters. If the service time for each customer exponential with mean 4 minutes and if the people arrive in a poisson fashion at the counter at the rate 10 per hour, then

(a) What is the probability of having to wait for service?

(b) What is the expected percentage of idle time for each girl?

(c) If a customer has to wait, what is the expected length of his waiting time?

Solution:-

This problem is of the model $(M/M/2):(\infty/\text{FCFS}/\infty)$

Here $c = 2\lambda = \frac{10}{60} = \frac{1}{6}$ people per minute, $\mu = \frac{1}{4}$ people per minute

$$\therefore \rho = \frac{\lambda}{c\mu} = \frac{\frac{1}{6}}{2\cdot\frac{1}{4}} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{2}{6} = \frac{1}{3}$$

$$\therefore c\rho = 2\cdot\frac{1}{3} = \frac{2}{3}$$

Now $$p_0 = \frac{1}{\sum_{n=0}^{c-1}\frac{(c\rho)^2}{\lfloor n} + \frac{(c\rho)^2}{\lfloor c(1-\rho)}} = \frac{1}{\sum_{n=0}^{1}\frac{\left(\frac{2}{3}\right)^n}{\lfloor n} + \frac{\left(\frac{2}{3}\right)^2}{\lfloor 2\left(1-\frac{1}{3}\right)}} = \frac{1}{1+\frac{\frac{2}{3}}{\lfloor 1} + \frac{\left(\frac{2}{3}\right)^2}{2\cdot\frac{2}{3}}} = \frac{1}{2}$$

(a) Now the probability of having to wait for service

$$= \text{prob}\,(n \ge 2) = \sum_{n=2}^{\infty} p_n = \frac{p_c}{1-\rho} = \frac{\frac{(c\rho)^c}{\lfloor c}\,p_0}{1-\rho} = \frac{\frac{\left(\frac{2}{3}\right)^2}{\lfloor 2}\cdot\frac{1}{2}}{1-\frac{1}{3}} = 0.167$$

(b) Expected number of girls who are idle

$$= 2\cdot p_0 + 1\cdot p_1 = 2\cdot\frac{1}{2} + 1\cdot\frac{1}{3} = 1 + \frac{1}{3} = \frac{4}{3}\left[p_1 = \frac{\lambda}{\mu}p_0 = \frac{\frac{1}{6}}{\frac{1}{4}}\cdot\frac{1}{2} = \frac{1}{3}\right]$$

Now the probability of any girl being idle

$$= \frac{\text{expected number of idle girls}}{\text{total number of girls}} = \frac{\frac{4}{3}}{2} = \frac{2}{3} = 0.67$$

Hence expected percentage of idle time for each girl $= 0.67 \times 100\% = 67\%$

(c) Expected length of the customer waiting time on the condition that the customer has to wait

$$= W / W > 0 = \frac{1}{\text{Mean service rate} - \text{mean arrival rate for } n \geq c} = \frac{1}{c\mu - \lambda} = \frac{1}{2\mu - \lambda} = 3$$

Note: We know that $L_s = \lambda \; W_s$ and $L_q = \lambda W_q$

Where λ is the mean arrival rate (given). These equations hold under general conditions that restrict the distribution of arrivals of service time however in the special case where customer arrive at the rate λ but not all arrivals can join the sysetm, the equation, the equations above are modified by redefining λ to include only those customers that actually join the system. Thus letting λ_{eff} = effective arrival rate for those who join the system we can have, for this type of system.

$L_s = \lambda_{eff} W_s$ and $L_q = \lambda_{eff} \dot{W}_s$

8.3 Model III: $(M / M / 1):(N / FCFS / \infty)$

This is a queuing model with Poisson arrival, Poisson service or departure, single channel, capacity of the system equal to N (fixed) and the service discipline is first come first serve.

Let n be the customers in the system. Let λ and μ be mean arrival and service rate of units respectively. In this model,

$$\lambda_n = \begin{cases} \lambda \text{ when } n < N \text{ i.e., } (n = 0,1,2,...N-1) \\ 0 \text{ when } n \geq N \end{cases}$$

and $\mu_n = \mu$ (independent of n)

Therefore the probability of one arrival during Δt is $\lambda_n \Delta t + O(\Delta t)$ for $\Delta t \; n \geq N$ and the probability of one departure or service during Δt is $\mu \Delta t + O(\Delta t)$.

To have 0 (zero) customer in the system in time $t+\Delta t$, there may be following two cases:

(i) no customer in the system in time t, no arrival in time Δt.

(ii) one customer in the system in time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of no customer in the system at time $t+\Delta t$ is given by

$$p_0(t+\Delta t) = p_0(t)(1-\lambda\Delta t) + p_1(t)(1-\lambda\Delta t)\mu\Delta t + O(\Delta t)$$

$$= p_0(t) - \lambda p_0(t)\Delta t + \mu p_1(t)\Delta t + O(\Delta t)$$

or, $$\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t} = -\lambda p_0(t) + \mu p_1(t) + \frac{O(\Delta t)}{\Delta t} \text{ for } n=0 \qquad (1)$$

To have n customers in the system at time $t+\Delta t$, there may be following three cases:

(i) n customers in the system at time t, no arrival in time Δt, no service in time Δt.

(ii) $(n-1)$ customers in the system at time t, one arrival in time Δt, no service in time Δt.

(iii) $(n+1)$ customers in the system at time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of n customers in the system at time $(t+\Delta t)$ is given by

$$p_n(t+\Delta t) = p_{n-1}(t)\lambda_{n-1}\Delta t(1-\mu\Delta t) + p_n(i)(1-\lambda_n\Delta t)(1-\mu\Delta t) + p_{n+1}(t)(1-\lambda_{n-1}\Delta t)\mu\Delta t + O(\Delta t)$$

$$p_n(t+\Delta t) - p_n(t) = p_{n-1}(t)\lambda_{n-1}\Delta t - \lambda_n p_n(t)\Delta t - \mu p_n(t)\Delta t + \mu p_{n+1}(t)\Delta t + O(\Delta t)$$

or, $$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t} = \lambda_{n-1}p_{n-1}(t) - (\lambda_n+\mu)p_n(t) + \mu p_{n+1}(t) + \frac{O(\Delta t)}{\Delta t} \qquad (2)$$

when $n<N$, then $\lambda_{n-1}=\lambda,\ \lambda_n=\lambda$

Hence from (2), we have

$$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t} = \lambda p_{n-1}(t) - (\lambda+\mu)p_n(t) + \mu p_{n+1}(t) + \frac{O(\Delta t)}{\Delta t} \qquad (3)$$

When $n=N,\ p_{n+1}=0,\ \ \lambda_{n-1}=\lambda,\ \ \lambda_n=0$

∴ From (2), we have

$$\frac{p_N(t+\Delta t)-p_N(t)}{\Delta t} = \lambda p_{N-1}(t) - \mu p_N(t) + \frac{O(\Delta t)}{\Delta t} \qquad (4)$$

Taking limit, as $\Delta t\to 0$ in the equation (1), (3) and (4), we have

$p_0'(t) = -\lambda p_0(t) + \mu p_1(t)$ for $n = 0$

$p_0'(t) = \lambda p_{n-1}(t) + (\lambda + \mu) p_n(t) + \mu p_{n+1}(t)$ for $0 < n < N$

$p_0'(t) = \lambda p_{N-1}(t) - \mu p_N(t)$ for $n = N$

Under steady state condition of the system, the above equation reduce to

$-\lambda p_0 + \mu p_1 = 0$ for $n = 0$ (5)

$\lambda p_{n-1} - (\lambda + \mu) p_n + \mu p_{n+1} = 0$ for $1 \le n \le N$ (6)

$\lambda p_{N-1} + \mu p_N = 0$ for $n = N$ (7)

Solution:

From (5), we have $p_1 = \frac{\lambda}{\mu} p_0 = \rho p_0$ where $\rho = \frac{\lambda}{\mu}$

Putting $n = 1$ in (6), we have

$$\lambda p_0 - (\lambda + \mu) p_1 + \mu p_2 = 0$$

or, $$p_2 = -\frac{\lambda}{\mu} p_0 + \left(1 + \frac{\lambda}{\mu}\right) p_1 = -\rho p_0 + (1+\rho) p_1 = -p_1 + p_1 + \rho p_1 = \rho p_1 = \rho^2 p_0$$

Putting $n = 2$ in (6), we have

$$\lambda p_1 - (\lambda + \mu) p_2 + \mu p_3 = 0$$

$$\therefore p_3 = -\frac{\lambda}{\mu} p_1 + \left(1 + \frac{\lambda}{\mu}\right) p_2 = -\rho p_1 + (1+\rho) p_2 = -p_2 + p_2 + \rho p_2 = \rho p_2 = \rho^3 p_0$$

Proceeding in this way, we have

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0 = \rho^n p_0 \text{ for } 0 \le n < N \quad (8)$$

$$\therefore p_{n-1} = \left(\frac{\lambda}{\mu}\right)^{n-1} p_0 = \rho^{n-1} p_0$$

From (7), we have $p_N = \frac{\lambda}{\mu} p_{N-1} = \rho p_{N-1} = \rho \rho^{N-1} p_0 = \rho^N p_0$

$$\therefore p_n = \rho^n p_0 \text{ for } 0 \le n \le N \tag{9}$$

But since the capacity of the system in N

$$\sum_{n=0}^{N} p_n = 1$$

or, $p_0 + p_1 + p_2 + ... + p_N = 1$

or, $p_0 + \rho p_0 + \rho^2 p_0 + ... + \rho^N p_0 = 1$

or, $p_0\left(1 + \rho + \rho^2 + ... + \rho^N\right) = 1$

or, $p_0 \dfrac{1-\rho^{N+1}}{1-\rho} = 1$ for $\rho \ne 1$ and $(N+1)p_0 = 1$ for $\rho = 1$

or, $$p_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{N+1}} & \text{for } \rho \ne 1 \\ \dfrac{1}{N+1} & \text{for } \rho = 1 \end{cases}$$

From (9), we have

$$p_n = \rho^n p_0 = \begin{cases} \dfrac{(1-\rho)\rho^n}{1-\rho^{N+1}} & \text{for } 0 \le n \le N \text{ and } \rho \ne 1 \\ \dfrac{1}{N+1} & \text{for } 0 \le n \le N \text{ and } \rho = 1 \end{cases}$$

Note: Here we do not require the condition that $\rho < 1$ i.e., the results hold even if $\rho > 1$ as the number of customers allowed in the system is control by the queue length ($=N-1$), not by the relative rates of arrival and departure, λ and μ.

Calculation of λ_{eff}:

Since the probability that a customer does not join the system is p_N, the probability that customers join the system is $(1-p_N)$. Hence, $\lambda_{eff} = \lambda\left(1-p_N\right)$.

Characteristics of the Model

(i) Expected line length i.e., average number of customers in the system.

The expected line length i.e., average number of customers in the system is given by

$$L_s = e(n) = \sum_{n=0}^{N} np_n = \sum_{n=0}^{N} n\frac{1-\rho}{1-\rho^{N+1}}\rho_n = \frac{1-\rho}{1-\rho^{N+1}}\sum_{n=0}^{N} n\rho^n \text{ when } \rho \neq 1$$

$$= \frac{1-\rho}{1-\rho^{N+1}}\left(\rho + 2\rho^2 + 3\rho^3 + \ldots + N\rho^N\right)$$

[Let $S = \rho + 2\rho^2 + 3\rho^3 + \ldots\ldots + N\rho^N$

$\therefore \rho S = \rho^2 + 2\rho^3 + \ldots\ldots + N\rho^{N+1}$

Substracting second from first, we have

$$S(1-\rho) = \rho + \rho^2 + \rho^3 + \ldots .n \text{ times } - N\rho^{N+1}$$

or, $$S(1-\rho) = \frac{\rho\left(1-\rho^N\right)}{1-\rho} N\rho^{N+1}$$

or, $$S = \left[\frac{\rho\left(1-\rho^N\right)}{1-\rho} - N\rho^{N+1}\right]\frac{1}{1-\rho}]$$

$$= \frac{1-\rho}{1-\rho^{N+1}}\left[\frac{\rho\left(1-\rho^N\right)}{1-\rho} - N\rho^{N+1}\right]\frac{1}{1-\rho}$$

$$= \frac{1}{1-\rho^{N+1}}\left[\frac{\rho - \rho^{N+1} - N\rho^{N+1} + N\rho^{N+2}}{1-\rho}\right]$$

$$= \frac{\rho\left[1-(N+1)\rho^N + N\rho^{N+1}\right]}{(1-\rho)\left(1-\rho^{N+1}\right)} \text{ when } \rho \neq 1$$

For $\rho = 1, L_s = \sum_{n=0}^{N} np_n = \sum_{n=0}^{N}\frac{n}{N+1} = \frac{1}{N+1}(1+2+\ldots+N) = \frac{N}{2}\left[\text{For } \rho=1, p_n = \frac{1}{N+1}\right]$

(ii) Average number of customers in the queue

Average number of customers in the queue is given by

$$= L_q = \sum_{n=1}^{N}(n-1)\,p_n = \sum_{n=0}^{N} np_n - \sum_{n=1}^{N} p_n = \sum_{n=0}^{N} np_n - \sum_{n=0}^{N} p_n + p_0 = L_s - 1 + p_0 \left[\because L_s = \sum_{n=0}^{N} np_n\right]$$

$$= \frac{\rho\left[1-(N+1)\rho^N + N\rho^{N+1}\right]}{(1-\rho)\left(1-\rho^{N+1}\right)} - \left(1 - \frac{1-\rho}{1-\rho^{N+1}}\right)$$

$$= \rho^2\left[1 - N\rho^{N-1} + (N-1)\rho^N\right]/(1-\rho)\left(1-\rho^{N-1}\right)$$

(iii) Waiting time in the queue

The waiting time in the queue is

$$W_q = \frac{L_q}{\lambda_{eff}} = \frac{L_q}{\lambda(1-P_N)}\left[\because \lambda_{eff} = \lambda(1-p_N)\right]$$

$$\frac{L_q}{\lambda\left[1 - \frac{(1-\rho)\rho^N}{1-\rho^{N+1}}\right]} = \frac{\rho^2\left[1-N\rho^{N-1}+(N-1)\rho^N\right]}{\lambda\left(1-\rho^N\right)(1-\rho)}$$

(iv) Waiting time in the system:

The waiting time in the system is given by

$$W_s = W_q + \frac{1}{\mu}\left(\text{or } \frac{L_s}{\lambda_{eff}}\right)$$

Example 4

In a car wash service facility information gather indicates that cars arrive for service according to a poissons distribution with mean five per hour. The time for washing and cleaning for each car varies but is found to follow an exponential distribution with mean 10 minutes per car. The facility can not handle more than one car at a time and has a total of 5 parking spaces. If the parking spot is full, newly arriving cars balk to six services elsewhere.

(a) How many customers the manager of the facility is loosing due to the limited parking space?

(b) What is the expected waiting time until a car is washed?

Solution: In this system, there may be five car's in the five parking spot and one can be serviced i.e., the capacity of this system is $N = 5 + 1 = 6$

Hence this problem is of the model $(M/M/1)(6/FCFS/\infty)$

Here $\lambda = 5$ per hour $\mu = \frac{1}{10} \times 60 = 6$ per hour $\therefore \rho = \frac{\lambda}{\mu} = \frac{5}{6}$

Now $p_N = p_6 = \frac{1-\rho}{1-\rho^{6+1}} \rho^6 = 0.774$ $\left[\text{Since } p_N = \frac{1-\rho}{1-\rho^{N+1}} \rho^N\right]$

(a) Therefore the rate at which the cars balk

$= \lambda - \lambda_{eff} = \lambda - \lambda(1 - p_N) = \lambda p_6 = 5 \times .0774 = 0.387$ cars/hour

Assuming 8 working hours per day a manager will loose $0.387 \times 8 = 3.096 = 3$ cars a day

(b) Now the expected waiting time until a car is wash is given by

$$W_s = \frac{L_s}{\lambda_{eff}}$$

Now $L_s = \dfrac{\rho\left[1-(N+1)\rho^N + N\rho^{N+1}\right]}{(1-\rho)\left(1-\rho^{N+1}\right)} = 2.29$ hours

$\lambda_{eff} = \lambda(1 - p_N) = \lambda(1 - p_6) = 4.613$

Hence $W_s = \frac{L_s}{\lambda_{eff}} = 0.496$ hours

8.4 Model IV: $(M/M/c):(N/FCFS/\infty)$

This is queuing model with Poisson arrival, Poisson service or departure, c channels, capacity of the system equal to N (fixed) and the service discipline is first come, first serve. Let n be the number of customers in the system and λ, μ be the mean arrival rate and mean service rate of unit respectively.

In this mode,

$$\lambda_n = \begin{cases} \lambda & \text{when } 0 \le n < N \\ 0 & \text{when } n \ge N \end{cases}$$

and $\mu_n = \begin{cases} n\mu & \text{when } 0 \le n < c \\ c\mu & \text{when } c \le n \le N \end{cases}$

Hence the probability of one arrival during $\Delta t = \lambda_n \Delta t + O(\Delta t)$ and the probability of one departure or service during $\Delta t = \mu_n \Delta t + O(\Delta t)$

To have zero customer in the system in time $t + \Delta t$, there may be the following two cases:

(i) no customer in the system in time t, no arrival in time Δt.

(ii) one customer in the system in time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of no customer in the system at time $t + \Delta t$ is given by

$$p_0(t+\Delta t) = p_0(t)(1-\lambda_0 \Delta t) + p_1(t)(1-\lambda_1 \Delta t)\mu_1 \Delta t + O(\Delta t)$$

or, $p_0(t+\Delta t) - p_0(t) = -\lambda p_0(t)\Delta t + \mu p_1(t)\Delta t + O(\Delta t)$ $\quad \left[\begin{array}{l}\text{Here } \lambda_0 = \lambda \\ \lambda_1 = \lambda,\ \mu_1 = \mu\end{array}\right.$

or, $\dfrac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = \lambda p_0(t) + \mu p_1(t) + \dfrac{O(\Delta t)}{\Delta t}$ for $n = 0$

Taking limit as $\Delta t \to 0$, we have

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t) \text{ for } n = 0 \qquad (1)$$

To have n customers in the system in time $t + \Delta t$, there may be following three cases:

(i) n customers in the system at time t, no arrival in time Δt, no service in time Δt.

(ii) $(n-1)$ customers in the system at time t, one arrival in time Δt, no service in time Δt.

(iii) $(n+1)$ customers in the system at time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of n customers in the system at time $(t + \Delta t)$ is given by

$$p_n(t+\Delta t) = p_n(t)(1-\lambda_n \Delta t)(1-\mu_n \Delta t) + p_{n-1}(t)\lambda_{n-1}\Delta t(1-\mu_{n-1}\Delta t)$$
$$+P_{r+1}(t)(1-\lambda_{n+1}\Delta t)\mu_{n+1}\Delta t + O(\Delta t)$$

or, $p_n(t+\Delta t) - p_n(t) = -(\lambda_n + \mu_n)p_n(t)\Delta t + \lambda_{n-1}p_{n-1}(t)\Delta t + \mu_{n-1}p_{n+1}(t)\Delta t + O(\Delta t)$

or, $\dfrac{p_n(t+\Delta t) - p_n(t)}{\Delta t} = -(\lambda_n + \mu_n)p_n(t) + \lambda_{n-1}p_{n-1}(t) + \mu_{n-1}p_{n+1}(t) + \dfrac{O(\Delta t)}{\Delta t}$ (2)

(a) For $0 < n < c$

$$\lambda_{n-1} = \lambda_n = \lambda,\ \mu_n = n\mu,\quad \mu_{n+1} = (n+1)\mu$$

Hence from (2), we have

$$\frac{p_n(t+\Delta t) - p_n(t)}{\Delta t} = -(\lambda + n\mu)p_n(t) + \lambda p_{n-1}(t) + (n+1)\mu p_{n+1}(t) + \frac{O(\Delta t)}{\Delta t}$$

Taking limit as $\Delta t \to 0$ of both sides, we have

$$p_n'(t) = -(\lambda + n\mu)p_n(t) + \lambda p_{n-1}(t) + (n+1)\mu p_{n+1}(t) \tag{3}$$

(b) For $c \le n < N$

$$\lambda_{n-1} = \lambda,\ \lambda_n = \lambda,\ \mu_n = c\mu,\ \mu_{n+1} = c\mu$$

Hence from (2), we have

$$\frac{p_n(t+\Delta t) - p_n(t)}{\Delta t} = -(\lambda + c\mu)p_n(t) + \lambda p_{n-1}(t) + c\mu p_{n+1}(t) + \frac{O(\Delta t)}{\Delta t}$$

Taking limit as $\Delta t \to 0$ of both sides, we have

$$p_n'(t) = -(\lambda + c\mu)p_n(t) + \lambda p_{n-1}(t) + c\mu p_{n+1}(t) \text{ for } c \le n < N \tag{4}$$

(c) For $n = N$ $\lambda_n = \lambda_N = 0,\ \lambda_{n-1} = \lambda_{N-1} = \lambda,\ \mu_n = c_N = c\mu,\ p_{n+1}(t) = 0$

Hence from (2), we have

$$\frac{p_n(t+\Delta t) - p_n(t)}{\Delta t} = -(o + c\mu)p_n(t) + \lambda p_{n-1}(t) + \frac{O(\Delta t)}{\Delta t}$$

Taking limit as $\Delta t \to 0$ of both sides, we have

$$p_n'(t) = -c\mu p_n + \lambda p_{n-1}(t)$$

$$p_N'(t) = -c\mu p_n + \lambda p_{N-1}(t) \text{ for } n = N \tag{5}$$

Under steady state condition of the system i.e. $\lim_{t\to\infty} p_n(t) = p_n$ & $\lim_{t\to\infty} p_n'(t) = 0$, equations (1), (3), (4), (5) reduce to

$$-\lambda p_0 + \mu p_1 = 0 \text{ for } n = 0 \tag{6}$$

$$\lambda p_{n-1} - (\lambda + n\mu)p_n + (n+1)\mu p_{n+1} = 0 \text{ for } 0 < n < c \tag{7}$$

$$\lambda p_{n-1} - (\lambda + n\mu)p_n + c\mu p_{n+1} = 0 \text{ for } c \le n < N \tag{8}$$

$$\lambda p_{N-1} - c\mu p_N = 0 \text{ for } n = N \tag{9}$$

$$\therefore p_n = \begin{cases} \frac{1}{\lfloor n}\left(\frac{n}{\mu}\right)^n p_0 = \frac{1}{\lfloor n}(c\rho)^n p_0 & \text{for } 0 \le n \le c \\ \frac{1}{c^{n-c}\lfloor c}\left(\frac{\lambda}{\mu}\right)^n p_0 = \frac{c^c \rho^n}{\lfloor c} p_0 & \text{for } c \le n \le N \end{cases}$$

$$\text{where } p_0 = \left\{\sum_{n=0}^{c-1}\frac{1}{\lfloor n}\left(\frac{\lambda}{\mu}\right)^n + \sum_{n=c}^{N}\frac{1}{c^{n-c}\lfloor c}\left(\frac{\lambda}{\mu}\right)^n\right\}^1 = \left\{\sum_{n=0}^{c-1}\frac{1}{\lfloor n}(c\rho)^n + \sum_{n=c}^{N}\frac{c^c\rho^n}{\lfloor c}\right\}^{-1}$$

$$= \left[\sum_{n=0}^{c-1}\frac{(c\rho)^N}{\lfloor n} + \frac{c^c}{\lfloor c} + \left(\rho^c + \rho^{c+1} + \ldots + p^N\right)\right]^{-1} = \left[\sum_{n=0}^{c-1}\frac{(c\rho)^N}{\lfloor n} + \frac{c^c\rho^c}{\lfloor c} + \left(1 + \rho + \ldots + p^{N-c}\right)\right]^{-1}$$

$$= \left[\sum_{n=0}^{c-1}\frac{(c\rho)^n}{\lfloor n} + \frac{(c\rho)^c}{\lfloor c}\cdot\frac{1-\rho^{N-c+1}}{1-\rho}\right]^{-1} \quad \text{for } \rho = \frac{\lambda}{c\mu} \ne 1$$

$$= \left[\sum_{n=0}^{c-1}\frac{(c\rho)^n}{\lfloor n} + \frac{(c\rho)^c}{\lfloor c}(N-c+1)\right]^{-1} \quad \text{for } \rho = \frac{\lambda}{c\mu} = 1$$

Remark: If we talk $N \to \infty$ and consider $\frac{\lambda}{c\mu} < 1$, then the above result are same as those of the model $(M/M/C):(\infty/FCFS/\infty)$ and the putting $c = 1$, we get the results of the model $(M/M/1):(\infty/FCFS/\infty)$.

Characteristic of the model:

Average queue length is

$$L_q = \sum_{n=c+1}^{N}(n-c)\,p_n = \sum_{n=c}^{N}(n-c)\,p_n = \sum_{n=c}^{N}(n-c)\frac{c^2\rho^n}{\lfloor c}p_o$$

$$= \frac{c^2}{\lfloor c}p_o\sum_{n=c}^{N}(n-c)\rho^n = \frac{c^2}{\lfloor c}p_o\rho^c\sum_{n=c}^{N}(n-c)\rho^{n-c}$$

$$= \frac{(c\rho)^c}{\lfloor c} p_o \sum_{x=0}^{N-c} x\rho^x \text{ putting } n-c=x$$

$$= \frac{(c\rho)^c}{\lfloor c} p_o \sum_{n=0}^{N-c} \rho \frac{d}{d\rho}(\rho^x)$$

$$= \frac{(c\rho)^c}{\lfloor c} p_o \rho \frac{d}{d\rho} \left\{ \sum_{x=0}^{N-c} \rho^x \right\}$$

$$= \frac{(c\rho)^c}{\lfloor c} p_o \rho \frac{d}{d\rho} \left[\frac{1-\rho^{N-c+1}}{1-\rho} \right]$$

$$= \begin{cases} \dfrac{(c\rho)^c}{\lfloor c} p_o \rho \dfrac{1-\rho^{N-c+1}-(1-\rho)(N-c+1)\rho^{N-c}}{(1-\rho)^2} & \text{for } \rho = \dfrac{\lambda}{c\mu} \neq 1 \\ p_o \dfrac{c^c (n-c)(N-c+1)}{L\lfloor c} \text{ for } \lambda = \dfrac{\lambda}{c\mu} = 1 \end{cases}$$

L = average number of customer in the system.

$$= E(n) = \sum_{n=0}^{N} x p_n = \sum_{n=0}^{c-1} n p_n + \sum_{n=0}^{N} n p_n$$

$$= \sum_{n=c}^{N} (n-c) p_n + \sum_{n=c}^{N} c p_n + \sum_{n=0}^{c-1} n p_n$$

$$= L_q + c \left[\sum_{n=c}^{N} p_n - \sum_{n=0}^{c-1} p_n \right] + \sum_{n=0}^{c-1} n p_n$$

$$= L_q + c - c \sum_{n=c}^{c-1} p_n + \sum_{n=c}^{c-1} n p_n = L_q + c + \sum_{n=c}^{c-1} (n-c) p_n$$

$$= L_q + c - p_0 \sum_{n=c}^{c-1} \frac{(c-n)(\rho c)^n}{\lfloor n}$$

Calculation of $\lambda_{eff} = \lambda(1-p_N) = \lambda_{eff}$

Let $\bar{c}$ = expected number of idle servers

$$\therefore \bar{c} = \sum_{n=0}^{c} (c-n) p_n$$

$\therefore c - \bar{c}$ = the expected number of busy servers

$\therefore \mu(c-\bar{c})$ = Actual number of customer served per unit time

$\therefore \lambda_{eff} = \mu(c-\bar{c})$

Now $W_s = \dfrac{\lambda s}{\lambda_{eff}}$ and $W_q = \dfrac{L_q}{\lambda_{eff}}$

Machine Repair Problem

Whenever a machine breaks down, it will result a great loss to the organization. So the machine repair problems are very important problems in queuing theory when a machine breaks down, the machine start its repairing. If during this time another machine breaks down then it will be attended after completion of the repair of the first machine. Thus the broken machines form a queue and wait for their repair. There are various situations. There may be one or more than one machines. If there is one machine then it is called the problem of single channel and in case there are more than one machines, then it is called multi channel problem. The machines may be repaired in a single phase or in k-phases.

8.5 Model-V: (M/M/R): (k/GD/k), k>R

In this queuing system, there are k-machines which are serviced by R repairmen or mechanics. The arrivals (break down) and service follow Poisson distribution with parameter λ and μ respectively. A broken machine which is in service can not be a candidate for a new customer and as such the calling source is finite in number equal to k. Let n be the number of machines in breakdown situation.

Hence the approximate probability of a single service during the time Δt is $\mu_n \Delta t$ where

$$\mu_n = \begin{cases} n\mu \text{ for } n \le R \\ R\mu \text{ for } R \le n \le k \\ 0 \text{ for } n > k \end{cases}$$

If λ is the rate of breakdown per machine, then the probability of a single arrival during the time Δt (when there are n machines in breakdown situation) is approximately $\Delta t\ \lambda_n$ for $n \le k$ where

$$\lambda_n = \begin{cases} (k-n)\lambda \text{ for } 0 \le n < k \\ 0 \text{ for } n \ge k \end{cases}$$

To have 0 customer in the system in time $t+\Delta t$, there may be following two cases

(i) no customer in the system at time t, no arrival in time Δt.

(ii) one customer in the system at time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of no customers in the system at time $t+\Delta t$ is given by

$$p_0(t+\Delta t) = p_0(t)(1-\lambda_0\Delta t) + p_1(t)(1-\lambda_1\Delta t)\mu_1\Delta t + O(\Delta t)$$

$$= p_0(t)(1-k\lambda\Delta t) + p_1(t)\{1-(k-1)\lambda\Delta t\}\mu\Delta t + O(\Delta t)$$

or, $p_0(t+\Delta t) - p_0(t) = -k\lambda p_0(t)\Delta t + \mu p_1(t)\Delta t + O(\Delta t)$

or, $\dfrac{p_0(t+\Delta t)-p_0(t)}{\Delta t} = -k\lambda p_0(t) + \mu p_1(t) + \dfrac{O(\Delta t)}{\Delta t}$ (1)

To have n customers in the system at time $t+\Delta t$, there may be following three cases:

(i) n customers in the system at time t, no arrival in time Δt, no service in time Δt.

(ii) $(n-1)$ customers in the system at time t, one arrival in time Δt, no service in time Δt.

(iii) $(n+1)$ customers in the system at time t, no arrival in time Δt, one service in time Δt.

Therefore the probability of n customers in the system at time $(t+\Delta t)$ is given by

$$p_0(t+\Delta t) = p_n(t)(1-\lambda_n\Delta t)(1-\mu_n\Delta t) + p_{n-1}(t)\lambda_{n-1}\Delta t(1-\mu_{n-1}\Delta t)$$

$$+p_{n-1}(t)(1-\lambda_{n-1}\Delta t)\mu_{n-1}\Delta t + O(\Delta t)$$

or, $p_n(t+\Delta t) - p_n(t) = -\lambda_n p_n(t)\Delta t - \mu_n p_c(t)\Delta t + \lambda_{n-1}p_{c-1}(t)\Delta t - \mu_{n+1}p_{t+1}(t)\Delta t + O(\Delta t)$

or, $\dfrac{p_n(t+\Delta t)-p_n(t)}{\Delta t} = -(\lambda_n+\mu_n)p_n(t) + \lambda_{n-1}p_{n-1}(t) + \mu_{n+1}p_{n+1}(t) + \dfrac{O(\Delta t)}{\Delta t}$ (2)

[dividing bothsides by Δt]

When $0 < n = R$, then

$$\lambda_{n-1} = \{k-(n-1)\}\lambda$$

$$\lambda_n = (k-n)\lambda,\ \lambda_{n+1} = \{k-(n+1)\}\lambda$$

$$\mu_n = n\mu,\ \mu_{n+1} = (n+1)\mu$$

Hence from (2), we have

$$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t}=-\{(k-n)\lambda+n\mu\}\,p_n(t)+\{k-(n-1)\}\,\lambda p_{n-1}(t)$$

$$+(n+1)\mu p_{n+1}(t)+\frac{O(\Delta t)}{\Delta t} \qquad (3)$$

For $R \le n < k,\ \lambda_{n-1}=\{k-(n-1)\}\lambda,\ \lambda_n=(k-n)\lambda,\ \mu_n=R\mu,\ \mu_{n+1}=R\mu$

Hence from (2), we have

$$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t}=-\{(k-n)\lambda+R\mu\}\,p_n(t)+\{k-(n-1)\}\,\lambda p_{n-1}(t) \qquad (4)$$

$$+R\mu p_{n+1}(t)+\frac{O(\Delta t)}{\Delta t} \quad \text{for } R\le n<k$$

For $n = k$

$$\lambda_{n-1}=\{k-(n-1)\}\,\lambda=\{k-(k-1)\}\,\lambda=\lambda[\therefore n=k$$

$$\lambda_n=(k-n)\lambda=0$$

$$\mu_n=R\mu$$

Hence from (2), we have

$$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t}=-(o+R\mu)\,p_n(t)+\lambda p_{n-1}(t)+\frac{O(\Delta t)}{\Delta t} \quad \text{for } n=k$$

i.e., $$\frac{p_k(t+\Delta t)-p_k(t)}{\Delta t}=\lambda p_{k-1}(t)+R\mu p_k(t)+\frac{O(\Delta t)}{\Delta t} \qquad (5)$$

Now taking limit as $\Delta t\to 0$ in (1), (3), (4), (5), we have

$$p_n'(t)=-\mu\lambda p_0(t)+\mu p_1(t) \quad \text{for } n=0$$

$$p_n'(t)=-\{(k-n)\lambda+n\mu\}\,p_n(t)+\{k-(n-1)\}\,\lambda p_{n-1}(t)+(n+1)\mu p_{n+1}(t) \quad \text{for } 0<n<R$$

$$p_n'(t)=-\{(k-n)\lambda+R\mu\}\,p_n(t)+\{k-(n-1)\}\,\lambda p_{n-1}(t)+R\mu p_{n+1}(t) \quad \text{for } R\le n<k$$

$$p_n'(t)=\lambda p_{k-1}(t)+R\mu p_k(t) \quad \text{for } n=k$$

Under steady state condition of this system i.e., $\lim_{t\to\infty} p_n(t) = p_n$ and $\lim_{t\to\infty} p_n'(t) = 0$, the above four equations reduces to

$$-k\lambda p_0 + \mu p_1 = 0 \text{ for } n = 0 \quad (6)$$

$$\{k-(n-1)\}\lambda p_{n-1} - \{(k-n)\lambda + n\mu\} p_n + (n+1)\mu p_{n+1} = 0 \text{ for } 0 < n < R \quad (7)$$

$$\{k-(n-1)\}\lambda p_{n-1} - \{(k-n)\lambda + R\mu\} p_n + R\mu p_{n+1} = 0 \text{ for } R \le n < k \quad (8)$$

$$\lambda p_{k-1} - R\mu p_k = 0 \text{ for } n = k \quad (9)$$

From (6), $\mu p_1 = k\lambda p_0$

or, $p_1 = k\frac{\lambda}{\mu} p_0$ or $p_1 = k\rho p_0$ where $\rho = \frac{\lambda}{\mu}$

Now putting $n = 1$ in (7) we have

$$k\lambda p_0 - \{(k-1)\lambda + \mu\} p_1 + 2\mu p_2 = 0$$

or, $2\mu p_2 = \{(k-1)\lambda + \mu\} p_1 + k\lambda p_0$

or, $2p_2 = \left\{(k-1)\frac{\lambda}{\mu} + 1\right\} p_1 - k\frac{\lambda}{\mu} p_0$

$$= \{(k-1)\rho + 1\} p_1 - k\rho p_0 \left[\because \rho\frac{\lambda}{\mu}\right]$$

$$= (k-1)\rho p_1 + p_1 - p_1 [\because p_1 = k\rho p_0]$$

$$= (k-1)\rho p_1 = (k-1)\rho k\rho p_0 = k(k-1)\rho^2 p_0$$

$$p_2 = \frac{k(k-1)}{2}\rho^2 p_0 = \left(\frac{k}{2}\right)\rho^2 p_0$$

Again putting $n = 2$ in (7), we have

$$3p_3 = \frac{(k-2)(k-1)k\rho^3 p_0}{2}$$

or, $p_3 = \frac{(k-2)(k-1)k\rho^3 p_0}{3.2} = \frac{k}{3}\rho^3 p_0$

By induction, $p_n = \binom{k}{n} \rho^n p_0$ for $0 \le n \le R$

Putting $n = R$ in (8), we have

$$\{k-(R-1)\}\lambda p_{R-1} - \{(k-R)\lambda + R\mu\} p_R + R\mu p_{R+1} = 0$$

or, $$Rp_{R-1} = \{(k-R)\rho + R\} p_R - (k-R+1)\rho p_{R-1}$$

$$= \{(k-R)\rho + R\}\binom{k}{R}\rho^R p_0 - (k-R+1)\rho\binom{k}{R-1}\rho^{R-1}p_0$$

$$= \{(k-R)\rho + R\}\binom{k}{R}\rho^R p_0 - R\binom{k}{R}\rho^R p_0$$

$$= \binom{k}{R}\rho^R p_0\left[(k-R)\rho + R - R\right]$$

$$= \binom{k}{R}(k-R)\rho^{R+1}p_0$$

$$= \binom{k}{R-1}(R+1)\rho^{R+1}p_0$$

$$\therefore p_{R+1} = \binom{k}{R+1}\frac{R+1}{R}\rho^{R+1}p_0$$

Again putting $n = R+1$ in (8), we have

$$(k-R)\rho p_R - \{(k-\bar{R}+1)\rho + R\} p_{R+1} + Rp_{R+1} = 0$$

or, $$Rp_{R+2} = \{(k-R-1)\rho + R\} p_{R+1} - (k-R)\rho p_R$$

$$= (k-R-1)\rho p_{R+1} + Rp_{R+1} - (k-R)\rho p_R$$

$$\therefore p_{R+2} = \frac{(k-R-1)(R+1)}{R^2}\binom{k}{R+1}\rho^{R+2}p_0$$

$$= \frac{(k-R-1)(R+1)}{R^2}\frac{\lfloor\underline{k}}{\lfloor\underline{R+1}\lfloor\underline{k-R-1}}\rho^{R+2}p_0$$

$$= \frac{(R+1)(R+2)}{R^2\lfloor\underline{R+2}}\frac{\lfloor\underline{k}}{\lfloor\underline{k-(R+2)}}\rho^{R+2}p_0$$

$$= \binom{k}{R+2}\frac{(R+1)(R+2)}{R^2}\rho^{R+2}p_0 = \binom{k}{R+2}\frac{\lfloor\underline{R+2}}{\lfloor\underline{R}R^2}\rho^{R+2}p_0$$

$$\therefore p_{R+1} = \binom{k}{R+i}\frac{\lfloor\underline{R+i}}{\lfloor\underline{R}R^i}\rho^{R+i}p_0 \text{ for } R \le R+i < k$$

From (9), we have

$$R\mu p_k = \lambda p_{k-1}$$

or, $$Rp_k = \frac{\lambda}{\mu}p_{k-1} = \rho\binom{k}{k-1}\frac{\lfloor\underline{k-1}}{\lfloor\underline{R}R^{k-R-1}}\rho^{k-1}p_0$$

$$\binom{k}{1}\frac{\lfloor\underline{k-1}}{\lfloor\underline{R}R^{k-R-1}}\rho^k p_0 = \binom{k}{k}\frac{k\lfloor\underline{k-1}}{\lfloor\underline{R}R^{k-R-1}}\rho^k p_0 = \binom{k}{k}\frac{\lfloor\underline{k}}{\lfloor\underline{R}R^{k-R-1}}\rho^k p_0$$

$$\therefore p_k = \binom{k}{k}\frac{\lfloor\underline{k}}{\lfloor\underline{R}R^{k-R}}\rho^k p_0$$

$$\text{Thus } p_n = \begin{cases} \binom{k}{n}\rho^n p_0 & \text{for } 0 \le n \le R \\ \binom{k}{n}\frac{\lfloor\underline{n}}{\lfloor\underline{R}R^{n-R}}\rho^n p_0 & \text{for } R \le n \le k \end{cases}$$

Now, $\sum_{n=0}^{n} p_n = 1$

or, $$\sum_{n=0}^{R-1} p_n + \sum_{n=R}^{k} p_n = 1$$

or, $$\sum_{n=0}^{R-1}\binom{k}{n}\rho^n p_0 + \sum_{n=R}^{k}\binom{k}{n}\frac{\lfloor\underline{n}}{\lfloor\underline{R}R^{n-R}}\rho^n p_0 = 1$$

or, $$p_0 = \frac{1}{\sum_{n=0}^{R-1}\binom{k}{n}\rho^n + \sum_{n=R}^{n}\binom{k}{n}\frac{\rho^n\, n!}{R^{n-R}\, R!}}$$

Characteristics of the model

(i) Average number of customers in the system (L_s)

Average number of customers in the system is given by

$$L_s = E(n) = \sum_{n=0}^{k} np_n = \sum_{n=0}^{R-1} np_n + \sum_{n=R}^{k} np_n$$

$$= \sum_{n=0}^{R\cdot 1} n\binom{k}{n}\rho^n + \frac{1}{R!}\sum_{n=R}^{k} n\binom{k}{n}\frac{\rho^n\, n!}{R!\,R^{n-R}}$$

$$= p_0\left[\sum_{n=0}^{n-1} n\binom{k}{n}\rho^n + \frac{1}{R!}\sum_{n=R}^{k} n\binom{k}{n}\frac{\rho^n\, n!}{R^{n-R}}\right]$$

(ii) Expected queue length (L_q)

Expected queue length is given by

$$L_q = \sum_{n=R+1}^{k} (n-R)p_n = \sum_{n=R+1}^{k} np_n - \sum_{n=R+1}^{k} np_n$$

$$= \sum_{n=0}^{k} np_n - \sum_{n=0}^{R} np_n - R\left\{\sum_{n=0}^{k} p_n - \sum_{n=0}^{R} p_n\right\}$$

$$= L_s - \sum_{n=0}^{R} np_n - R\left\{1 - \sum_{n=0}^{R} p_n\right\} \left[\because \sum_{n=0}^{k} p_n = 1\right]$$

$$= L_s - R - \sum_{n=0}^{R} np_n + \sum_{n=0}^{R} p_n$$

$$= L_s - R + \sum_{n=0}^{R}(R-n)\,p_n$$

$= L_s - \left(R - \overline{R}\right)$ where $\overline{R} = \sum_{n=0}^{R}(R-n)\,p_n$ = expected number of idle machine repairmen.

(iii) Effective arrival rate (λ_{eff})

In this queuing system, the arrivals occur with a rate λ but all arrivals do not join the system. This situation occurs when the maximum allowable queue length is reached. In that case, no new arrivals are allowed to join the queue. So, we shall define λ considerating those arrivals which join the system. This arrival rate is known as effective arrival rate and it is denoted by λ_{eff}.

Hence the effective arrival rate is given by

$$\lambda_{eff} = \mu(R-\bar{R})$$

or, $$\lambda_{eff} = \sum_{n=0}^{k} \lambda_n p_n = E[\lambda_n] = E[\lambda(k-n)] [\text{Since } \lambda_n = \lambda(k-n)]$$

$$= \lambda E[(k-n)] = \lambda[k-E(n)] = \lambda[k-L_s]$$

(iv) Expected waiting time

Expected waiting time of a customer in the system is given by

$$W_R = \frac{L_s}{\lambda_{eff}}$$

whereas expected waiting time of a customer in the queue is given by

$$W_q = \frac{L_q}{\lambda_{eff}}$$

Example 5

There are five machines, each of which, when running, suffers break downs at an average rate of 2 per hour. There are two servicemen and only one man can work on a machine at a time. If n machines are out of order when $n>2$, the $n-2$ of them wait until a serviceman is free. Once a serviceman starts work on a machine, the time to complete the repair has an exponential distribution with mean 5 minutes. Find the distribution of the number of machines out of action at a given time. Also, find the average time an out-of-action machine has to spend waiting for the repairs to start.

Solution:

Here

k = total number of machines in the system = 5

R = number of servicemen = 2

$\lambda = 2$ per hour, $\mu = 1$ per 5 minutes hour

Hence $\rho = \lambda'\mu = \frac{2}{12} = \frac{1}{6}$

Let n be the number of machines out of order.

Hence

$$p_n = \begin{cases} \binom{5}{n}\left(\frac{1}{6}\right)^n p_0 & \text{for } 0 \le n \le 2 \\ \binom{5}{n}\frac{\lfloor n}{2^{n-2}\lfloor 2}\left(\frac{1}{6}\right)^n p_0 & \text{for } 2 \le n \le 5 \end{cases}$$

$$= \begin{cases} \binom{5}{n}\left(\frac{1}{6}\right)^n p_0 & \text{for } 0 \le n \le 2 \\ \binom{5}{n} 2\lfloor n\left(\frac{1}{12}\right)^n p_0 & \text{for } 2 \le n \le 5 \end{cases}$$

$$\text{where } p_0 = \left[\sum_{n=0}^{2-1}\binom{5}{n}\left(\frac{1}{6}\right)^n + \sum_{n=2}^{5}\binom{5}{n}2\lfloor n\left(\frac{1}{12}\right)^n\right]^{-1} = \frac{648}{1493}$$

Average number of machines out of action is given by

$$L_q = \sum_{n=2+1}^{5}(n-2)\,p_n = \sum_{n=3}^{5}(n-2)p_n = p_3 + 2p_4 + 3p_5 = \frac{165}{1493}$$

Average time an out-of-action machine has to spend waiting for the repairs to start is $W_q = L_q / \lambda_{eff}$

But, $\lambda_{eff} = \sum_{n=0}^{k}\lambda_n p_n = \sum_{n=0}^{k}\lambda(k-n)p_n = \lambda\sum_{n=0}^{5}(5-n)\,p_n = \frac{6 \times 2050}{1493}$.

Hence W_q = 55/4100 hrs. = 33/41 minutes.

9. SELF ASSESSMENT QUESTIONS/EXERCISE

(a) What is queuing theory? What information can be obtained by analyzing a queuing system?

(b) Give a brief description of the types of queue discipline commonly found.

(c) Explain the important characteristics of queuing system.

(d) Derive and solve the steady state difference equations governing the queuing system $(M/M/1):(\infty/FCFS/\infty)$.

(e) For the queuing model $(M/M/1):(N/FCFS/\infty)$, derive and solve the steady state differential – difference equation.

(f) Obtain the expected waiting time for a customer in the queue for the queuing model $(M/M/1):(N/FCFS/\infty)$.

(g) Derive and solve the steady state difference equations for the queuing system $(M/M/C):(N/FIFO)$.

(h) For the M/M/C queue system derive expressions for (i) probability that a person will not have to wait; (ii) average number of persons in the system.

(i) A telephone exchance has two long distance operators. The telephone company finds that during the peak period, long distance calls arrive in poisson fashion at an average rate of 15 per hour. The length of service on these calls is approximately exponentially distributed with mean length 5 minutes.

What is the probability that a subscriber will have to wait for his long distance call during the peak hours of the day?

If the subscribers wait and are serviced in turn, what is the expected waiting time?

(j) In a steady state system with Poisson input and Poisson output, find the probability that there are m customers waiting in the queue with a single channel and 'first in first out' discipline and hence find the average number of customers in the system (the capacity of the system is infinite).

(k) Patients arrive at a clinic according to a Poisson distribution at the rate of 30 patients per hour. The waiting room does not accommodate more than 14 patients. Examination time per patient is exponential with mean rate 20 per hour.

(i) Find the effective arrival rate at the clinic.

(ii) What is the probability that an arriving patient will not wait?

(iii) What is the expected waiting time until a patient is discharged from the clinic?

(l) A bank has two tellers working on savings accounts. The first teller handles withdrawals only. The second teller handles depositors only. It has been found that the service time distributions of both deposits and withdrawals are exponentials with a man service time of 3 minutes per customer. Depositors and withdrawers are found to arrive in a Poisson fashion throughout the day with mean arrival rate of 16 and 14 per hour. What would be the effect on average waiting time for each with depositors and withdrawers if each teller could handle both withdrawals and depositors?

(m) In a railway marshalling yard, good trains arrive at a rate of 30 trains per day. Assuming that the interval-arrival time follows an exponential distribution and the service time (the time taken to jump a train) distribution is also exponential with average of 36 minutes. Calculate the expected queue size and the probability that the queue size exceeds 10.

10. SUGGESTED FURTHER READINGS

* Taha, H.A., Operations Research - An Introduction, PHI.
* Bronson, R. and Naadimuthu, G., Theory and Problems of Operations Research, Schuam's Outline Series, MGH.
* Swarup, K., Gupta, P.K. and Man Mohan, Operations Research, Sultan Chand & Sons.
* Sharma, J.K., Operations Research - Theory and Applications, Macmillan.
* Gupta, P.K. and Hira, D.S., Operations Research, S. Chand & Co. Ltd.

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X

Special Paper : OR

Module No. - 115(a)
Reliability

1. INTRODUCTION

Reliability is the capability of a device to work without any breakdown. It is defined with respect to a time horizone and combines the time dimension with the performance level of equipment. Formally, it is defined as the probability of a device performing satisfactority for a given period of time when it is operated in the manner specified and for the purpose it is intended.

Use of the concept of reliability dates back to the time when multiengine aircraft emerged between the World Wars I and II. Probability concepts were used to evaluate the chances of failure of one out of two engines or one out of four engines during flights. A complex system such as an aircraft or a computer system not only requires smooth functioning of he components but also depends upon its design as well as on varous interconnections and switchover devices. The environmental conditions like pressure, humidity, vibrations and shocks also affect the reliability of a system. Although originaed from aeronautical and various defence equipments, the concept of reliability is also being used in industry specially in manufacturing instruments.

The idea iof reliability is of help to the manufacturer is copying with the occurrence of malfunctions or failures of the products in service. While the quality of the components of a manufacturing item is very important for the finished product, it has been observed that the performance level as measured by reliability can be significantly improved by using an apropriate design.

The reliability of a system is a function of its components. The occurrence of failure in a component can not be predicted with certainty as it is a statistical variable. The study of reliability of various components requires a significant amount of data collection and its analysis. The feedback of information goes from the user to the repairer, the manufacturer right upto the designer.

The reliability of equipment is estimated from the reliability of its components. Statistical procedures such as Life Testing can be used to estimate the reliability of the different components. Sometimes failure rates under conditions of extreme stress tend to infer the rates under normal working conditions. Reliability estimating and allied problems have thus become an important branch of statistics.

Structure :

2. OBJECTIVES

The objectives of this module are to

* discuss the basic concepts relating to reliability;
* introduce the measures of reliability;
* derive the reliability of series system, parallel system, parallel-series system and series-parallel system.

3. KEYWORDS

Reliability, probability, early failures, chance failures, wear-out failures, measures of reliability, failure density, MTBF, MTTF, Hazard rate, system reliability, series system, parallel system, series-parallel system, parallel-series system, stand-by redundancy.

4. RELIABILITY

Reliability is defined as the probability of a device performing its intended purpose adequately for the period of time intended under the operating conditions encountered. The reliability is the probability with which the devices will not fail to perform a required operation for a certain length of time. Such problem is known as the problem of survival.

This definition brings into the focus of four important factors viz.,

(i) The reliability of a device is expressed as a probability.

(ii) The device is required to give adequate performance.

(iii) The duration of adequate performance is specified.

(iv) The environmental or operating conditions are specified.

In practice, even the best design manufacturing and maintenance efforts do not completely eliminate the occurrence of failure. During the life of a system we may experience three distinct types of failures - early failures, random failures and wear out failures.

Early failures :

Early failures are those which occur in the early life of a sysem operations.

Random or Chance failures :

Chance failure are predominant during the actual operating of the system and occur at random, irregularly and unexpectedly. The phase during which only chance failures occur is called the useful life of the system.

Wear-out failure :

Wear–out failures are caused due to ageing any wearing out of components. These failures occur if the system maintain properly or not maintain at all. The frequency of such failures increases rapidly with time.

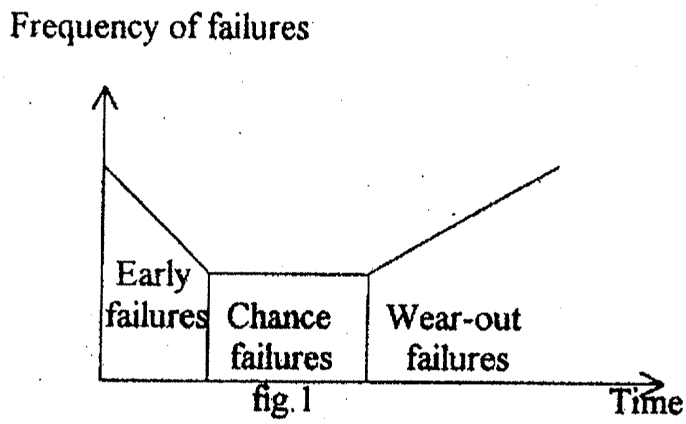


Fig. -1 : Life characteristics curve

4.1 Measures of reliability

The prediction of system reliability is based on a number of factors such as : life characteristics, operating conditions and the failure distribution.

If a random sample of items are taken from a population and are put to test (or use) under a set of fixed (or given) environmental or operating conditions, some number of sample will fail successively in time. The data so obtain will represent the length of each item. The length of life can be measured depending on whether the item is repairable (Radio, TV, Aeroplate etc.) or non-repairable (bulb, fuse, missile, rocket etc.). For repairable items, the life can be measured by failure rate or mean time between failures whereas for non repairable items the life can be measured by mean time to failure.

Failure rate : The failure rate λ is defined as the number of failure in a given time interval

$$\lambda = \frac{\text{Number of failure}}{\text{Total unit of operating hour}} = \frac{f}{T}$$

Failure density :

This is the ratio of the number of failures during a given unit interval of time to the total number of items at the very beginning of the test.

Example 1 :

Let the total number of items at the beginning of the test i.e., the total initial population was 1000 and during the first unit interval, the number of components that failures 130 (say). The failure density during the first unit interval, $fd_1 = \frac{n_1}{N} = \frac{130}{1000} = 0.13.$

Similarly, the failure density during the second unit interval ca.. also be calculated.

Generalization

Let n_1 be the number of components that fail during the first unit interval and n_2 be the number that fail during the second unit interval and so on.

Let N be the total population. Then the failure density during the first unit interval is $fd_1 = \frac{n_1}{N}$, failure density during the second unit interval is $fd_1 = \frac{n_1}{N}$ and so on.

Let l be the last interval after which there is no survivor, then

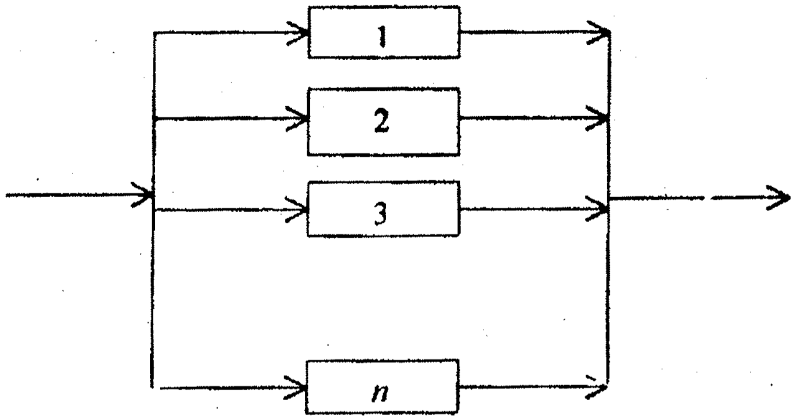


Fig. 8

The reliability of the system can be calculated considering the conditions for system failure. Let $X_1, X_2, ..., X_n$ represent the successful operation of units 1, 2, ..., n respectively. Similarly, let $\bar{X}_1, \bar{X}_2, ..., \bar{X}_n$ represent their unsuccessful operation respectively. If $p(X_1)$ is the probability of successful operation of unit then $p(\bar{X}_1)$ is the probability of successful operation of unit then $p(\bar{X}_1)$ is the probability of its failure.

Hence, $p(\bar{X}_1) = 1 - p(X_1)$.

As the system will fail if all n units of the system fail simultaneously. The probability of failure of he system is given by

$$p(\bar{S}) = p(\bar{X}_1\bar{X}_2\bar{X}_3...\bar{X}_n)$$

$$= p(\bar{X}_1)p(\bar{X}_2/\bar{X}_1)p(\bar{X}_3/\bar{X}_1\bar{X}_2) --- p(\bar{X}_n/\bar{X}_1\bar{X}_2 - \bar{X}_{n-1})$$

As the failure of the units are independent of one another then $p(\bar{S}) = p(\bar{X}_1)p(\bar{X}_2) - p(\bar{X}_n)$.

Hence, the probability of success of the system is given by

$$p(S) = 1 - p(\bar{S})$$

$$= 1 - p(\bar{X}_1)p(\bar{X}_2) - p(\bar{X}_n)]$$

$$= 1 - [\{1 - p(X_1)\}\{1 - p(X_2)\} --- \{1 - p(X_n)\}]$$

Hence the reliability of the system comprising of n components connected in parallel is given by

$$Z(t) = \lim_{h\to 0}\left[\frac{R(t)-R(t+h)}{hR(t)}\right]$$

$$= -\frac{1}{R(t)}\lim_{h\to 0}\left[\frac{R(t+h)-R(t)}{h}\right]$$

$$= -\frac{1}{R(t)}\frac{dR}{dt}$$

$$= \frac{f(t)}{R(t)} \qquad \left[\text{since } f(t) = -\frac{dR}{dt}\right]$$

Relationship between *Z*(*t*) and *R*(*T*)

Now, $Z(t) = \frac{f(t)}{R(t)} = \frac{dQ(t)}{dt}\frac{1}{R(t)}$ [since $Q(t)+R(t)=1$]

which implies $\int_0^t Z(t)dt = \int_0^t \frac{1}{R(t)}\frac{dQ}{dt}dt$

$$= \int_0^t \frac{1}{1-Q(t)}dQ$$

$$= -\log\left[1-Q(t)\right]_{t=0}^{t} = -\log R(t)\Big|_0^t = -\log R(t) \qquad [\text{Since } R(0)=1]$$

$$\therefore R(t) = \exp\left\{-\int_0^t Z(t)dt\right\}$$

$$\therefore Q(t) = 1-R(t) = 1-\exp\left\{-\int_0^t Z(t)dt\right\}$$

$$\therefore f(t) = -\frac{dR(t)}{dt} = Z(t)\exp\left\{-\int_0^t Z(t)dt\right\}$$

Particular case

When the failure rate is constant then $Z(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$

This means that for constant failure rate, the hazard rate is also constant and is equal to the failure rate.

The expected value $E(t)$ of a MTBF of a continuous random variable t is given by

$$E(t) = MTBF = \int_0^\infty tf(t)\,dt$$

where $f(t)$ is the failure density function.

We have $f(t) = \dfrac{dQ(t)}{dt} = \dfrac{dR(t)}{dt}$

Hence, $MTBF = \int_o^\infty -t\dfrac{dR(t)}{dt}dt = -\int_{t=0}^\infty tdR$

$$= tR(t)\Big|_{t=0}^\infty + \int_0^\infty R(t)\,dt \qquad \text{[by parts]}$$

$$= 0 + \int_0^\infty R(t)\,dt$$

$$\therefore MTBF = m = \int_0^\infty R(t)\,dt$$

Particular case

For a constant failure rate we have $R(t) = e^{-\lambda t}$

$$\therefore MTBF = m = \int_0^\infty R(t)\,dt = \int_0^\infty e^{-\lambda t}\,dt = \frac{1}{\lambda}$$

Example 2

A device with 1000 hours useful life at a constant failure rate 0.0001 per hour in a given experiment. What is the reliability per 10 hours operation of this device. Find the value of MTBF?

Solution : We know that the reliability per 10 hours is $R(t) = e^{-\lambda t}$

Here l = 0.0001 and $t = 10$

Hence, $R(t) = e^{-\lambda t} = e^{-00018*10} = e^{-.001} = 0.999$

$\therefore MTBF = \frac{1}{\lambda} = \frac{1}{0.0001} = 10,000$ hours.

Example 3

For an equipment, the reliability per 100 hours of operation has been estimated to be 0.999. What is the failure rate of the equipment? Calculate MTBF?

Solution :

It is given that $R(t = 0.999$

Here $e^{-\lambda t}$ is constant, so $e^{-\lambda t} = 0.999$ as $R(t) = e^{-\lambda t}$

$\Rightarrow 1 - \lambda t = 0.999$, neglecting square and higher powers of λt.

$\Rightarrow 1 - 100\lambda = 0.999$ as $t = 100$

$\therefore \lambda = 0.00001$.

i.e., Failure rate = 0.00001 per hours.

= 10^{-5} per hours.

Hence, MTBF = $1/\lambda = 10^5$ hours.

4.2 System Reliability

Generally to determine the reliability factor of a system, the system is broken down to sub systems and elements whose individual reliability factors can be estimated or determined. Depending on the manner in which these subsystems and elements are connected to constituted the given system. The combinatorial rules are appolied to obtain the system reliability. Hence the basic steps are as follows:

(i) First identify the elements and subsystems which constitute the given system and whose individual reliability factors can be estimated.

(ii) Next draw a block diagram or a circuit diagram to represent the logical manner or configuration in which these units are connected to form the system.

(iii) The condition for the successful operation of the system is then determined i.e., it may be decided as to how the units should function.

(iv) Finally the combinatorial rules of probability theory i.e. addition, multiplication and their combinations are applied to arrival at the system reliability factor.

4.2.1 Reliability of series system

Series system

In this system, a large number of components of the system are connected in series which means that if any one of the components fails, the system fails. In other words, if the system is operated, each component connected in series should successfully be operated. The system comprising of $n-$ components in series is represented as

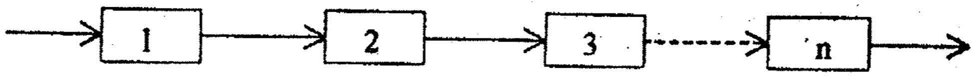


Fig. 6

Let the successful operation of these individual units be represented by $X_1, X_2, ..., X_n$. and their respective probabilities $p(X_1), p(X_2), ..., p(X_n)$. For the successful operation of the system, it is necessary that all n units function satisfactorily.

Hence if the units are not independent one another then the system reliability is

$$p(S) = p(X_1X_2X_3....X_n)$$

$$= p(X_1)p(X_2/X_1)p(X_3/X_1X_2).........p(X_n/X_1X_2...X_{n-1})$$

If the successful operation of each unit is independent of the successful operation of the remaining units the $p(S) = p(X_1)p(X_2)...p(X_n)$.

If R_i is the reliability of the i-th components in series in the system and R_s is the reliability of the system having n components in series then $R_s = R_1R_2....R_n = \prod_{i-1}^{n} Ri$

If $R_1 = R_2 = R_3 = ... = R_n = R$ (say) then

$R_s = R^n = (1-Q)^n$ where Q is the probability of failure of each unit. When ech component has an exponential time to failure density, then $R_s = R_1 R_2 ... R_n = e^{-\lambda_1 t} e^{\lambda_2 t} ... e^{-\lambda_n t}$

$$R_s = R_1 R_2 ... R_n = e^{-\lambda_1 t} e^{\lambda_2 t} ... e^{-\lambda_n t} \text{ where } \lambda_s = \sum_{i=1}^{n} \lambda_s i$$

The mean time between failures for the system having n-components in series is given by

$$m_s = \int_0^\infty R_s(t) dt$$

$$= \int_0^\infty e^{-(\lambda_1+\lambda_2+--\lambda_n)t} dt$$

$$= \frac{1}{\lambda_1 + \lambda_2 + ... + \lambda_n} = \frac{1}{\lambda_s}$$

if $\lambda_1 = \lambda_2 = ... = \lambda_n = \lambda$ (say) then $R_s(t) = e^{-(n\lambda)t}$ and $m_s = \frac{1}{m\lambda}$

If $n = 1$, then $R_s = e^{-\lambda t}$ and $m = 1/\lambda$.

Example 4

An electronic circuit consists of 5 silicon transistor, 3 silicon diodes, 10 composite register and 2 ceramic capacitor in series configuration. The hourly failure rate of each component is

for transistor : $\lambda_t = 4 \times 10^{-5}$

for diode : $\lambda_d = 3 \times 10^{-5}$

for resistor : $\lambda_r = 2 \times 10^{-4}$

for capacity : $\lambda_c = 2 \times 10^{-4}$

Calculate the reliability of the circuit for 10 hours when the components follow exponential distribution.

Solution :

Components	Nos. (N_i)	λ_i (per hour)	$N_i\lambda_i$

Transistor	5	4×10^{-5}	20×10^{-5}
Diode	3	3×10^{-5}	9×10^{-5}
Resister	10	2×10^{-4}	20×10^{-4}
Capacitor	2	2×10^{-4}	4×10^{-4}
			$\sum N_i\lambda_i = 0.00269$

Let λ_s and R_s be the failure rate and reliability of the system.

As each component follows exponential time to failure density, then

$\lambda_s = \sum \lambda_i = 5\lambda_t + 3\lambda_d + 10\lambda_r + 2\lambda_c = 20\times10^{-5} + 9\times10^{-5} + 20\times10^{-4} + 4\times10^{-4}$

$= 0.00269$ per hour

$\therefore R_s = e^{-\lambda_s t} = e^{-.00269t}$

The estimated reliability of the circuit for ten hours is $R_s(10) = e^{-0.00269\times10} = e^{-0.0269} = 0.9735$

This means that the circuit is expected to operate on an average without failure 9735 times and would fail (10000-9735) i.e., 265 times out of 10,000 operations of 10 hours each.

In this case, MTBF is $m = 1/\lambda = 1/0.00269 = 371.75$ hours.

This means that the circuit is expected to operate without failure for 372 hours.

Example 5

The system connected is series consist of three independence parts A, B and C which have MTBF of 100, 400 and 800 hours respectively. Find MTBF of the system and reliability of the systems for 30 hours. How much MTBF of the parts A has to be increased to get and improvement of MTBF of the system by 30%.

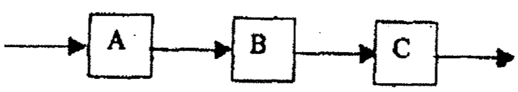


Fig. 7

Solution :

Here, MTBF of the part A = m_1 = 100 hours

MTBF of the part B = m_2 = 400 hours

MTBF of the part C = m_3 = 800 hours

λ_1 = failure rate of $A = 1/m_1 = 1/100$ failure hours

λ_2 = failure rate of $B = 1/m_2 = 1/400$ failure hours

λ_3 = failure rate of $C = 1/m_3 = 1/800$ failure hours

Hence failure rate of system $\lambda_s = \lambda_1 + \lambda_2 + \lambda_3$

= 1/100+ 1/ 400 + 1/800

= 11/800 failure / hour

Therefore, MTBF of the system $= 1/\lambda_s$

= 1/(11/800)

= 72.75 hours.

And reliability of the system $= R_s(30) = e^{\lambda_s \times 30} = e^{-11/800 \times 30} = e^{-33/80}$

Second part : MTBF of the system = 72.75 hors = m_s (say)

Let m'_s be the new MTBF of the system.

$$\therefore m'_s = m_s + 30\% \text{ of } m_s = \frac{800}{11} + \frac{30}{100} \times \frac{800}{11} = 1040/11$$

Again, failure rate of the system $= \lambda'_s = 1/m_s = 11/1040$ hours.

Let x hours will be the MTBF of parts A to improve the MTBF of the system by 30%.

Hence, 1/(100+x) + 1/400 + 1/800 = 11/1040

i.e., x = 46.5

4.2.2 Reliability of parallel system

In this system a large number of components of the system are connected in parallel which means that the successful operation of the system depends on the satisfactory functioning of any one of the components. A system comprising of n - components connected in parallel in represented as

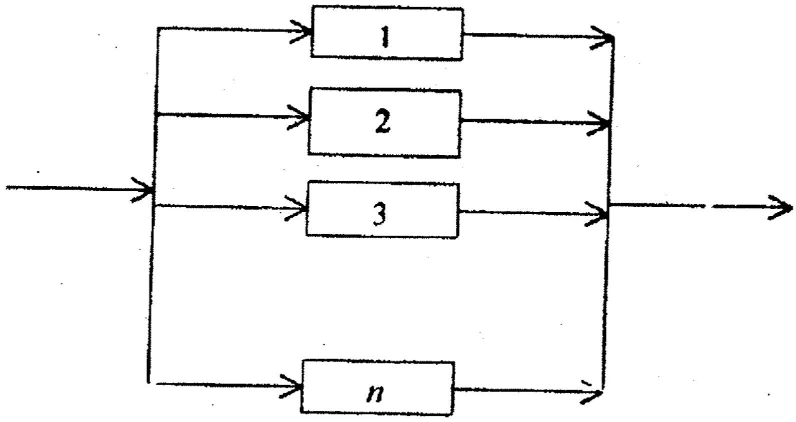


Fig. 8

The reliability of the system can be calculated considering the conditions for system failure. Let $X_1, X_2, ..., X_n$ represent the successful operation of units 1, 2, ..., n respectively. Similarly, let $\bar{X}_1, \bar{X}_2, ..., \bar{X}_n$ represent their unsuccessful operation respectively. If $p(X_1)$ is the probability of successful operation of unit then $p(\bar{X}_1)$ is the probability of successful operation of unit then $p(\bar{X}_1)$ is the probability of its failure.

Hence, $p(\bar{X}_1) = 1 - p(X_1)$.

As the system will fail if all n units of the system fail simultaneously. The probability of failure of he system is given by

$$p(\bar{S}) = p(\bar{X}_1 \bar{X}_2 \bar{X}_3 ... \bar{X}_n)$$

$$= p(\bar{X}_1) p(\bar{X}_2/\bar{X}_1) p(\bar{X}_3/\bar{X}_1\bar{X}_2) --- p(\bar{X}_n/\bar{X}_1\bar{X}_2 - \bar{X}_{n-1})$$

As the failure of the units are independent of one another then $p(\bar{S}) = p(\bar{X}_1) p(\bar{X}_2) - p(\bar{X}_n)$.

Hence, the probability of success of the system is given by

$$p(S) = 1 - p(\bar{S})$$

$$= 1 - p(\bar{X}_1) p(\bar{X}_2) - p(\bar{X}_n)]$$

$$= 1 - [\{1 - p(X_1)\}\{1 - p(X_2)\} --- \{1 - p(X_n)\}]$$

Hence the reliability of the system comprising of n components connected in parallel is given by

$$R_s(t) = 1 - \prod_{i=1}^{n} Q_i(t)$$

where $q_i(t)$ is the failure of the i-th component $= 1 - \prod_{i=1}^{n} \{1 - R_i(t)\}$

In n units are identical and the unit failures are independent of one another then

$$R_s(t) = 1 - \prod_{i=1}^{n} \{1 - R_i(t)\}$$

$$= 1 - \{1 - R(t)\}^n$$

If the failure rate of each unit is exponential time to failure distribution, then

$$R_s(t) = 1 - \prod_{i=1}^{n} \{1 - R_i(t)\} = 1 - \prod_{i=1}^{n} \{1 - e^{-\lambda_i t}\}$$

where λ_i is the failure rate of the i-th component.

For identical units $R_s(t) = 1 - \left(1 - e^{\lambda t}\right)^n$

Mean Time between Failure

The mean time between failure of a system having two components connected in parallel can be obtained by integrating the reliability function over the range of t from 0 to ∞,

i.e., $m_s = \int_0^{\infty} R_s(t)dt = \int_0^{\infty} \left[1 - \{1 - R_1(t)\}\{1 - R_2(t)\}\right]dt$

$$= \int_0^{\infty} \left[1 - \{1 - e^{-\lambda_1 t}\}\{1 - e^{-\lambda_2 t}\}\right]dt = \int_0^{\infty} \left[e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1 + \lambda_2)t}\right]dt$$

$$= 1/\lambda + 1/\lambda - 1/(\lambda_1 + \lambda_2)$$

If these two components are identical i.e. $\lambda_1 = \lambda_2 = \lambda$ (say) then $m_s = 1/\lambda + 1/\lambda - 1/2\lambda = 1.5/\lambda$. This means that the mean time between failure of a parallel system consisting of two components of equal failure rate is 1.5 times the MTBF of a single component.

The system reliability can be improved either by improving the desin or by providing redundancy of the system. The system with redundancy has a number of stand-by units which take over if the other components of the system in fail. This a reserve stock of stand-by's improves system reliability but at a higher cost. The compromise between increase cost due to stand-by arrangement and increase given reliability will help to decide the optimal redundancy of the system.

Example 6

The failure rate of an electronic subsystem is 0.0005 failure/hour. If a operational period of 500 hours with probability of success 0.95 is desired. What label of parallel redundancy is needed?

Solution : Given that $\lambda = 0.0005$ failures/hour, R_s = reliability of the system = 0.95.

Let there be n subsystems connected in parallel.

As the subsystems are connected in parallel, $R_s = 1-(1-R)^n$ where R is the reliability of the system.

Again, $R = e^{-\lambda t} = e^{-.0005\times 500} = e^{-0.25}$

Therefore, $0.95 = 1-(1-e^{-0.25})^n$

which implies $n = \log(0.05)/\log(0.2214) \approx 2$

Hence the label of parallel redundancy is 2.

Example 7

How many identical components each of which is 90% reliable over a period of 50 hours be used to obtain a 99.99% parallel redundancy system over 50 hours. If we want to obtain the same system reliability over a period of 100 hours, how many components should be added?

Solution : Given that R = reliability of each component = 90% = 0.9, R_s = reliability of the parallel system = 99.99% = 0.999 and t = 50 hours.

Let there be n label of parallel redundancy.

$\therefore R_s = 1-(1-R)^n \Rightarrow 0.9999 = 1-(1-0.9)^n \Rightarrow (0.1)^n = .0001 \Rightarrow n = 4$

Hence 4 identical components are required to be connected in parallel.

Second part :

As the reliability of the system per 100 hours is given, we have to evauate the reliability of each components per 100 hours.

As the reliability of each component per 50 hours is 0.9,

$0.9 = e^{-\lambda * 15}$ or $0.9 - 1 - 50\lambda$ [neglecting higher powers of λ] $\therefore \lambda = 0.002$.

Let R' be the reliability of each component per 100 hours

$\therefore R' = e^{-\lambda t} = e^{-.002 \times 100} = e^{-.2} = 0.81$.

$\therefore R_s = 1 - (1 - R')^n$ where n is the label of redundancy.

$= 1 - (1 - 0.81)^n$

$\Rightarrow 0.9999 = 1 - (0.19)^n \Rightarrow n \approx 6$

Hence for the 2nd case, 6–4 i.e., 2 identical components are to be added.

Example 8

Show that MTBF of the system of n identical units is parallel connected is $\frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i}$ where λ is the failure rate of each component.

Solution : Let the parallel redundant system is having n units with failure rate $\lambda_1, \lambda_2, ..., \lambda_n$ (say). Hene the reliability of the system is given by

$$R_s = 1 - (1 - R_1)(1 - R_2)...(1 - R_n).$$

$$= 1 - \left(1 - e^{-\lambda_1 t}\right)\left(1 - e^{-\lambda_2 t}\right)...\left(1 - e^{-\lambda_n t}\right)$$

$$= 1 - \prod_{i=1}^{n}\left(1 - e^{-\lambda_i t}\right)$$

Now, MTBF of the system $= \int_0^\infty R_s(t)\,dt = \int_0^\infty \left[1 - \left(1 - e^{-\lambda_1 t}\right)\left(1 - e^{-\lambda_2 t}\right)...\left(1 - e^{-\lambda_n t}\right)\right]dt$

$$= \int_0^\infty \left[\left(e^{-\lambda_1 t} + e^{-\lambda_2 t} + ... + e^{-\lambda_n t}\right) - \left\{e^{-(\lambda_1+\lambda_2)t} + e^{-(\lambda_1+\lambda_3)t}\right\} + ...\right]dt$$

$$= (1/\lambda + 1/\lambda_2 + ... + 1/\lambda_n) \pm \{1/(\lambda_1 + \lambda_2) + 1/(\lambda_1 + \lambda_3) + ...\} + \{1/(\lambda_1 + \lambda_2\lambda_3) + 1/(\lambda_1 + \lambda_3 + \lambda_4) + ...\}$$

$$+ ... + (-1)^{n+1}\frac{1}{\sum_{i=1}^{n}\lambda_i}$$

For identical units, $\lambda_1 = \lambda_2 = ... = \lambda_n$

In that case, MTBF $= 1/\lambda - 1/2\lambda + 1/n\lambda = \frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i}$

2.3 Mixed Configuration

Parallel series system

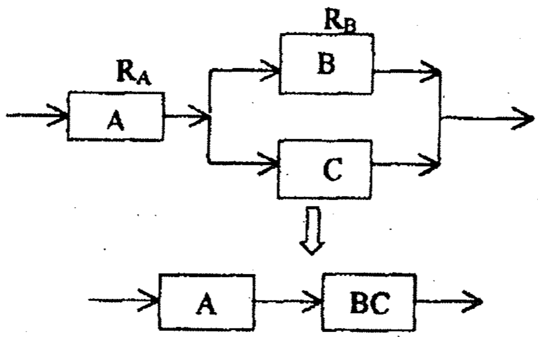


Fig. 9

As the unit B and C are connected in parallel, these units may be replaced by a unit BC (say) with reliability R_{BC} where

$$R_{BC} = 1 - (1 - R_B)(1 - R_e)$$

and the changed configuration of the system will be as

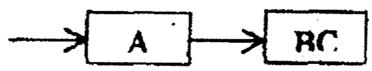


Fig. 10

Now the units A and BC are connected in series. The reliability of the system can be calculated as

$$R_s = R_A R_{BC} = R_A\{1 - (1 - R_B)(1 - R_C)\}$$

Another mixed configuration

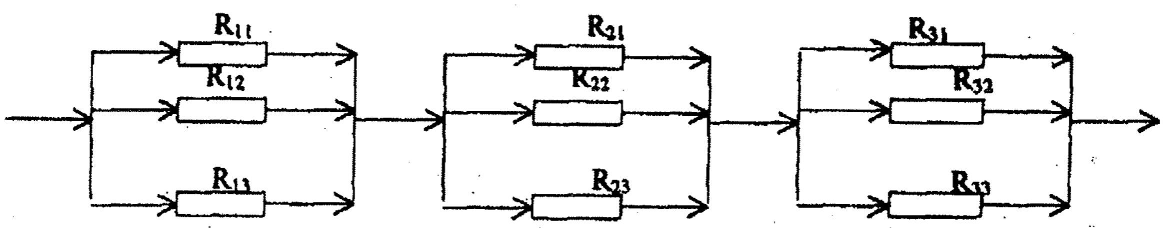


Fig. 11

The above system can be represented by an equivalent system in the following figure.

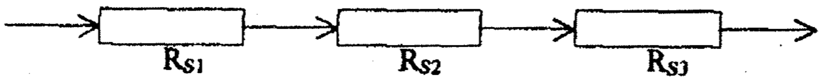


Fig. 12

In this case,

$$R_{S1} = 1-(1-R_{11})(1-R_{12})(1-R_{13})$$

$$R_{S2} = 1-(1-R_{21})(1-R_{22})(1-R_{23})$$

$$R_{S3} = 1-(1-R_{31})3(1-R_{33})$$

Hence the reliability of the system will be $R_s = R_{S1}R_{S2}R_{S3}$

Series parallel system

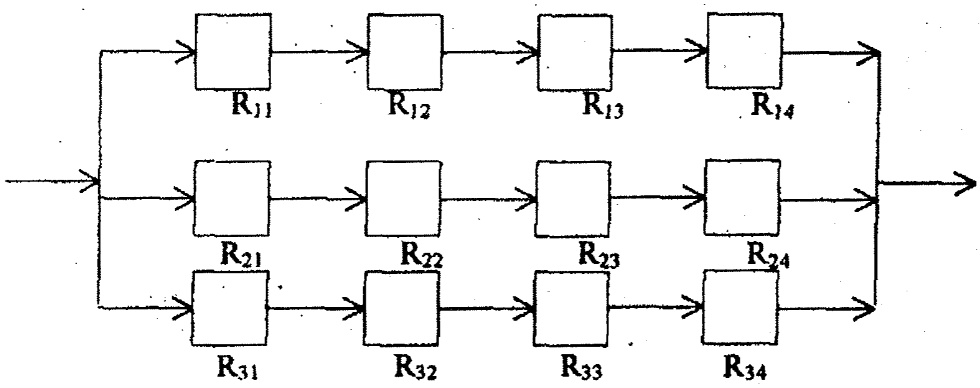


Fig. 13

The above system can be represented by an equivalent in Fig. 14.

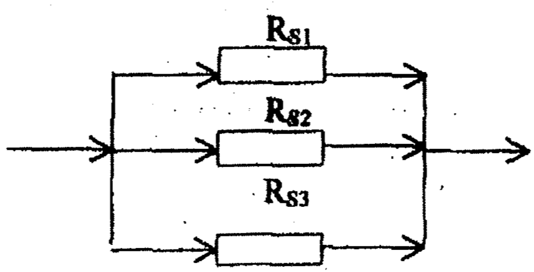


Fig. 14

Here $R_{S1} = R_{11}\,R_{12}\,R_{13}.R_{14}$, $R_{S2} = R_{21}\,R_{22}\,R_{23}\,R_{24}$ and $R_{S3} = R_{31}\,R_{32}\,R_{33}\,R_{34}$

So, R_S = reliability of the system = $= 1-(1-R_{S1})(1-R_{S2})(1-R_{S3})$.

Example 9

A system consisting of 4 identical subsystems connected in parallel. Each subsystem consists of 3 identical units connected in series. If for each unit the probability of each over a certain period of time is 0.95 obtain the system of reliability.

The given system can be represented by an equivalent system in Fig. 16.

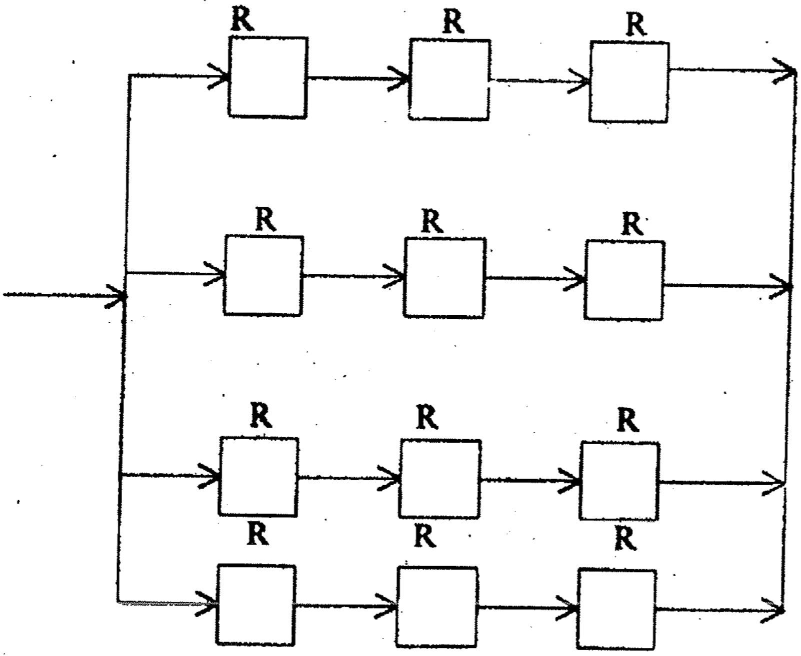


Fig.15

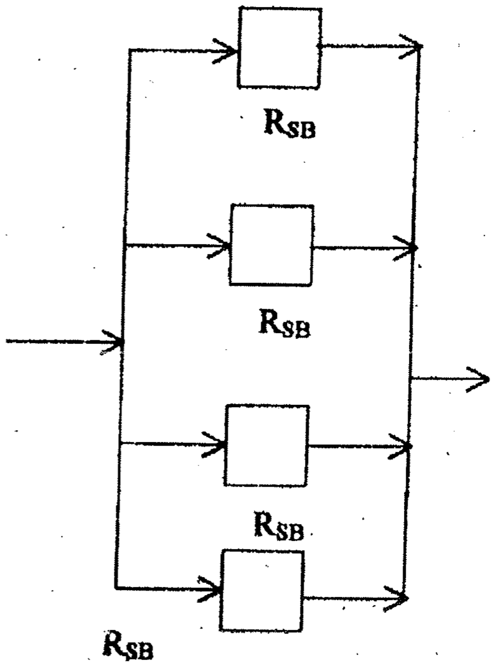


Fig. 16

Here $R = 0.95$ (given)

Hence the reliability of each sub system is $R_{SB} = R^3 = (0.95)^3 = 0.857$

Hence the system reliability $(R_S) = 1-(1-R_{SB})^4, = 1-(1-0.857)^4 = 0.999582$

4.2.4 Stand-by Redundancy

In parallel configuration, all the components operate simultaneously and the experiences wear and tear during the operation of the system. In the stand-by-system, there is a primary active element which is operating and one or more components are standing-by to take over the operation one after another when the first one fails. The operation of stand-by components is sequential i.e., each of the duplicate elements becomes active and is energized onoy after the failure of the previous active elements. The system will survive until the required time t if either the following two condition holds.

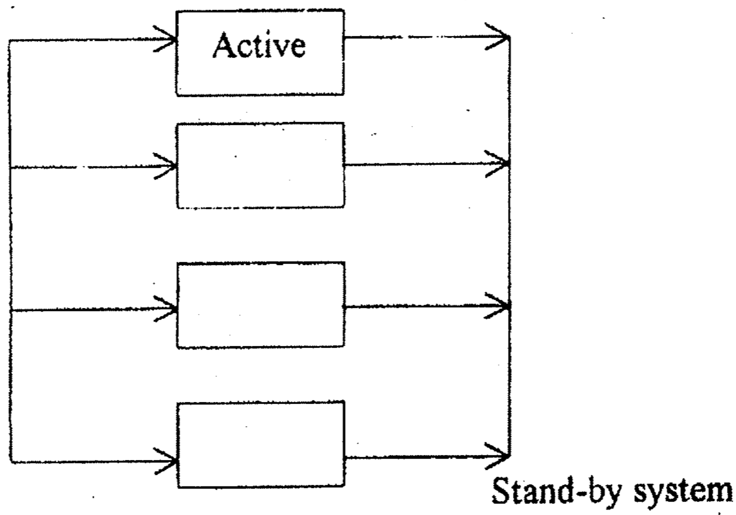


Fig. 17

(i) Primary element is successful operating upto time t.

(ii) Primary element fails at t_1 and the active stand-by element takes over and survives from t_1 to t.

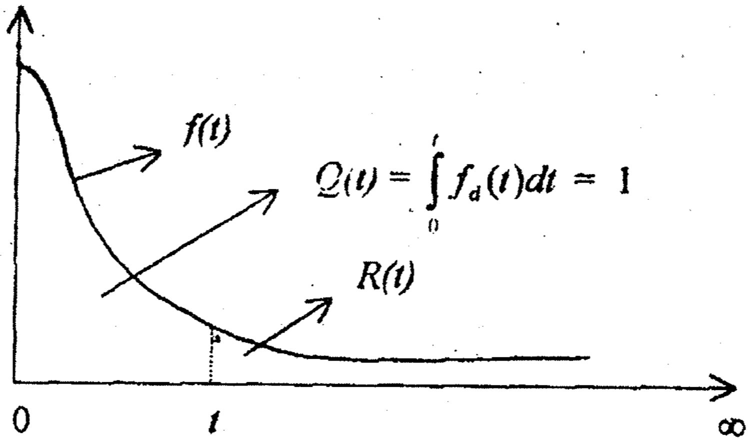
Assuming the failure density function of the primal active and the stand-by unit as $f_p(t)$ and $f_d(t)$. respectively we have

(a) probability of primary element surviving upto the time t is $1-\int_0^t f_p(t)\,dt=\int_t^\infty f_p(t)\,dt$

(b) probability of primary element failing at time t_1 is $f_p(t_1)$

(c) probability of stand-by element working successfully from t_1 to $t=1-\int_0^{t-t_1} f_d(t)\,dt=\int_{t-t_1}^\infty f_d(t)\,dt$

Hence the reliability of the system is given by $R_2(t) = \int_t^{\infty} f_p(t)dt \int_0^t f_p(t_1)\left[\int_{t-t_1}^{\infty} f_d(t)dt]dt_1\right].$



Fiig.- 18

For the exponential case, $f_p(t) = \lambda_p e^{\lambda_p t}, f_d(t) = \lambda_d e^{-\lambda dt}$

where λ_p and λ_d are the failure rates of the primarily active and stand-by elements respectively.

Hence the reliability of the system $R_t(t) = \int_t^{\infty} \lambda_p e^{-\lambda_p t} dt + \int_0^t \lambda_p e^{-\lambda_p t_1} \left\{ \int_{t-t_1}^{\infty} \lambda_d e^{-\lambda_d t} dt \right\} dt_1$

$$= e^{-\lambda_p t} + \int_0^t \lambda_p e^{-\lambda_p t_1} e^{-\lambda_d (t-t_1)} dt_1 \quad = e^{-\lambda_p t} + \lambda_p \int_0^t e^{-\lambda_d t} e^{(\lambda_d - \lambda_p)} dt_1 = \left(\frac{\lambda_d}{\lambda_d - \lambda_p} \right) e^{-\lambda_p t} - \left(\frac{\lambda_p}{\lambda_d - \lambda_p} \right) e^{-\lambda_d t}$$

For $\lambda_p = \lambda d = \lambda, R_2(t) = e^{-\lambda t} + \lambda \int_0^t e^{-\lambda t} dt_1$

$$= e^{-\lambda t} + \lambda e^{-\lambda t} \int_0^t dt = e^{-\lambda t} + \lambda t e^{-\lambda t}$$

Now the failure density function is exponential. Number of failures follows Poisson distribution i.e., prob (Number of failures $= i) = p(i) = e^{-\lambda t} \cdot (\lambda t)^i / i$

From this, $p(0) = e^{-\lambda t}, p(1) = \lambda t e^{-\lambda t}$ and so on.

So, $R_2(t) = e^{-\lambda t} + \lambda t e^{-\lambda t} = p(0) + p(1)$

Similarly, it can be show that

$$R_3(t) = p(0) + p(1) + p(3) = e^{-\lambda t} + \lambda t e^{-\lambda t} + \frac{(\lambda t)^2}{\lfloor 2}e^{-\lambda t} \quad \left[\text{As } R_3(t) = \sum_{i=0}^{2} \frac{e^{-\lambda t}(\lambda t)^i}{\lfloor i}\right]$$

Hence, for a stand-by redundant system with n components of equal failure rate, the reliability of the system is given by $R(t) = \sum_{i=0}^{n-1} \frac{e^{\lambda t}(\lambda t)^i}{\lfloor i} = e^{-\lambda t} \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{\lfloor i}$

MTBF

Let us consider the MTBF of a stand-by redundant system with one active and one stand-by. Then

$$m_2 = \int_0^{\infty} R_2(t)\,dt = \int_0^{\infty} (1 + \lambda t)e^{-\lambda t}dt = \int_0^{\infty} e^{-\lambda t}dt + \int_0^{\infty} \lambda t e^{-\lambda t dt} = \frac{1}{\lambda} + \frac{\lambda}{\lambda_2} = \frac{2}{\lambda}$$

Hence the mean time between the failures in the case of a stand-by redundant system consisting of two components is twice that of the single component.

Stand-by system with imperfect sensing over device

Let sensing over device is not perfectly reliable and hence its probability of the failure must be considered. Let the primary and secondary components be not similar and higher different failure rate and the system consist of two dissimilar components placed in stand-by redundancy. The failure rates of both components follow the exponential failure time distribution. The primary components have a failure rae λ_1 (say) and the stand-by component λ_2 (say).

Let r_{ss} be the reliability of the sensing and switching over device.

The reliability of such a system is given by $R(t) = e^{-\lambda_1 t} + r_{ss}\left(\frac{\lambda}{\lambda_2 - \lambda_1}\right)\left[e^{-\lambda_1 t} - e^{-\lambda_2 t}\right]$

When $\lambda_1 = \lambda_2 = \lambda$, then $R(t) = e^{-\lambda t} + r_{ss}\left(\lambda t e^{-\lambda t}\right) = e^{-\lambda t}\left[1 + r_{ss}(\lambda t)\right]$

In general for a stand-by system of n components which have equal failure rate and when one unit is operating activity and the rest $(n-1)$ units are standing-by to take over the operation in succession then MTBF of a system is $m_n = n/\lambda$.

Example 10

An industrial process in controlled by a computer and two similar components are operated in stand-by redundancy such that if a computer fails another is instantaneously brought into use in its plae. The failure rate of each computer is given by $\lambda = 0.01$ failure/hour. Compare the improvement in reliability over a single computer when one and then two computers are used in a stand-by. The operating period is 100 hours and the switch is considered to be perfect.

(i) In the first case, only one computer is in the system. Hence the reliability of the system having single computer is given by

$$R_1(t) = e^{-\lambda t} = e^{-.01\times 100} = 1/e = 0.37$$

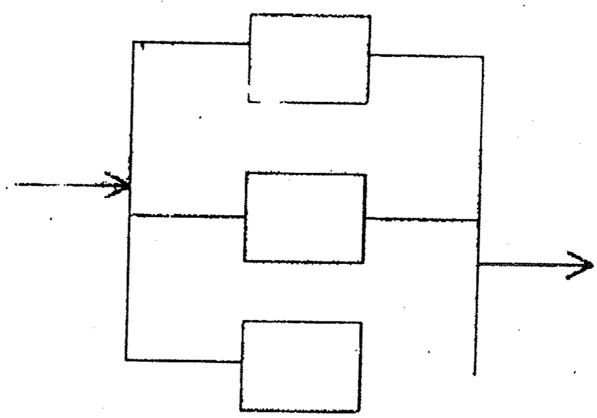


Fig. 19

(ii) For the system having one computer along with another one as stand-by, the reliability of the system is given by

$$R_2(t) = \sum_{i-0}^{i-1} \frac{e^{-\lambda t} \cdot (\lambda t)^i}{\lfloor i}$$

$$= e^{-\lambda t} + (\lambda t)e^{-\lambda t} = 0.37 + 0.01 \times 100 \times 0.37 = 0.37 + 0.37 = 0.74$$

As $R_2(t) = 2R_1(t)$, hence, there is 100% improvement in reliability by using one stand-by computer.

(iii) Again, for system having one computer along with two computer as stand-by, the reliability of the system is given by

$$R_3(t) = \sum_{i=0}^{3-1} \frac{e^{-\lambda t} \cdot (\lambda t)^i}{\lfloor i} = e^{-\lambda t} + \lambda t\left(e^{-\lambda t}\right) + \frac{e^{-\lambda t} \cdot (\lambda t)^2}{\lfloor 2} = 0.74 + \frac{0.37 \times 1}{2} = 0.74 + 0.185 = 0.925$$

Now, $R_3(t)/R_1(t) = 0.925/0.37 = 2.5$

Hence by using two stand-by computers, reliability is improved by 2.5 times i.e., there is a 150% improvement in reliability over the single computer.

Example 11

An electronic device has a failure rate of 500 failures per 10^6 hours. One identical stand-by unit is added to increase the reliability of the basic device. The operating time is 1000 hours. The failure rate of the sensing and switching element is 0.97. What is the system reliability? What will be the system reliability if the sensing and switching element is 100% reliable?

Solution :

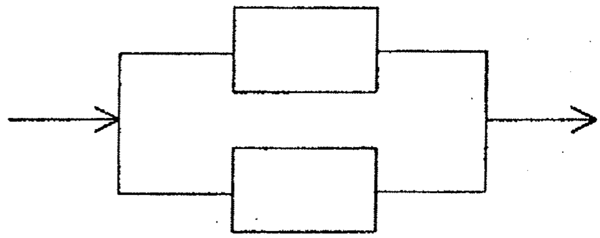


Fig. 20

Here, the failure rate $\lambda = 500$ failures $/10^6$ hours

$= 500/10^6$ failure per hour

$= 0.0005$ failure per hours,

and the operating tim e $t = 1000$ hours.

Hence $\lambda t = 0.0005 \times 1000 = 0.5$

Given that the failure rate of the sensing and switching element is 0.97.

Hence the reliability of sensing and switching element is $r_{SS} = 1 - 0.97 = 0.03$.

So, when the sensing and switching element is not perfect, the system reliability is given by

$$R_2(t) = e^{-\lambda t}\left[1 + r_{ss}(\lambda t)\right] = e^{-0.5}(1 + 0.015) = 0.6156$$

When the sensing and switching element is perfect, the system reliability is iven by

$$R_2(t) = e^{-\lambda t}(1 + \lambda t) = e^{-0.5}(1 + 0.05) = 0.909$$

5. Self Assessment questions / exercise

5.1 Define the reliability of a system. Compare the reliability of a series and parallel system.

5.2 Show that the reliability function for random failures is an exponential distribution. How system reliability can be improved?

5.3 Define system Reliability. Find the reliability of a system with two components of which one is a stand-by. The components are connected in parallel.

5.4 What is MTBF? The failure rate of an electronic sub-system is 0.0005 failurs/hour. If an operational period of 500 hours with probability of success $p = 0.95$ is desired, what level of parallel redundancy is needed?

5.5 Show that $R(t) = \exp\left[-\int_0^t \lambda(t)\,dt\right]$ where $R(t)$ is the reliability function and $\lambda(t)$ represents the failure rate.

5.6 In a system, there are n number of components connected in parallel with reliability $R_i(t), i = 1, 2, \ldots, n$. Find the reliability of the system. If $R_1(t) = R_2(t) = \ldots R_n(t) = e^{-\lambda t}$ then what will be the expression of system reliability?

5.7 How many identical components each of which is 90% reliable over a period of 50 hrs. be used to obtain a 99.99% parallel redundancy system over 100 hrs.

5.8 Show that the reliability of an item can be expressed as an exponential function i.e., $\exp\left[-\int_0^t \lambda(t)\,dt\right]$.

5.9 An industrial process is controlled by a computer and two similar components are operated in stand-by redundancy, such that, if a computer fails, another is instantaneously brought into use in its place. The failure rate of each computer is given by $\lambda = 0.01$ failure/hr. Compare the improvement in reliability over a single computer when one and then two computers are stand-by. The operating period is 100 hrs. and the switch system is 100% perfect.

5.10 A system consists of 5 identical and independent units with one unit operating and 4 units stand-by. One of the stand-by units takes over when any operating unit fails. Assuming that the switching device is perfect. Obtain the system reliability for a period of 100 hours if each of the 5 units has a failure rae of 100 failures/10^6 hours.

5.11 Obtain reliability over 100 hours period of a system consisting of two subsystem A and B connected in parallel where A consist of 4 identical components in series and B consist of identical components in parallel. Each component has a reliability 0.90 over a period of 100 hours.

5.12 The system connected in series of 500 transistors, 10500 resistors and 500 capacitors. Failure rate of these components are as follows:

transistor : $\lambda_t = 0.7 \times 10^{-7}$ per hour diode : $\lambda_d = 0.3 \times 10^{-6}$ per hour

resistor : $\lambda_r = 0.1 \times 10^{-6}$ per hour

capacity : $\lambda_c = 2 \times 0.2 \times 10^{-6}$ per hour

What is the failure rate of the system? What is the reliability system of 100 hours.

6. Suggested Further Readings

* Aggarwal, K.K., Reliability Engineering, 1993.
* Ramakumar Ramachandra, Enginering Relialiability Fundamentals and Applications, 1993.
* Mustafi, C.K., Statistical Methods in Managerial Decisions, Macmillan, India.
* Sinha, S.K. and Kale B.K., Life Testing and Reliability Estimation, Wiley Eastern, India.
* Swarup, K., Gupta, P.K. and Man Mohan, Operations Research, Sultan Chand & Sons.

--- 0 ---

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X Special Paper–OR

Module No. - 115(b)
GEOMETRIC PROGRAMMING

1. INTRODUCTION

In this module, we shall focus our attention on a rather interesting technique called 'Geometric Programming' for solving a special type of non-linear optimization problems. This technique was developed by R. Duffin and C. Zener (1964) for finding the solution of special type problems by considering the associated dual problem. This technique was given the name 'Geometric Programming' because the geometric-arithmetic mean inequality was the basis of its development. The advantage of it is that it is usually much simplier to work with the dual problem than with the primal.

Structure :

2. OBJECTIVES

The objectives of this module are to discuss a methodology called Geometric Programming for solving constrained and unconstrained optimization problems with posynomial form of objective functions and constraints.

3. KEYBOARDS

Geometric Programming, Monomial, Posynomial, Primal, Dual, Degre of difficulty, constained, unconstrained, optimization problem.

4. BASIC GEOMETRIC PROGRAMMING

4.1 Monomial and posynomial functions

Let $x_1, x_2, ..., x_n$ denote n real positive variables and $x = (x_1, x_2, ..., x_n)$ a vector with components x_i. A real valued function f of x, with the form

$$f(x) = cx_1^{a_1}x_2^{a_2}....x_n^{a_n} \tag{1}$$

where $e > 0$ and $a_1 \in \Re$, is called a monomial function or a monomial. The constant c is called the coefficient of the monomial where as the constants $a_1, a_2, .., a_n$ be the exponents of the monomial. As an example, $6.3x_1^3x_2^{-0.65}$ is a monomial of the variable x_1, x_2, with coefficient 6.3.

Monomials are closed under multiplication and division, i.e. if f and g are both monomials then fg and f/g are also monomials. A monomial raised to any power is also a monomial :

$$f(x)^\beta = \left(cx_1^{a_1}x_2^{a_2}...x_n^{a_n}\right)^\beta = c^\beta x_1^{\beta a_1}x_2^{\beta a_2}...x_n^{\beta a_n}$$

The term 'monomial', as used here (in the context of geometric programming) is similar to, but differs from the standard definition of 'monomial' used in algebra. In algebra, a monimial has the form (1), but the xponents $a_i\ (i = 1, 2, ..., n)$ must be nonnegative integers and the coefficient c is one. In this module, monomial refers to the definition mentioned earlier, in which the coefficient can be any positive number and the exponents can be any real numbers, including negative and fractional.

A sum of one or more monomials, i.e., a function of the form

$$f(x) = \sum_{j=1}^{m} c_j x_1^{a_{1j}} x_2^{a_{2j}} ... x_n^{a_{nj}} \quad (2)$$

where $c_j > 0$, is called a posynomial function, or, simply, a posynomial. The term posynomial is meant to suggest a combination of 'positive' and 'polynomial'. Any monomial is also a posynomial. Posynomials are closed under addition, multiplication. Again, if a posynomial function is divided by a monomial, then the resulting function will also be a posynomial. If β is a nonnegative integer and *f* is posynomial, then f^{β} is a posynomial (since it is the product of β posynomials).

Let us take a few example. Suppose *x*, *y* and *z* are (positive) variables. The functions (or expressions)

$$5y, 0.56, 2x\sqrt{y/z}, 6x^4 y^{-0.75} z^2$$

are monomials (hence, also posynomials). The functions

$$0.45 + x/z, 3(7 + yz)^{12}, 3x + 7y + 2z$$

are posynomials but not monomials. The functions

$$-2.5, 4(1 + yz)^{4.1}, 2x - 7y + z$$

are not posynomials or monomials. The function *f*(*x*) given by

$$f(x, y, z) = 12x^2 y^3 z^{0.5} + 16x^{-2} y^3 + 7zy$$

is a posynomial.

4.2 Geometric-Arithmetic mean inequality

The general arithmetic mean-Geometric mean inequality also known as Cauchy's inequality for any *n* nonnegative numbers $x_1, x_2, ..., x_n$ is

$$\frac{x_1, x_2, \ldots, x_n}{n} \geq (x_1 x_2 \ldots x_n)^{\frac{1}{n}} \tag{3}$$

with the equality sign holding true if any one if $x_1 = x_2 = \ldots = x_n$.

If we set first m_1 of the numbers x_i equal to the same value, say u_1, the next m_2 of the numbers x_i equal to the same value, say u_2, ... and the last m_n of the numbers x_i equal to the same value, say u_n, then the inequality (3) reduces to

$$\frac{m_1 u_1 + m_2 u_2 + \ldots + m_N u_N}{m_1 + m_2 + \ldots + m_N} \geq \left(u_1^{m_1} u_2^{m_2} \ldots u_N^{m_N}\right)^{\frac{1}{m_1 + m_2 + \ldots + m_N}}$$

$$\frac{m_1 u_1 + m_2 u_2 + \ldots + m_N u_N}{n} \geq \left(u_1^{m_1} u_2^{m_2} \ldots u_N^{m_N}\right)^{\frac{1}{n}} \text{ where } n = m_1 + m_2 + \ldots + m_N$$

This is valid for any set of nonnegative numbers $u_1, u_2 \ldots u_N$ and $m_1, m_2, \ldots, m_N$.

Let $y_1 = m_1/n, y_2 = m_2/n, y_N = m_N/n$

Then from the above inequality, we have

$$y_1 u_1 + y_2 u_2 + \ldots + y_N u_N \geq u_1^{y_1} u_2^{y_2} \ldots u_N^{y_N}$$

where $y_1 + y_2 + \ldots + y_N = 1$

Let $y_1 u_1 = U_1, y_2 u_2 = U_2, \ldots, y_N u_N = U_N$

Hence, $U_1 + U_2 + \ldots + U_n \geq \left(\frac{U_1}{y_1}\right)^{y_1} \left(\frac{U_2}{y_2}\right)^{y_2} \ldots \left(\frac{U_N}{y_N}\right)^{y_N}$

where $y_1 + y_2 + \ldots + y_N = 1$

4.3 Unconstrained Geometric Programming

The unconstrained Geometric Programming problem is defined as follows:

Find $X = (x_1, x_2, \ldots, x_n)^T$ that minimizes the objective function

$$f(X)=\sum_{j=1}^{N}U_j(X)=\sum_{j=1}^{N}c_j\left(x_1^{a_{1j}}x_2^{a_{2j}}...x_n^{a_{nj}}\right)=\sum_{j=1}^{N}c_j\prod_{i=1}^{n}x_i^{a_{ij}} \quad (4)$$

where $c_j>0, x_i>0$ and a_{ij} are real constant.

Solution :

From Geometric-Arithmetic mean inequality, we have

$$U_1+U_2+...+U_n\geq\left(\frac{U_1}{y_1}\right)^{y_1}\left(\frac{U_2}{y_2}\right)^{y_2}...\left(\frac{U_N}{y_N}\right)^{y_N} \quad (5)$$

where $y_1+y_2+...+y_N=1$ (6)

The left hand side of the inequality (5) i.e., original function $f(X)$ is called the primal function, the right hand side of (5) is called the pre-dual function.

Since $U_j=c_j\prod_{i=1}^{n}x_i^{a_{ij}}, j=1,2,...,N$

Then the pre-dual function can be written as

$$\left(\frac{U_1}{y_1}\right)^{y_1}\left(\frac{U_2}{y_2}\right)^{y_2}...\left(\frac{U_N}{y_N}\right)^{y_N}=\left(\frac{c_1\prod_{i=1}^{n}x_i^{a_{i1}}}{y_1}\right)^{y_1}\left(\frac{c_2\prod_{i=1}^{n}x_i^{a_{i2}}}{y_2}\right)^{y_2}....\left(\frac{c_N\prod_{i=1}^{n}x_i^{a_{iN}}}{y_N}\right)^{y_N}$$

$$=\left(\frac{c_1}{y_1}\right)^{y_1}\left(\frac{c_2}{y_2}\right)^{y_2}...\left(\frac{c_N}{y_N}\right)^{y_N}\left\{\left(\prod_{i=1}^{n}x_i^{a_{i1}}\right)^{y_1}\left(\prod_{i=1}^{n}x_i^{a_{i2}}\right)^{y_2}...\left(\prod_{i=1}^{n}x_i^{a_{iN}}\right)^{y_N}\right\}$$

$$=\left(\frac{c_1}{y_1}\right)^{y_1}\left(\frac{c_2}{y_2}\right)^{y_2}...\left(\frac{c_N}{y_N}\right)^{y_N}\left\{\left(\sum_{x_1^{j=1}}^{N}a_{1}y_j\right)\left(\sum_{x_2^{j=1}}^{N}a_{2}y_j\right)...\left(\sum_{x_n^{j=1}}^{N}a_{nj}y_j\right)\right\} \quad (7)$$

If we select the y_j in such a way that y_j satisfy the relation (6) and also the relation

$$\sum_{j=1}^{N}a_{ij}y_j=0, i=1,2,...,n \quad (8)$$

Then the expression (7) reduces to

$$\left(\frac{U_1}{y_1}\right)^{y_1}\left(\frac{U_2}{y_2}\right)^{y_2}\cdots\left(\frac{U_N}{y_N}\right)^{y_N}=\left(\frac{c_1}{y_1}\right)^{y_1}\left(\frac{c_2}{y_2}\right)^{y_2}\cdots\left(\frac{c_N}{y_N}\right)^{y_N} \tag{9}$$

Equation (6) is called the normality condition and the equation (8) are called the orthogonality condition for the weight $y_1, y_2, ..., y_N$.

Thus the inequality (5) becomes

$$U_1+U_2+...U_N \geq \left(\frac{c_1}{y_1}\right)^{y_1}\left(\frac{c_2}{y_2}\right)^{y_2}\cdots\left(\frac{c_N}{y_N}\right)^{y_N}$$

i.e. $f(x)\geq\varphi(y)$ (say) (10)

the right hand side of (10) i.e., $\varphi(y)$ is called the dual function and y is a vector with components $y_1, y_2, ..., y_N$.

We can be proved that

$$\underset{x}{\text{Minimize}}\, f(x)=\underset{y}{\text{Maximize}}\ \varphi(y)$$

In this way, we have converted the problem of minimizing $f(x)$ into a new problem of maximizing $\varphi(y)$ the original problem of minimization is referred to as primal one and the related problem of maximization as its dual.

The maximum of dual function subject to the orthogonality and normality conditions is a sufficient condition for $f(x)$ to be a global minimum.

For equality in the geometric-arithmetic mean inequality (5), all the $\frac{U_j}{y_j}$ must be equal. Therefore, if we let

$$\frac{U_1}{y_1}=\frac{U_2}{y_2}=...=\frac{U_N}{y_N}=R,\ \text{i.e., } \frac{U_j}{y_j}=R\ \text{(say)}$$

Then $\sum_{j=1}^{N}U_j=R\sum_{j=1}^{N}y_j=R$ [by (6) $\sum_{j=1}^{N}y_j=1$]

Hence minimum $f(x)=Min\sum_{j=1}^{N}U_j=R=f(x^*)$

Then $\frac{U_j}{y_j} = f(x^*), j = 1,2,...,N$

Hence, $U_j = f(x^*) y_j, j = 1,2,...,N$ (11)

or, $c_j \prod_{i=1}^{n} x_i^{a_{ij}} = f(x^*) y_j, j = 1,2,...,N$ (12)

From these relations, the optimum values of x_i are obtained.

Degree of Difficulty

The quantity N-n-1 (where N and n denote the total number of terms in he posynomial and number of decision variables respectively) is termed as degree of difficulty in geometric programming. If N-n-1=0, the problem is said to have a zero difficulty. In this case, the unknows, $y_j\ (j = 1,2,...,N)$ can be determined uniquely from the orthogonality and normality conditions. If $N - n - 1 > 0$, there will be an infinite number of solutions of $y_j\ (j = 1,2,...,N)$. For problem with negative degree of difficulty, geometric programming is not applicable.

Primal and Dual problems

For unconstrained Geometric Programming, the Primal and Dual problems are as follows:

Primal Problem :

Find $x = (x_1, x_2, ..., x_n)^T$

so that $f(x) = \sum_{j=1}^{N} c_j x_1^{a_{1j}} x_2^{a_{2j}} ... x_n^{a_{nj}}$ is minimum

and $x_1 > 0, x_2 > 0, ..., x_n > 0.$

Dual Problem :

Find $y = (y_1, y_2, ..., y_N)^T$ so that

$$\varphi(y) = \left(\frac{c_1}{y_1}\right)^{y_1} \left(\frac{c_2}{y_2}\right)^{y_2} ... \left(\frac{c_N}{y_N}\right)^{y_N} = \prod_{j=1}^{N} \left(\frac{c_j}{y_j}\right)^{y_j}$$

$$\log \varphi(y) = \log \left[\prod_{j=1}^{N} \left(\frac{c_j}{y_j}\right)^{y_j} \right]$$

is maximum subject to the constraints

$y_1 + y_2 + ... + y_N = 1$

and $\sum_{j=1}^{N} a_{ij} y_j = 0, i = 1, 2, ..., n$

Example 1

Solve the following GP problem :

Minimize $f(x) = 7x_1x_2^{-1} + 7x_2x_3^{-2} + 5x_1^{-3}x_2x_3 + x_1x_2x_3$

Solution : Here, $c_1 = 7, c_2 = 3, c_3 = 5, c_4 = 1$

and $U_1 = 7x_1x_2^{-1}, U_2 = 3x_2x_3^{-2}, U_3 = 5x_1^{-3}x_2x_3, U_4 = x_1x_2x_3$

For this problem, the normality condition is

$$y_1 + y_2 + y_3 + y_4 = 1 \quad (1)$$

where y_1, y_2, y_3, y_4 are the dual variables.

And the orthogonality conditons are

$$\begin{pmatrix} 1 & 0 & -3 & 1 \\ -1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = 0$$

i.e., $y_1 - 3y_3 + y_4 = 0$ (ii)

$-y_1 + y_2 + y_3 + y_4 = 0$ (iii)

$-2y_2 + y_3 + y_4 = 0$

In this case, $n = 3, N = 4$. So, $N - n - 1 = 0$ i.e., the degree of difficulty is zero. Therefore, we wil get the unique solution of y_j's.

Now, subtracting (iii) from (i), we have $y_1 = \frac{1}{2}$.

Again, subtracting (iv) from (i), $y_1 + 3y_2 = 1$ which implies $y_2 = \frac{1}{6}$ as $y_1 = \frac{1}{2}$

Now, by (i) – (ii), we have $y_2 + 4y_3 = 1$ or, $y_3 = \frac{5}{24}$.

Hence, from (ii), $y_4 = y_1 - y_2 - y_3 = \frac{1}{2} - \frac{1}{6} - \frac{5}{24} = \frac{1}{8}$

$$\therefore \min f(x) = f(x^*) = \left(\frac{c_1}{y_1}\right)^{y_1}\left(\frac{c_2}{y_2}\right)^{y_2}\left(\frac{c_3}{y_3}\right)^{y_3}\left(\frac{c_4}{y_4}\right)^{y_4}$$

$$= \left(\frac{7}{1/2}\right)^{1/2}\left(\frac{3}{1/6}\right)^{1/6}\left(\frac{5}{5/24}\right)^{5/24}\left(\frac{1}{1/8}\right)^{1/8}$$

$$= (14)^{1/2}(18)^{1/6}(24)^{5/24}(8)^{1/8} = 15.23$$

We have $U_j = y_j f(x^*)$

$\therefore U_1 = y_1 f(x^*)$ or, $7x_1x_2^{-1} = (1/2)\times 15.23$ or, $7x_1x_2^{-1} = 7.62$ (v)

Again, $U_2 = y_2 f(x^*)$ or, $3x_2x_3^{-2} = (1/6)\times 15.23$ or, $3x_2x_3^{-2} = 2.54$ (vi)

$U_3 = y_3 f(x^*)$ or, $5x_1^{-3}x_2x_3 = (5/24)\times 15.23$ or, $5x_1^{-3}x_2x_3 = 3.17$ (vii)

$U_4 = y_4 f(x^*)$ or, $x_1x_2x_3 = (1/8)\times 15.23$ or, $x_1x_2x_3 = 1.90$ (viii)

Now dividing (viii) by (vii), we have

$$\frac{x_1^4}{5} = \frac{1.90}{3.17} \text{ or, } x_1 = 1.3157$$

From (v), $7x_1x_2^{-1} = 7.62$ or, $7.62x_2 = 7x_1$ or, $x_2 = 1.209$

From (vi), $3x_1x_3^{-2} = 2.54$ or, $2.54x_3^2 = 3x_2$ or, $x_3 = 1.195$

Hence the optimal solution of the problem is $x_1^* = 1.3157, x_2^* = 1.2097, x_3^* = 1.195$.

Example 2

Solve the following problem

Maximize $f(x) - 5x_1x_2^{-1} + 2x_1^{-1}x_2 + 5x_1 + x_2^{-1}$

Solution : Here $c_1 = 5, c_2 = 2, c_3 = 5, c_4 = 1$ and $U_1 = 5x_1x_2^{-1}, U_2 = 2x_1^{-1}x_2, U_3 = 5x_1, U_4 = x_2^{-1}$

Now the normality condition is

$y_1 + y_2 + y_3 + y_4 = 1$ (i)

and the orthogonality conditions are

$$\begin{pmatrix} 1 & -1 & 1 & 0 \\ -1 & 1 & 0 & -1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = 0$$

$$\text{i.e.,} \quad \left.\begin{aligned} y_1 - y_2 + y_3 &= 0 \\ -y_1 + y_2 - y_4 &= 0 \end{aligned}\right\} \qquad \text{(ii)}$$

Here, $n = 2, N = 4$.

Hence $N - n - 1$ i.e., the degree of difficulty is one.

In this case, we express three y_j's in terms of remaining one.

Let us solve y_1, y_2, y_3 in terms of y_4.

Hence from (*ii*) and (*i*), we have

$$y_1 - y_2 + y_3 = 0 \qquad \text{(iii)}$$

$$-y_1 + y_2 = y_4 \qquad \text{(iv)}$$

$$y_1 + y_2 + y_3 = 1 - y_4 \qquad \text{(v)}$$

Now by (v) – (iii), we have $2y_2 = 1 - 2y_4$ or, $y_2 = (1-y_4)/2$.

Putting the above value of y_2 in (iv), we have $y_1 = (1-3y_4)/2$.

Again, adding (iii) and (iv), we have $y_3 = y_4$.

Hence, the corresponding dual problem of the given problem can be written as

$$Max\ \phi(y) = \left(\frac{c_1}{y_1}\right)^{y_1}\left(\frac{c_2}{y_2}\right)^{y_2}\left(\frac{c_3}{y_3}\right)^{y_3}\left(\frac{c_4}{y_4}\right)^{y_4} = \left(\frac{5}{\frac{1-3y_4}{2}}\right)^{\frac{1-3y_4}{2}}\left(\frac{2}{\frac{1-y_4}{2}}\right)^{\frac{1-y_4}{2}}\left(\frac{5}{y_4}\right)^{y_4}\left(\frac{1}{y_4}\right)^{y_4}$$

$$= \left(\frac{10}{1-3y_4}\right)^{\frac{1-3y_4}{2}}\left(\frac{4}{1-y_4}\right)^{\frac{1-y_4}{2}}\left(\frac{5}{y_4}\right)^{y_4}\left(\frac{1}{y_4}\right)^{y_4}$$

Since the maximization of $\varphi(y)$ is equivalent to the maximizaton of $\log\varphi(y)$, we will maximize $\log\varphi(y)$ for convenience.

Thus

$$\log\varphi(y) = (1-3y_4)\{\log 10 - \log(1-3y_4)\}/2 + (1-y_4)\{\log 4 - \log(1-y_4)\}/2 + y_4\{\log 5 - \log y_4\} - y_4 \log y_4$$

Since $\log\varphi(y)$ is expressed as a function of y_4. The necessary condition of maximum of $\log\varphi(y)$ gives

$$\partial[\log\phi(y)]/\partial y_4 = 0$$

$$\text{i.e. } \log\left[\frac{(1-3y_4)^{3/2}(1-y_4)^{1/2}}{y_4^2}\right] = \log\left(\frac{2\times 10^{3/2}}{5}\right)$$

$$\text{or, } \sqrt{\frac{(1-3y_4)^3(1-y_4)}{y_4^2}} = \frac{2\times 10^{3/2}}{5} \text{ or, } \sqrt{\frac{(1-3y_4)^3(1-y_4)}{y_4^2}} = 12.6 \qquad \text{(vi)}$$

From (vi), the value of y_4^* can be obtained by using a trial and error process given in Table 1,

Table 1 : Computation of y_4^* by trial and error method

Value of y_4^*	Value of left hand side expression of (vi)
0.25	1.732
0.20	5.657
0.18	8.72
0.16	13.42
0.163	12.59
0.162	12.85

From Table 1, we find that $y_4^* = 0.163$ (approx.)

Hence $y_1^* = (1-3y_4^*)/2 = 0.26, y_2^* = (1-y_4^*)/2 = 0.42, y_3^* = y_4^* = 0.163$

The optimal value of the objective function is given by

$$\phi(y_4^*) = f(x^*) = \left(\frac{5}{0.26}\right)^{0.26}\left(\frac{2}{0.42}\right)^{0.42}\left(\frac{5}{0.163}\right)^{0.163}\left(\frac{5}{0.163}\right)^{0.163} \approx 9.661$$

Now, using $U_j = y_j f(x^*)$, we have

$$U_3 = y_3^* f(x^*) \text{ i.e., } 5x_1^* = 0.16\times 9.661 = 1.546 \text{ or, } x_1^* = 1.546/5 = 0.309$$

$$U_4 = y_3^* f(x^*) \text{ i.e., } x_2^{*-1} = 0.16\times 9.661 = 1.546 \text{ or, } x_2^* = 0.647$$

Hence the required optimal solution is $x_1^* = 0.309, x_2^* = 0.647$ and the minimum value of the given objective function is 0.9661.

4.4 Constrained Optimizaton Problem

The constrained optimization problem is as follows:

Find $x = (x_1, x_2, \ldots, x_n)$ which minimizes the objective function

$$f(x) = \sum_{j=1}^{N_0} c_{0j} \prod_{i=1}^{n} x_i^{a_{0ij}}$$

and satisfies the constraints

$$g_k(x)=\sum_{j=1}^{N_k} c_{kj}\prod_{i=1}^{n} x_i^{a_{kij}} <,>,=1, k=1,2,...,m$$

where the coefficient $c_{0j}(j=1,2,...,N_0)$ and $c_{kj}(k=1,2,...m; j=1,2,...,N_k)$ are positive numbers, the exponents $a_{0ij}(i=1,2,...n; j=1,2,...,N_0)$ and $a_{kij}(k=1,2,...,m; i=1,2,...,n; j=1,2,...,N_k)$ are any real numbes, m indicates the total number of constraints, N_0 represents the number of terms in the objective function and N_k denotes the number of terms in the k-th constraint. The decision variables $x_1, x_2, ..., x_n$ are assumed to take only positive valeus.

4.4.1 Problem with less than type inequalities

Primal problem :

Find $x=(x_1,x_2,...,x_n)$
such that $f(x)$ is minimum
subject to the constraints

$g_1(x)\le 1, g_2(x)\le 1,...,g_m(x)\le 1$ and $x_i>0, i=1,2,...,n, c_{kj}>0$ and a_{kij} are real numbes

[Note : $g_0\to$ primal function, $x_1,x_2,...,x_n\to$ primal variables, $g_k(x)\le 1(k=1,2,...,m)\to$ primal constraints $x_i>0(i=1,2,...,n)\to$ positive restrictions, n = number of primal variables, m = number of primal constraints, $N=N_0+N_1+...+N_m$ = total number of terms in the posynomials, N–n–1=degree of difficulty]

Dual problem :

Find $\lambda=(\lambda_{01},\lambda_{02},...,\lambda_{0N_0},\lambda_{01},\lambda_{11},\lambda_{12},...,\lambda_{1N_1},...,\lambda_{m1},\lambda_{m2},...,\lambda_{nN_m})$

such that $\phi(\lambda)=\prod_{k=0}^{m}\prod_{j=1}^{N_k}\left(\frac{c_{kj}}{\lambda_{kj}}\sum_{l=1}^{N_k}\lambda_{kl}\right)^{\lambda_{kj}}$ is maximum subject to the constraints

$\sum_{j=1}^{N_0}\lambda_{0j}=1$ [normality condition]

$\sum_{k=0}^{m}\sum_{j=1}^{N_k}a_{kij}\lambda_{kj}=0, i=1,2,...,n$ [orthogonality condition]

and $\lambda_{kj}\ge 0, c_{kj}>0$ and a_{kij} are real numbers.

[Note : $\phi(\lambda)\to$ dual function $\lambda_{01},\lambda_{02},...,\lambda_{mN_m}\to$ dual variables $\lambda_{kj}\ge 0,(j=1,2,...,N_k, k=1,2,...,m)$ are nonnegative restrictions, $N=N_0+N_1+...+N_m$ = number of dual variables, $n+1$ = number of dual constraints]

Optimal Decision variables :

For problems with zero degree of difficulty, the solution for λ^* is unique. Once the optimum values of λ_{kj} are obtained, the maximum of the dual function $\varphi(\lambda^*)$ are obtained, which is also the minimum of the primal function $f(x^*)$. The decision variables x^* can be obtained by solving the following equations simultaneously.

$$\lambda_{0j}^* = \frac{c_{0j}\prod_{i=1}^{n}(x_i^*)^{a_{0ij}}}{f(x^*)}, j = 1,2,...,N_0$$

$$\frac{\lambda_{kj}^*}{\sum_{l=1}^{N_k}\lambda_{kl}^*} = c_{kj}\prod_{i=1}^{n}(x_i^*)^{a_{kij}}, j = 1,2,..,N_k; k = 1,2,...,m$$

Example 3

Solve the following constrained optimiztion problem.

Minimize $Z = 0.188x_1x_3$ subject the constraints

$$1.75x_1x_2^{-1}x_3^{-1} \le 1$$

and $900x_1^{-2} + x_1^{-2}x_2^2 \le 1$

Solution : Here the number of primal variables (n) = 3, N = total number of terms in the posynomials = 4. Therefore, the degree of difficulty = N − n − 1 = 0.

The given problem can be written as

$$Z = c_{01}\prod_{i=1}^{3} x_i^{a_{0i1}}$$

subject to the constraints

$$c_{11}\prod_{i=1}^{3} x_i^{a_{1i1}} \le 1, \sum_{j=1}^{2} c_{2j}\prod_{i=1}^{3} x_i^{a_{2ij}} \le 1$$

where

$c_{01} = 0.188, a_{011} = 1, a_{021} = 0, a_{031} = 1, c_{11} = 1.75, a_{111} = 1, a_{121} = -1, a_{131} = -1, c_{21} = 900, c_{22} = 1,$

$a_{211} = -2, a_{221} = 0, a_{231} = 0, a_{212} = -2, a_{222} = 2, a_{232} = 0$

Here $N_0 = 1, N_1 = 1, N_2 = 2$

The corresponding dual problem is to find $\lambda = (\lambda_{01}, \lambda_{11}, \lambda_{21}, \lambda_{22},)$ so as to

$$\text{Maximize } \phi(\lambda) = \left\{\frac{c_{01}}{\lambda_{01}}(\lambda_{11})\right\}^{\lambda_{01}} \left\{\frac{c_{11}}{\lambda_{11}}(\lambda_{11})\right\}^{\lambda_{11}} \left\{\frac{c_{21}}{\lambda_{21}}(\lambda_{21}+\lambda_{22})\right\}^{\lambda_{21}} \left\{\frac{c_{22}}{\lambda_{22}}(\lambda_{21}+\lambda_{22})\right\}^{\lambda_{22}}$$

$$= (0.188)^{\lambda_{01}} (1.75)^{\lambda_{11}} \left\{ \frac{900}{\lambda_{21}} (\lambda_{21} + \lambda_{22}) \right\}^{\lambda_{21}} \left\{ \frac{1}{\lambda_{22}} (\lambda_{21} + \lambda_{22}) \right\}^{\lambda_{22}}$$

subject to

$$\sum_{j=1}^{N_0} \lambda_{0j} = 1 \Rightarrow \lambda_{01} = 1 \text{ [Since } N_0 = 1]$$

$$\sum_{k=0}^{m} \sum_{j=1}^{N_k} a_{kij} \lambda_{kj} = 0 \Rightarrow \sum_{k=0}^{2} \sum_{j=1}^{N_k} a_{kij} \lambda_{kj} = 0$$

i.e., $a_{011}\lambda_{01} + a_{111}\lambda_{11} + a_{211}\lambda_{21} + a_{212}\lambda_{22} = 0$

$a_{021}\lambda_{01} + a_{121}\lambda_{11} + a_{221}\lambda_{21} + a_{222}\lambda_{22} = 0$

$a_{031}\lambda_{01} + a_{131}\lambda_{11} + a_{231}\lambda_{21} + a_{232}\lambda_{22} = 0$

Putting the values of a_{kij}

$\lambda_{01} + \lambda_{11} - 2\lambda_{21} - 2\lambda_{22} = 0$ (i)

$-\lambda_{11} + 2\lambda_{22} = 0$ (ii)

$\lambda_{01} - \lambda_{11} = 0$ (iii)

Adding (i) and (ii), we have $\lambda_{01} - 2\lambda_{21} = 0$ or, $\lambda_{21} = 1/2$.

From (iii), $\lambda_{01} = \lambda_{11}$ i.e., $\lambda_{11} = 1$.

From (ii), $\lambda_{11} = 2\lambda_{22}$ i.e., $\lambda_{22} = \frac{1}{2}$

Hence, $\lambda_{01} = 1, \lambda_{11} = 1, \lambda_{21} = 1/2, \lambda_{22} = 1/2$ (iv)

Hence Min $f(x) = f(x) =$ Max $\varphi(\lambda^*)$

$$= (0.188)^{\lambda_{01}} (1.75)^{\lambda_{11}} \left\{ \frac{900}{\lambda_{21}} (\lambda_{21} + \lambda_{22}) \right\}^{\lambda_{21}} \left\{ \frac{1}{\lambda_{22}} (\lambda_{21} + \lambda_{22}) \right\}^{\lambda_{22}}$$

$= 19.74$

The optimal value of decision variables are obtained as

$$\lambda_{01}^* = \frac{c_{01} (x_1^*)^{a_{011}} (x_2^*)^{a_{021}} (x_3^*)^{a_{031}}}{f(x^*)} \text{ i.e.,} 1 = \frac{0.188}{19.74} x_1^* x_3^* \text{ i.e., } x_1^* x_3^* = \frac{19.74}{0.188} \quad \text{(v)}$$

$$\frac{\lambda_{11}^*}{\lambda_{11}^*} = c_{11} (x_1^*)^{a_{111}} (x_2^*)^{a_{121}} (x_3^*)^{a_{131}} \text{ i.e.,} 1.75 x_1^* x_2^{*-1} x_3^{*-1} = 1 \quad \text{(vi)}$$

$$\frac{\lambda_{21}^*}{\lambda_{21}^*+\lambda_{22}^*}=c_{21}\left(x_1^*\right)^{a_{211}}\left(x_2^*\right)^{a_{221}}\left(x_3^*\right)^{a_{231}}\ i.e., 900x_1^{*-2}=\frac{1}{2} \qquad \text{(vii)}$$

$$\frac{\lambda_{22}^*}{\lambda_{21}^*+\lambda_{22}^*}=c_{22}\left(x_1^*\right)^{a_{212}}\left(x_2^*\right)^{a_{222}}\left(x_3^*\right)^{a_{232}}\ i.e., x_1^{*-2}x_2^{*2}=\frac{1}{2} \qquad \text{(viii)}$$

From equations (vi) – (viii), we can easily obtain $x_1^*=30\sqrt{2}, x_2^*=30, x_3^*=1.75\sqrt{2}$.

Hence the required optimal solution is $x_1^*=30\sqrt{2}, x_2^*=30, x_3^*=1.75\sqrt{2}$ and the minimum value of the given objective function is 19.74.

4.4.2 Problme with equality constraints
Primal problem :

Find $x=(x_1,x_2,...,x_n)$
such that $g_0(x)$ is minimum
subject to the constraints

$$g_i(x)=\sum_{r=1}^{P_i}C_{ir}P_{ir}(x)=1$$

Where P_i denotes the number of terms in the i-th constraints and

$$P_{ir}(x)=\prod_{j=1}^{k}(x_j)^{a_{irj}}$$

[Note : $g_0 \to$ primal function $x_1,x_2,...,x_n \to$ primal variables, $g_k(x)=1(k=1,2,...,m)\to$ primal constraints, $x_i>0(i=1,2,...,n)\to$ positive restrictions, n = number of primal variables, m = number constraints, $N=N_0+N_1+...+N_m$ = total number of terms in the posynomials, N–n– 1=degree of difficulty]

Dual problem
Find λ

such that $$\phi(\lambda)=\prod_{j=1}^{n}\left(\frac{c_j}{\lambda_j}\right)^{\lambda_j}\prod_{i=1}^{m}\left[\sum_{r=1}^{P_i}\left(\frac{C_{rj}}{\lambda_{rj}}\right)^{\lambda_j}\right]\prod_{i=1}^{m}(V_i)^{v_i}$$

where $V_i=\sum_{r=1}^{P_i}y_{ir}$

is maximum subject to the constraints

$$\sum_{j=1}^{N_0} \lambda_{0j} = 1 \text{ [normality condition]}$$

$$\sum_{k=0}^{m} \sum_{j=1}^{N_k} a_{kij} \lambda_{kj} = 0, i = 1, 2, ..., n \text{ [orthogonality condition]}$$

and $\lambda \geq 0, c_{kj} > 0$ and a_{kij} are real numbers.

[Note : $\phi(\lambda \rightarrow)$ dual function, $\lambda_{01}, \lambda_{02}, ..., \lambda_{mN_m} \rightarrow$ dual variables $\lambda_{kj} \geq 0, (i = 1, 2, ..., N_k, k = 1, 2, ..., m)$ are nonnegative restrictions, $N = N_0 + N_1 + ... + N_m$ = number of dual variables, n+1 = number of dual constraints]

5. SELF ASSESSMENT QUESTIONS/ EXERCISE

5.1 What is Geometric Programming?

5.2 State whether each of the following function is a polynomial, posynomial or both.

(a) $f = 4 - x_1^2 + 6x_1x_2 + 3x_2^2$

(b) $f = 12 + 2x_1^2 + 5x_1x_2 + x_2^2$

(c) $f = 4 + 2x_1^2x_2^{-1} + 3x_2^{-4} + 5x_1^{-1}x_2^3$.

5.3 How is the degree of difficulty defined for a constrained geometric programming problem?

5.4 What is arithmetic-geometric inequality?

5.5 What is normality condition in a geometric programming problem?

5.6 Define a complementary geometric programming problem.

5.7 Solve the following problems :

(a) Minimize $Z = 5x_1x_2^{-1} + 2x_1^{-1}x_2 + 5x_1 + x_2^{-1}$

(b) Minimize $Z = 60x_1^{-3}x_2^{-2} + 50x_1^3x_2 + 20x_1^{-3}x_2^3$

(c) Minimize $Z = 20x_1x_3 + 40x_2x_3 + 80x_1x_2$ subject to $\dfrac{8}{x_1x_2x_3} \leq 1$

(d) Minimize $Z = 2x_1x_2^{-3} + 4x_1^{-1}x_2^{-2} + 32x_1x_2/3$ subject to $10x_1^{-1}x_2^2 = 1$

6. SUGGESTED FURTHER READINGS

* Duffin, R.J., Peterson, E. and Zener, C., *Geometric Programming*, Wiley, New York, 1967.
* Rao, S.S., *Engineering Optimization-Theory and Practice*, New Age International, 2001.
* Swarup, K., Gupta, P.K. and Man Mohan, *Operations Research*, Sultan Chand & Sons.
* Sharma, J.K., *Operations Research – Theory and Applications*, Macmillan.
* Sharma, S.D., *Operations Research, Kedar Nath Ram Nath & Co.*, India.

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART - II

Module No - 116

SEQUENCING

1. Introduction : The problem of sequencing is to determine an appropriate order/sequence for a series of jobs to be done on a finite number of service facilities, in some preassigned order, so as to optimize the total involved cost/time. A practical situation may correspond to an industry producing a number of products, each of which are processed through a finite number of different machines.

Let, in an industry, there be n jobs, each of which is to be performed one at a time, at each of m different machines over a variety of combinations. We are given the order in which these machines are to be used for processing each job and also the actual or expected processing time taken by each job on the machines. Then the general sequencing problem is to determine the sequence, out of $(n!)^m$ possible ones, that minimizes the time from the starting of the first job upto the completion of the last job.

This module is organised as follows.

1. Introduction
2. Objectives
3. Keywords
4. Terminology and Notations
5. Basic Assumptions
6. Processing *n* Jobs Through 2 machines
   - 6.1 Optional Sequence algorithm
   - 6.2 Illustrative examples
7. Processing *n* Jobs through 3 Machines

2. Objectives :

Main objectives of this module is to make the readers familiar with the basic idea of sequencing a given number of jobs, each of which is to be processed through a given number of machines, i.e. the service facilities. For different number of jobs and different number of machines, the cases are discussed through illustrative examples.

3. Keywords :

Processing order, processing time, idle time, total elapsed time, optional sequence.

4. Terminology and Notations :

The following terminology and notations will be used in what follows :-

i) Number of Machines :

It means the service facilities through which a job must pass before it is completed.

e.g., in 'publishing a book', it has to be processed through composing, printing, binding, etc. In this case, 'publishing a book' constitutes the job and the different processes constitute the number of machines.

ii) Processing Order :

It refers to the order in which a job has to be processed through different machines.

e.g., when a tailor has to make a suit, he has to process the job in the order :

a) Marking on the cloth using measuring tape
b) Cutting
c) Stitching
d) Ironing.

iii) Processing Time :

This is the time for which a job is to be processed on a machine. The notation Tij will be used to denote the processing time required by the ith job on the jth machine.

iv) Total Elapsed Time :

This is the total time required by all machines from the starting og the first job to the completion of the last job, including idle time, if any. it will be denoted by T.

v) Idle Time :

It means the time for which a machine emains idle during the total elapsed time. The notatin Xij will be used to denote the idle time of the jth machine from the end of the (i-1)th job to the start of ith job.

vi) No Passing Rule :

This rule means that passing of a job to a machine other than the one scheduled in the processing order is not allowed. If each of the n jons is to be prossed through two machines A and B in the order AB, then this rule says that each job will go to machine A first and then to the machine B.

5. Basic Assumptions :

for a sequencing problem, following assumptions are made.

i) No machine can process more than one job at a time.

ii) Processing times Tij s are indepemndent of the processing order in which the jobs are processed.

iii) Each job, once started on a machins must be continued till the completion of it on that machine.

iv) Each job must be completed before any other job, which it has to precede, can begin.

v) A job is processed as soon as possible subject to ordering requirements.

vi) The time involved in moving a job from one machine to another is negligibly small.

vii) There is only one fo each type of machine.

viii) All jobs are known and are ready to start processing before the period under consideration begins.

6. Processing n jobs Through 2 Machines :

Let there be n jobs, each of which is to be processed thrugh two machines, say A_1 and A_2 in the order A_1A_2 That is, each job will go to machine A_1 first and then to A_2; or in other words, passing is not allowed. Let Tij be the time required for processing the i th job on the Ajth machine, i = 1,2,........,n; j =1,2.

Since passing is not allowed, it is obvious that all the n jobs are to be processed on machine A_1 without any idle time for it. On the other hand, machine A_2 is subject to its remaining idle at various stages. Let X_{i2} be the time for which machine A_2 remains idle after finishing (i - 1) the job and before starting to process the i th job, i = 1,2,.......,n. This, the total elapsed time T is given by $T=\sum_{i=1}^{n} T_{i2}+\sum_{i=1}^{n} X_{i2}$,

Where some of the X_{i2}'s may be zero.

Now the problem is to minimize the total elapsed time T. However, since $\sum_{i=1}^{n} x_{i2}$ is the total time for which the machine A_2 has to work and is thus fixed, it does not perform any role in minimizing T. Thus the problem reudces to that o minimizing $\sum_{i=1}^{n} x_{i2}$ A very convenient procedure for obtaining a sequence of performing jobs so as to minimize $\sum_{i=1}^{n} x_{i2}$ is well illustrated by the following Gantt Chart.

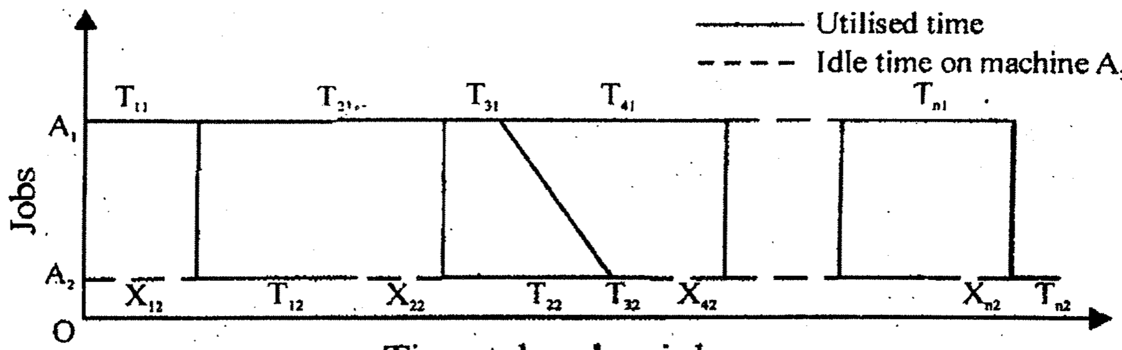


Figure 6.1. Gantt Chart

From the chart, it is clear that

$X_{12} = T_{11}$

$X_{22} = \{ T_{11} + T_{21} + - X_{12} - T_{12}$ if $T_{11} + T_{21} > X_{12} + T_{12}$

0 otherwise

$$= \max \{ T_{11} + T_{21} - X_{12} - T_{12}, 0\}$$

$$\therefore X_{12} + X_{22} = \max \{T_{11} + T_{21} - T_{12}, T_{11}\} \text{ Since } X_{12} = T_{11.}$$

Similarly, $X_{32} = \max \{ T_{11} + T_{21} + T_{31} - T_{12} - T_{22} - X_{12} - X_{22}, 0\}$

and so $X_{12} + X_{22} + X_{32} = \max \left\{\left(\sum_{i=1}^{3} Ti_1 - \sum_{i=1}^{2} Ti_2\right), \sum_{i=1}^{2} Xi_1\right\}$

$$= \max \left\{\left(\sum_{i=1}^{3} Ti_1 - \sum_{i=1}^{2} Ti_2\right), \left(\sum_{i=1}^{2} Ti_1 - Ti_1\right), T_{11}\right\}$$

In general, we have $\sum_{i=1}^{w} X_{i2_i} = \max\left\{\left(\sum_{i=1}^{w} T_{i1} - \sum_{i=1}^{n-1} T_{i2}\right), \left(\sum_{i=1}^{n-1} T_{i1} - \sum_{i=1}^{n-2} T_{i2}\right), \ldots, \left(\sum_{i=1}^{2} T_{i2} - T_{12}\right), T_{11}\right\}$

$$= \max_{k\varepsilon\{1,2,\ldots,n\}} \left\{\sum_{i=1}^{k} T_{i1} - \sum_{i=1}^{k-1} T_{i2}\right\}$$

Now, if we denote $\sum_{i=1}^{w} X_{i2}$ by D_n (S) , then the problem becomes that of finding the sequence S^* for processing the jobs 1,2,.......,n so that D_n (S^*) is minimum over all D_n(S) for any such sequence S, i.e., the inequlality $D_n(S^*) \leq D_n$ (S) holds for any sequence S of processing the jobs. In other words, we have to determine an optimal sequence S^* so as to minimise interchange of the consecutive jobs. Each such interchange of jobs gives a value of D_n(S) smaller than or equal to its value before the change.

6.1 Optimal sequence Algorithm :

An terative process for determining the optimal sequence for n-job 2-machine sequencing problem can be given stepwise as follows -

Step 1. :- Examine the T_{i1}'s and T_{i2}'s for i = 1, 2,..., n and find out $\min_{i\varepsilon\{1,2,\ldots,n\}} \{T_{i1}, T_{i2}\}$

Step 2(a) :- If this minimum be T_{k1} for some i = k, process the k-th job first of all.

(b) :- If T_{r2} be the minimum for some i = r, process the r-th job last of all.

Step 3(a) :- If there is a tie for minima $T_{k1} = T_{r2}$, process the k-th job first of all and the r th one in the last.

(b) :- If the tie for the minimum occurs among the T_{i1}'s, select the job corresponding to the minimum of T_{i1}'s and process if first of all.

(c) :- If the tie for minimum occurs among the T_{i2}'s, select the job corresponding to theminimum of T_{i2}'s and process it in the last.

Step 4 :- Cross out the jobs already assigned and repeat steps 1 to 3, arranging the jobs next to first or next to last, until all the jobs have been assigned.

6.2 Illustrative examples :

6.2.1 There are seven jobs each of which has to go thorugh the machines A and B in the order AB. Processing times (in hours) are given as follows.

Job	:	1	2	3	4	5	6	7
Machine A	:	3	12	15	6	10	11	9
Machine B	:	8	10	10	6	12	1	3

Determine a sequence of these jobs that will minimize the total elapsed time T.

Solution. Clearly, Min $\{T_{i1}, T_{i2}\}$ Which is corresponding to T_{62}. Thus the job 6 to be processed in the last. so, in list of seven jobs, we place the job 6 at the last place as shown below:

						6

Now the problems preduces to the following 6 jobs and 2 machines problem.

Job	:	1	2	3	4	5	7
Machine A	:	3	12	15	6	10	9
Machine B	:	8	10	10	6	12	3

Here $\underset{1}{\text{Min}} \{T_{i1}, T_{i2}\} = 3$ which corresponds to T_{11} and T_{72} since there is a tie for T_{11} and T_{72}, the Job I will be processed first and the job 7 will be processed just before the last job. So we place them in the list as below :

1					7	6

Now the reduced problem is –

Job	:	2	3	4	5
Machine A	:	12	15	6	10
Machine B	:	10	10	6	12

The minimum of T_{i1}'s and T_{i2}'s now 6, which corresponds to T_{41} and T_{42}. Thus there is a tie for the minimum for the same job. We may choose arbitrarily to process job 4 either next to job 1 or before the job 7. Here we select it to process after the job 1, as below.

1	4				7	6

The Problem now reduces to

Job	:	2	3	5
Machine A	:	12	15	10
Machine B	:	10	10	12

Here $\min_i\{T_{i1}, T_{i2}\}$ is 10 which corresponds to T_{51}, T_{22} and T_{32}. So here also a tie occurs. Since, $T_{21} \leq T_{31}$, we place job 5 next to job 4 and job 2 before 7. The only remaining place is thus occupied by job 3 and the optimal sequence is given by :-

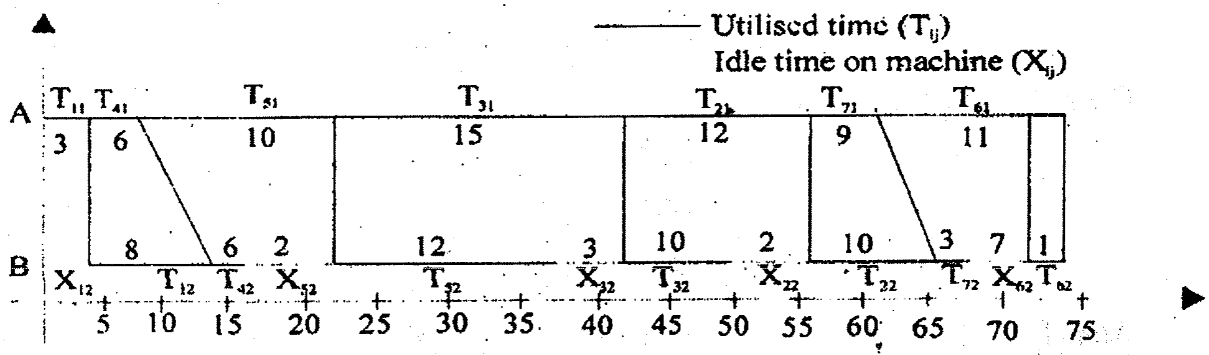
1	4	5	3	2	7	6

The minimum elapsed time can thus be computed through the following table.

	Machine	**A**	**Machine**		**B**
Job	**Time in**	**Timeout**	**Time in**	**Time out**	**Idle time of B**
1	0	3	3	11	3
4	3	9	11	17	0
5	9	19	19	31	2
3	19	34	34	44	3
2	34	46	46	56	2
7	46	55	56	59	0
6	55	66	66	67	7

Thus the minimum elasped time is 67 hours and idle time of machine b is 17 hours and for machine A is 1 hour, since the machine A completes its assigned work in 66 hours which is 1 hour less than the total elapsed time.

The Gantt Chart for this problem can be drawn as below



Time taken by the jobs (in hour)

6.2.2 A bookbinder has one printing press, one binding machine and the manuscripts of five different books. The times required to perform the printing and bidnding operations for each book are known as below.

Book	:	I	II	III	IV	V
Printing press (A)	:	3	7	4	5	7
Binding machine(B)	:	6	2	7	3	4

Determine the order in which the books should be processed, in order to minimize the total time requierd to turn out all the books. Use Gantt Chart to obtain the minimum total elapsed time.

Solution : Here, $\underset{i}{Min}\{T_{i1}, T_{i2}\} = 2$ which corresponds to T_{22}. So, the job II will be precessed in the last. Thus, in the order of precessing the books, II will be placed in the last.

				II

Now, the problem reduces to

Books	:	I	III	IV	V
(A)	:	3	4	5	7
(B)	:	6	7	3	4

Here, $\underset{i}{Min}\{T_{i1}, T_{i2}\} = 4$, which corresponds to T_{11} and T_{42}. A tie in the minima occurs here.

So, job I will be performed first and the job IV will be performed just before the last job. Thus the order of processing becomes.

I			IV	II

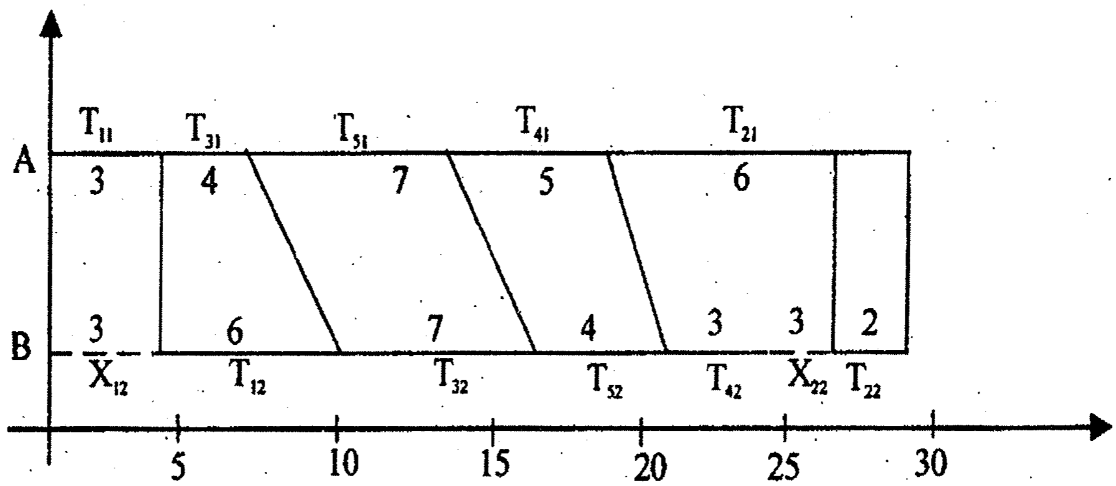
The reduced problem is -

Books :	III	V
(A) :	4	7
(B) :	7	4

$\underset{i}{Min}\{T_{i1}, T_{i2}\} = 4$ corresponds to T_{31} and T_{52}. Here also a tie is occured. So, job III will be prossed next to the first job and the job V will be precessed just before the job IV. Thus the optimal order for processing the books is given by

I	III	V	IV	II

The Gantt Chart of the given problems can be drawn as below.



From the Gantt Cahrt, it is clear that the minimum total elapsed time is 28 unit. Also the idle times of machine A and B are 2 units and 6 units respectively.

7. Processing n Jobs Through 3 Machines : Let there be n jobs, each of which is to be processed thorugh 3 machines, say A_1 and A_3 in the order A_1 A_2 A_3. Transfer of jobs is not i.e., the order of processing the jobs over each machine will be strictly followed. Also let Tij be the exact or expected precessin time of the i th job on the A_j th machine. The interative procedure for obtaining an optimal sequence of jobs is as follows :

Step 1. Find $\underset{i \in \{1,2,...,n\}}{Minimum}\ T_{i1}$, $\underset{i \in \{1,2,...,n\}}{Minimum}\ T_{i2}$ and $\underset{i \in \{1,2,...,n\}}{Minimum}\ T_{i3}$

Step 2. Check whether i) minimum $T_{i1} \geq$ maximum T_{i2}

or ii) minimum $T_{i3} \geq$ maximum T_{i2}

If either or both inequalityes in step 2. is/are satisfied, then the method given in section 6. can be extended, as given in step 3 below, to obtain an optimal sequence. But if neither of the inequalities in step2. holds, no general precedure is available, so far, to obtain the optimal sequence.

Step 3. Convert the three machine problem into two -machine problem by introducing two fictitious machines, say G and H, such that the processing time of the jobs in these two machines are $T_{iG} = T_{i1} + T_{i2}$

and $T_{iH} = T_{i2} + T_{i3}$ for i = 1,2,.......,n.

Step 4. Determine the optima sequence of n jobs for 2 machines by the method given in section 6.

7.1 Illustrative Examples:

7.1.1 Determine the optimal sequence of jobs that minimizes the total elapsed time based on the following information (Processing time on machines is given in hours and passing is not allowed.)

	Jobs							
		I	II	III	IV	V	VI	VII
Machine A_1	:	3	8	7	4	9	8	7
Machine A_2	:	4	3	2	5	1	4	3
Machine A_3	:	6	7	5	11	5	6	12

Solution . Here, $\underset{i}{Min}\ T_{i1} = 3$, $\underset{i}{Min}\ T_{i3} = 5$ and $\underset{i}{Min}\ T_{i3} = 5$

Since, $\underset{i}{Min}\ T_{i3} \geq \underset{i}{Min}\ T_{i2}$ is satisfied, the problem can be converted into a 7 jobs and 2 machines problem.

Thus if G and H are the two machines such that

$T_{iG} = T_{i1} + T_{i2}$

and $T_{iH} = T_{i2} + T_{i3}$, for i = 1,2,........,7,

The the problems can be rewritten as the following 7 jobs and 2 machines problem.

	Jobs						
	I	II	III	IV	V	VI	VII
Machine G :	7	11	9	9	10	12	10
Machine H :	10	10	70	16	6	10	15

Here $\underset{i}{Min}\{T_{iG}, T_{iH}\} = 6$ which corresponds to T_{5H}. Thus job V shall be precessed in the last. We allocate job V in the last cell as below :

						V

The problem now reduces to th4e following 6 jobs and 2 machines problem.

	Jobs					
	I	II	III	IV	VI	VII
Machine G :	7	11	9	9	12	10
Machine H :	10	10	70	16	10	15

Here $\underset{i}{Min}\{T_{iG}, T_{iH}\} = 7$ which corresponds to T_{IG} and T_{3H}. So job I will be processed at first and job 3 will be processed just before the job V.

Preceding in this way, the optimal sequence is obtained as

I	IV	VII	VI	II	III	V

To obtain the total elapsed time, we consturct the following table.

Machine A_1			Machine A_2			MachineA_3		
Job	in	out	in	out	Idle time	In	Out	Idle time
I	0	3	3	7	3	7	13	7
IV	3	7	7	12	0	13	24	0
VII	7	14	14	17	2	24	36	0
VI	14	22	22	26	5	36	42	0
II	22	30	30	33	4	42	49	0
III	30	37	37	39	4	49	54	0
V	37	46	46	47	7	54	59	0

This table indicates that the minimum total elapsed time is 59 hours. Since the machine A_1 completes its assigned work in 46 hours, idle time of machine A_1 is 13 hours. Since machine A_2 completes its assigned jobs in 47 hours and remains idle of 25 hours within that period, total idle time of machine A_2 is 25 + (59 - 47) hours = 37 hours and the idle time of machine A_3 is 7 hours.

7.1.2 There are five jobs, each of which must go throught machines A, B, C, in the order ABC. Processing times are given in the following table.

jobs	Machine A	Machine B	Machine C
1	8	5	4
2	10	6	8
3	6	2	8
4	7	3	6
5	11	4	5

Determine a sequence for five jobs that will minimize the total elapsed time T.

Solution : Here $\underset{i}{Min}\, T_{i1} = 6$, $\underset{i}{Min}\, T_{i3} = 4$ and $\underset{i}{Min}\, T_{i2} = 6$

Since $\underset{i}{Min}\, T_{i1} \geq \underset{i}{Min}\, T_{i2}$ holds, the problem can be solved by introducing two fictitous machines G and H with the processing times $T_{iG} = T_{i1} + T_{i2}$ and $T_{iH} = T_{i2} + T_{i3}$ as given below.

jobs	Machine G	Machine H
1	13	9
2	16	15
3	8	10
4	10	9
5	15	9

In this table, $Min\{T_{iG}, T_{iH}\} = 8$ which occurs for T_{3G}. Thus, the jobs 3 will be procesed at first.

The reduced problem of 4 jobs and 2 machines has the minimum entry 9 which occurs for T_{1H}, T_{4H} and T_{5H}, i.e., a tie occurs.

So either job 1, or job 4 or job 5 will be processed at last. Also, any of the other will preceed the last job.

Finally, remaining job 2 will be processed after the job 3.

So, we have some alternative solutions of this problem. The minimum elapsed time will be some for each of the alternative solutions, though the idle times for individual machines may be different in each case.

We consider the solution | 3 | 2 | 1 | 4 | 5 | and find the minimum elapsed time as follows :

	Machine	A	Machine		B	Machine		C
Job	time in	time out	time in	time out	Idle time	time In	time Out	Idle time
3	0	6	6	8	6	8	16	8
2	6	16	16	22	8	22	31	6
1	16	24	24	29	2	31	35	0
4	24	31	31	34	2	35	41	0
5	31	42	42	46	8	46	51	5

Thus the minimum total elapsed time is 51 units. Idle times of machine A, B, C, are 9 units, 31 units and 19 units respectively.

7.1.3. A readymade garment manufacturer has to process 7 terms through two stage of production, viz, cutting and searing. The time taken for each of these items at the different stages are given below in proper units.

		Items						
		1	2	3	4	5	6	7
Processng	Cutting :	5	7	3	4	6	7	12
time	Searing :	2	6	7	5	9	5	8

(i) Find an order in which these items are to be processed through these stages so as to minimize the total processing time.

(ii) Suppose as third stage of production is added, viz., pressing and packing with processing time for these items as follows:

Items	1	2	3	4	5	6	7
Times forpressing and packing	10	12	11	13	12	10	11

Find an order in which these seven items are to be processed so as to minimize the time taken to process all the items through all the three stages.

Solution : Let the cutting and searing machine be denoted by A and B respectively. Also let A_i & B_i be the processing time of the i th job on machine A and B respectively, where, i = 1,2,......7.

(i) The minimum processing time in the given table is 2 which is B_1 (i.e., processing time of item 1 on machine B). There fore we shall process the item 1 in the last and list the elements as follows-

						1

After assinging item 1, we are left with 6 intems with their processing time as follows :

	Items						
Machines		2	3	4	5	6	7
A	:	7	3	4	6	7	12
B	:	6	7	5	9	5	8

Minimum tiemin this table is 3 which is A_3 (i.e., processing time of item 3 on machine A). Therefore we shall process the item 3 at first and list the elements as follows :

3						1

Now the problem reduces to a problem of precessing 5 jobs on two machines with the processing time as given in the following table :

		Items				
Machines		2	4	5	6	7
A	:	7	5	6	7	12
B	:	6	5	9	5	8

Here the minimum processing time is 4, which is A_4 (i.e., processing time of item 4 on machine A). Therefore, the item 4 should be processed just after the first item to be processed, i.e., item 3. The items are thus listed as below :

3	4					1

Proceeding in this way, we get the optimal sequence as

3	4	5	7	2	6	1

Now the idle time of the machines can be calculated from the following table :

Machine		A	Machine			B
Items	in	out	Idletime	in	out	Idle time
3	0	3	0	3	10	3
4	3	7	0	10	15	0
5	7	13	0	15	24	0
7	13	25	0	25	33	1
2	25	32	0	33	39	0
6	32	39	0	39	44	0
1	39	44	2	44	46	0

So, the idle times of machine A & B are 2 & 4 units respectively.

(ii) Now we denote the pressing and packing machine by C. Also the processing times of the items on this machine be denoted by C_i, i = 1,2,......,7. Thus we have to process 7 items through three machines in the order A,B,C, with the processing times as follows

	Items						
Machines	1	2	3	4	5	6	7
A :	5	7	3	4	6	7	12
B :	2	6	7	5	9	5	8
C :	10	12	11	13	12	10	11

Here, $\underset{i}{Min}\ A_i = 3$

$\underset{i}{Max}\ B_i = 9$

and $\underset{i}{Min}\ C_i = 10$

∴ The condition $\underset{i}{Min}\ C_i \geq \underset{i}{Max}\ B_i$ is satisfied here. Thus we consider two fictitious machines G and H and the processing time on these two fictitious machines are given by $G_i = A_i + B_i$ and $H_i = B_i + C_i$ as follows :

	Items						
Machines	1	2	3	4	5	6	7
G :	7	13	10	9	15	12	20
H :	12	18	18	18	21	15	19

Then proceeding as (i) we get the optimal sequence as

1	4	3	6	2	5	7

We can calculate the idle times of the machines from the following table.

	Machine		A	Machine		B	Machine		C
Items	in	out	Idletime	in	out	Idle time	In	Out	Idle time
1	0	5	0	5	7	5	7	17	7
4	5	9	0	9	14	2	17	30	0
3	9	12	0	14	21	0	30	41	0
6	12	19	0	21	26	0	41	51	0

2	19	26	0	26	32	0	51	63	0
5	26	32	0	32	41	0	63	75	0
7	32	44	42	44	52	3+34 = 37	75	86	0

So the idle times of machine A, B and C are 42, 44 and 7 units respectively.

It should also be noted that the intreoductio of a new machine can change the optimal sequence of processing the jobs, as in this case.

8. Processing 2 Jobs Through m Machines :

Let us consider the situatio, when —

(i) There are m machines, denoted by $A_1, A_2, \ldots\ldots\ldots, A_m$;

(ii) There are only two jobs, i.e., job 1 and job 2;

(iii) The order of processing each of the two jobs through m machines is known in advance, but such ordering may not be same for both the jobs;

(iv) The exact or expected processin times T_{ij}'s, $i = 1,2; \; j = 1,2,\ldots\ldots,m$ are known.

The problem is to minimize the total elapsed time T.

8.1. Graphica method : In the two job m-machine problem, the graphical method is suitable to find a good sulutio. This method is quite simple to apply, but the only problem is that ithis method dies not always yeild an optimal sequence.

In the following example, graphica method is illustreted throughly.

8.1.1 Use the graphical method to minimize the time needed to process the following jobs on the machines shown below. Also find the total time needed to complete both the jobs.

Job 1	Sequence of machines:	A	B	C	D	E	F
	Time :	8	10	2	6	12	10
Job 2	Sequence of machines:	B	A	C	F	D	E
	Time :	12	6	4	8	6	10

Solutin : We draw the graph of the problem in the follwing manner :

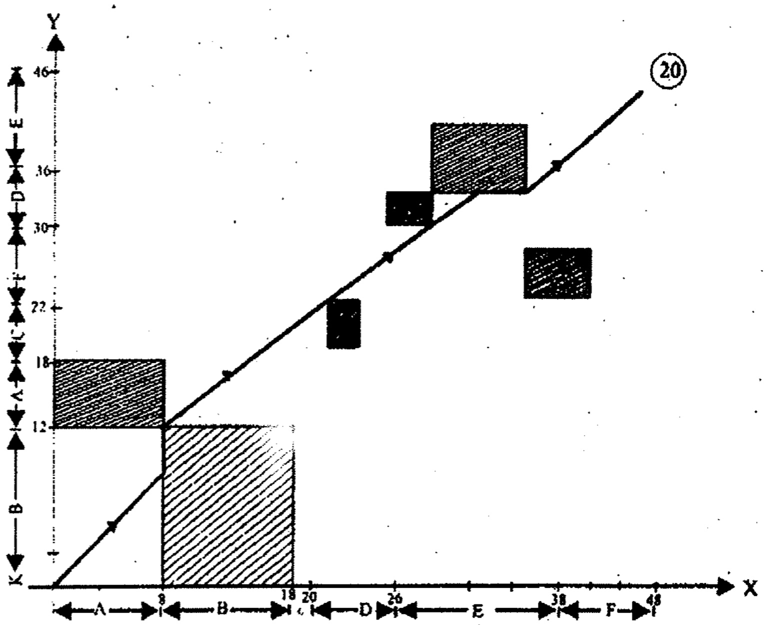
1. We draw horizontal and vertical lines OX and OY representing the processing time of

jon 1 and job 2 respectively.

2. Then we mark the processing times of the jobs on the machines in the given order as shown in the figure.

Machine A takes 8 units of time for job 1 and 6 units of time for job 2. We consturct the rectagle PQRS for the mechine A. Similarly, for other machines B, C, D, E, and F, we contruct the rectagles as shown in the figure.

3. Now starting from the starting point O, we move on doing jobs avoiding the shaded rectagular blocks unitl the finish point is reached. Here, it is important to note that we shall try to move as much as possible along a line inclined at an angle 45° with the horizontal. Whenever movement along this line is not possible we shall move only horizontally ansd vertically (as needed).



The best path is shown this figure by thick lines with arrows.

It is clear from the graph that we have processed :

Job 1 before Job 2 on machine A

Job 2 before Job 1 on machine B

Job 2 before Job 1 on machine C

Job 1 before Job 2 on machine D

Job 1 before Job 2 on machine E

Job 2 before Job 1 on machine F

Total elapsed time = Processing time of Job 1+ idle time of Job 1

= 48 Units + 4 Units

= 52 Units

= Processing time of Job 2 + Idle time of Job 2

= 46 Units + 6 Units

= 52 Units

9. Processing n Jobs Through m Machines .

Let each of the n jobs be processed through m machines, say A_1, A_2, ..., A_m in the order A_1A_2 .... A_m and T_{ij} denote the time taken by the i th machine to complete the j the job. In the following, we describe a step-by-step procedure to obtain an optimal sequence.

Step-1 : Find the values of - (i) $\underset{i}{Min}\, T_{ij}$, (ii) $\underset{i}{Min}\, T_{mj}$,

(iii) $\underset{i}{Max}\{T_{2j}, T_{3j}, ..., T_{(m-1)j}\}$ *for* $j = 1, 2, ..., n$

Step-2 : Then check whether -

(i) $\underset{j}{Min}(T_{ij}) \geq \underset{j}{Min}(T_{ij})$ *for* $i = 2, 3, ..., m-1$ *or*

(ii) $\underset{j}{Min}(T_{mj}) \geq \underset{j}{Max}(T_{ij})$ *for* $i = 2, 3, ..., m-1$ *or*

If either of these two inequalities does not hold, then this method fails to work. Other wise, go to next step.

Step - 3 : Convert the m - machine problem into 2 - machine problem by introducing two fictitions machines G and H, so that the processing time of the j th job on these machines are given by

$$T_{Gj} = \sum_{i=1}^{m-1} T_{ij} \text{ and } T_{Hj} = \sum_{i=2}^{m} T_{ij} \text{ for } j = 1, 2, ..., n$$

Step -4 : Now obtain the optional sequence of processing n jobs on these two machines by the algorithem given in section 6.1. This will also be the optimal sequence for processing n jobs through the m machines A_1, A_2, ... , A_m

Step - 5 : If in addition to the conditions given in Step - 2, it is seen that $\sum_{i=2}^{m-1} T_{ij}$

constant for all j = 1, 2, ..., n then it suffices to find the optional sequence for processing The n jobs through the only two machines A_1 and A_m in the order A_1 A_m by using the optional sequence algorithm.

It should be noted that this procedure for sequencing n jobs through m machines is not a general procedure. It is applicable to only those problems for which Step - 2 is satisfied.

9.1 Illustrative Examples.

9.1.1 There are four jobs, each of which has to go through the machines M_j, j = 1,2, ..., 6 in the order $M_1 M_2$ .... M_6. Processing times (in hours) of the jobs are given below

	M_1	M_2	M_3	M_4	M_5	M_6
Job						
A	18	8	7	2	10	25
B	17	6	9	6	8	19
C	11	5	8	5	7	15
D	20	4	3	4	8	12

Determine the sequence of these four jobs that minimizes the total elapsed time. Find also the minimum time.

Solution Here, $\underset{i}{Min}\, T_{1j} = 11, \underset{i}{Max}\, T_{2j} = 8, \underset{i}{Max}\, T_{3j} = 9$

$$\underset{j}{Max}\, T_{4j} = 6, \underset{j}{Max}\, T_{5j} = 10, \underset{j}{Min}\, T_{6j} = 12$$

The conditions $\underset{i}{Min}\, T_{1j} \geq \underset{i}{Max}\, \{T_{2j}, T_{3j}, T_{4j}, T_{5j}, T_{6j}\}$ and

$\underset{j}{Min}\, T_{6j} \geq \underset{j}{Max}\, \{T_{2j}, T_{3j}, T_{4j}, T_{5j}, T_{6j}\}$ both are satisfied. Therefore we can consider two factitions machines G and H with the processing times of the jobs on these two machines are given by—

Jobs $T_{Gj} = \sum_{i=1}^{5} T_{ij}$ $T_{Hj} = \sum_{i=2}^{6} T_{ij}$

A	45	52
B	46	48
C	36	40
D	39	31

Minimum entry in this table is 31 which is T_{H4} (i.e., the processing time of job D on machine H). Therefore, the job D will be processed at last and we place D at the last place in the sequence as follows :

			D

After assigning job 'D', we are left with 3 jobs with their processing time as follows :

Jobs	Machine G	Machine H
A	45	52
B	46	48
C	36	40

Now, minimum time in this table is 36 which is T_{3G} which is th eprocessing times of job 'C' on machine G. Therefore we shall do the job 'C' at first and make the firest eantry in the sequence as 'C' and proceeding in this way, we get the optimal sequence as C A B D , which is also the obtimal sequence of processing the job in the original problem. Minimum elapsed time and idle time of the machines are obtamed from the following table :

Job	Machine M_1			Machine M_2			Machine M_3			Machine M_4			Machine M_5			Machine M_6		
	In	Out	Idle time	In	Out	Idle time	In	Out	Idle time	In	Out	Idle time	In	Out	Idle time	In	Out	Idle time
C	0	11	0	11	16	11	16	24	16	24	29	24	29	36	29	36	57	36
A	11	29	0	29	37	13	37	44	13	44	46	15	46	56	10	56	81	5
B	29	46	0	46	52	9	52	61	8	61	67	15	67	75	11	81	100	0
D	46	66	46	66	70	14+42	70	73	9+39	73	77	6+35	77	85	2+27	100	112	0

The minimum total elapsed time is 112 hrs. and the idle times for machines M_1, M_2, ... M_6 are 46 hrs., 89 hrs., 85 hrs., 95 hrs., 79 hrs. and 41 hrs. respectively, out of which 46 hrs., 42 hrs., 39 hs., 35 hrs. and 27 hrs. will be the idle times of the machines M_1, M_2, ..., M_5 respectively, after the completion of the last job.

9.1.2 Solve the following sequencing problem giving an optimal solution when passing is not allowed.

Jobs / Machines	A	B	C	D	E
M_1	11	13	9	16	17
M_2	4	3	5	2	6
M_3	6	7	5	8	4
M_4	15	8	13	9	11

Solution : In this example, $\underset{j}{Min}\, T_{1j} = 9$, $\underset{j}{Min}\, T_{4j} = 8$, $\underset{j}{Min}\, \{T_{2j}, T_{3j}\} = 8$,

The conditions $\underset{i}{Min}\, T_{1j} \geq \underset{i}{Max}\, \{T_{2j}, T_{3j}\}$ *and*

$\underset{i}{Min}\, T_{4j} \geq \underset{i}{Max}\, \{T_{2j}, T_{3j}\}$ *both are satisfied*

Also, $T_{2j} + T_{3j} = 10$ for all j = 1,2,3,4,5.

So the given problem can be reduced to an optimal sequence problem involving five jobs and two machines M_1 and M_4 in the order M_1 M_4, which means that the machines M_2 and M_3 have no effects on the optimality of the sequence.

Using the optimal sequence algorithm given in section 6.1, we obtain the optimal Sequence : C $\rightarrow$ A $\rightarrow$ E $\rightarrow$ D $\rightarrow$ B

Now the total elapsed time can be calculated as follows :

Job	Machines			
	M_1	M_2	M_3	M_4
C	0 – 9	9 – 14	14 – 19	19 – 32
A	9 – 20	20 – 24	24 – 30	32 – 47
E	20 – 37	37 – 43	43 – 47	47 – 58
D	37 – 53	53 – 55	55 – 63	63 – 72
B	53 – 66	66 – 69	69 – 76	76 - 84

This shows that the total elapsed time is 84 Units.

10. Summary :

In this module, we have discussed the problem of sequencing a series of jobs on a finite member of machines. An algorithm to find and an optimal sequence of jobs on two machines is developed initially. Then with the help of this algorithm, optimal sequincing algorithm for processing n jobs through three machines, and n jobs through M machines are developed. The graphical method to obtain an order of two jobs which are to be processed on different machines. Illustrative examples are provided to elaborate each of the methods.

11. Exercise.

(i) Find the sequence that minimizes the total elapsed time required to complete the following tasks:

Tasks	:	A	B	C	D	E	F	G	H	I
Time of machine - I	:	2	5	4	9	6	8	7	5	4
Time of machine - II	:	6	8	7	4	3	9	3	8	11

(ii) There are sir jobs each of which must go first over machine 1 and then over machine 2. The order of completion of the jobs has no significance. The following table gives the machine times in hours for six jobs and two machines :

Job No.	:	1	2	3	4	5	6
Machine - I	:	5	9	4	7	8	6
Machine - II	:	7	4	8	3	9	5

Find the sequence of the jobs that minimizes the total elapsed time to complete the jobs. Find the minimum total elapsed time by using Gantt Chart.

(iii) We have find jobs each of which must go through two machines A and B in the order AB. Processing times are given in the following table:

Job No.	:	1	2	3	4	5
Machine - A	:	10	2	18	6	20
Machine - B	:	4	12	14	16	8

Determine a sequence for the five jobs that will minimize the total elapsed time find also the idle times of the machines.

(iv) Find the sequence that minimizes the total elapsed time to complete the following jobs :

Processing times in hours							
No. of Job	:	1	2	3	4	5	6
Machine-A	:	4	8	3	6	7	5
Machine-B	:	6	3	7	2	8	4

(v) Find the sequence that minimizes the total elapsed time (in hours) required to complete the folloing jobs on two machines M_1 and M_2 in the order M_1M_2.

a)

Job	:	A	B	C	D	E
M_1	:	5	1	9	3	10
M_2	:	2	6	7	8	4

b)

Job	:	A	B	C	D	E	F
M_1	:	30	120	50	20	90	110
M_2	:	80	100	90	60	30	10

(vi) Find the sequence that minimized the total elapsed time (in hours) to complete the following jobs on three machines M1, M2 and M3 in the order M1 M2 M3.

a)

		Job				
M1	:	A	B	C	D	E
M_1	:	3	8	7	5	2
M_2	:	3	4	2	1	5
M_3	:	5	8	10	7	6

b)

		Job				
Machines	:	A	B	C	D	E
M_1	:	5	7	6	9	5
M_2	:	2	1	4	5	3
M_3	:	3	7	5	6	7

(vii) Find the sequence that minimizes the total elapsed time required to complete the following tasks :

Tasks	A	B	C	D	E	F	G
Time on machine - I	3	8	7	4	9	8	7
Time on machine - II	4	3	2	5	1	4	3
Time on machine - III	6	7	5	11	5	6	12

(viii) Find the sequence that minimizes the total elapsed time required to complete the following tasks. Each job is processed in the order ACB.

	Job						
Machine	1	2	3	4	5	6	7
Machine - A	12	6	5	11	5	7	6
Machine - B	7	8	9	4	7	8	3
Machine -C	3	4	1	5	2	3	4

(ix) Use graphical method to minimize the time needed to process the following jobs on the machines shown below i.e., for each machine find the job which should be done first. Also calculate the total time needed to complete both the jobs.

			Machines				
Job - 1	Sequence	:	A	B	C	D	E
	Time	:	3	4	2	6	2
Job - 2	Sequence	:	B	C	A	D	E
	Time	:	5	4	3	2	6

(x) A machine shop has four machines A,B,C,D. Two jobs must be processed through each of these machines. The time (in hour) taken on each of the machines and the necessary sequence of jobs through the shop are given below :

			Machines			
Job - 1	Sequence	:	A	B	C	D
	Time	:	2	4	5	1
Job - 2	Sequence	:	B	C	A	D
	Time	:	6	4	2	3

(xi) Two jobs are to be processed on four machines a, b, c and d. The technological order for these jobs on machines as follows:

Job 1 :- a → b → c → d

Job 2 :- d → b → a → c

Processing times are given in the following table :

	a	b	c	d
Job 1	4	6	7	3
Job 2	4	7	5	8

Find the optional sequence of jobs on each of the machines, by Graphical method.

(xii) The following table gives the processing times of 10 items on 2 machines A and B. Each item has to be processed first on machine A and then on machine B. Use Jobnson's algorithm (optimal sequencing algorithm) to prepare a processing schedule which takes the beast processing time. Find this time also :

Item	:	1	2	3	4	5	6	7	8	9	10
Machine A (Processing time)	:	3	2	13	10	5	6	2	15	10	7
Machine B (Processing time)	:	5	8	5	12	11	10	13	7	5	12

(xiii) Write short notes on –

a) Waiting time of a job

b) Completion time of a job

c) No passing rule.

12. References

I. Sequencing Theory – S. Ashour (Springer-Verlay, Berlin, 1972)

II. Scheduling Algorithms – Peter Brucker (Springer, Berlin, 2001)

III. Handbook of Scheduling – Joseph Y-T. Leung (Chapman - Hall / CRC, 2004)

IV. Operations Research – S.D. Sharma

V. Operations Research –Hamdy A. Taha (Macmilar, 1982)

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X Special Paper : Operations Research

Module No. - 117

Advanced Optimization and Operations Research-II

(Replacement and Maintenance Models)

....

Module Structure :

117.1. Introduction

In this module, we shall discuss the problem of replacement. The problem of replacement is felt when the job performing units such as men, machines, equipments, etc. become less effective or useless due to either sudden or gradual deterioration in their efficiency, failure or breakdown. However such replacements would increase the need of capital cost for new ones.

In this chapter we shall discuss three types of replacement situations :

(i) Items which do not give any indication of deterioration with time but fail all of a sudden and become completely useless such as light bulbs and tubes, radio and television parts etc.

(ii) Items whose efficiency deteriorates with age due to constant use such as vehicles, tyres, machines etc.

(iii) The existing working staff in an organisation gradually reduces due to retirement, death and other reasons.

117.2. Objectives

Go through this module you will learn the following:

* Different types of failure
* Replacement policies
* Mortality theorem
* Individual replacement
* Group replacement
* Staffing problem
* Equipment renewal problem

117.3. Key-words

Replacement, individual policy, group replacement policy, other replacement policy.

117.4.1 Failure Mechanism of Items

There are two types of failure.

(i) Gradual failure and

(ii) Sudden failure.

117.4.1 Gradual failure

Gradual failure is progressive in nature. As the life of an item increases, its operational efficiency decreases resulting in

(i) increased maintenance and operating costs.

(ii) decrease in its productivity.

(iii) decrease in the resale or salvage value.

117.4.2 Sudden failure

As the name suggest this type of failure occurs suddenly. The period of giving desired service is not constant. It may follow some frequency distributión which may be progressive, retrogressive or random in nature.

(i) Progressive failure : If the probability of failure of an item increases with the increase in its life then such failure is called progressive failure. For example, light bulbs and tubes fail progressively.

(ii) Retrogressive failure : If the probability of failure in the beginning of the life of an item is more but as time passes the chances of its failure become less, then such failure is called retrogressive failure.

(iii) Random failure : In this type of failure, the constant probability of failure is associated with items that fail from random causes such as physical shocks, not related to age.

117.5 Replacement of Items Deteriorates with Time

When operational efficiency of an item deteriorates with time, it is economical to replace the same with a new one. In this section, we shall discuss various techniques for making comparisons under different conditions.

117.5.1 Replacement policy for items when money value remains constant.

The cost of maintenance of a machine is given as a function increasing with time and its scrap value is constant.

(a) If time is measured continuously, then the average annual cost will be minimized by replacing the machine when the average cost to data becomes equal to the current maintenance cost.

(b) If time is measured in discrete units, then the average annual cost will be minimized by replacing the machine when the next periods maintenance cost becomes greater than the current average cost.

Proof. The aim here is to determine the optimal replacement age of an equipment whose running cost increases with time and the value of money remains constant during that period. Let,

C = Capital (or purchase) cost of new equipment.

S = Scrap (or salvage) value of the equipment at the end of t years.

$R(t)$ = running cost of equipment for the year t

n = replacement age of the equipment.

(a) When time 't' is a constant variable :

If the equipment is used to t years, then the total cost incurred over this period is given by

TC = Capital (or purchase) cost – Scrap value at the end of t years + Running cost for t years

$$= C - S + \int_0^n R(t)\,dt$$

Therefore the average cost per unit time incurred over the period of n years is :

$$ATC_n = \frac{1}{n}\left\{C - S + \int_0^n R(t)\,dt\right\}$$

To obtain optimal value of n for which ATC_n is minimum, differentiate ATC_n with respect to n and set the first derivative equal to zero. That is, for minimum of ATC_n

$$\frac{d}{dn}\{ATC_n\} = -\frac{1}{n^2}\{C - S\} + \frac{R(n)}{n} - \frac{1}{n^2}\int_0^n R(t)\,dt = 0$$

or, $$R(n) = \frac{1}{n}\left\{C - S + \int_0^n R(t)\,dt\right\}, n \neq 0$$

or, $R(n) = ATC_n$

Hence the following replacement policy can be derived with the help of above equation.

(b) When time '*t*' is a discrete variable :

The average cost incurred over the period *n* is given by

$$ATC_n = \frac{1}{n}\left\{C - S + \sum_{t=0}^{n} R(t)\right\} \qquad \text{................ (1)}$$

If $C-S$ and $\sum_{t=0}^{n} R(t)$ are assumed to be monotonically decreasing and increasing respectively, then there will exist a value of *n* for existing value of *n* for which ATC_n is minimum. Thus we shall have inequalities

$$ATC_{n-1} > ATC_n < ATC_{n+1}$$

or, $ATC_{n-1} - ATC_n > 0$

and $ATC_{n+1} - ATC_n > 0$

Rewriting equation (1) for period $n + 1$, we get

$$ATC_{n+1} = \frac{1}{n+1}\left\{C - S + \sum_{t=0}^{n+1} R(t)\right\}$$

$$= \frac{1}{n+1}\left\{C - S + \sum_{t=0}^{n} R(t) + R(n+1)\right\}$$

$$= \frac{n}{n+1}\frac{\left\{C - S + \sum_{t=1}^{n} R(t)\right\}}{n} + \frac{R(n+1)}{n+1}$$

$$= \frac{n}{n+1} ATC_n + \frac{R(n+1)}{n+1}.$$

Therefore, $ATC_{n+1} - ATC_n = \frac{n}{n+1} ATC_n + \frac{A(n+1)}{n+1} - ATC_n$

$$= \frac{R(n+1)}{n+1} + ATC_n\left(\frac{n}{n+1} - 1\right)$$

$$= \frac{R(n+1)}{n+1} - \frac{ATC_n}{n+1}$$

Since $ATC_{n+1} - ATC_n > 0$, we get

$$\frac{R(n+1)}{n+1} - \frac{ATC_n}{n+1} > 0$$

or, $R(n+1) - ATC_n > 0$

or, $R(n+1) > ATC_n$

Similarly, $ATC_{n-1} - ATC_n > 0$, implies that $R(n+1) < ATC_{n-1}$. This proves the following replacement policy.

Example 117.1 : A firm is considering replacement of a machine, whose cost price is Rs. 12,200, and the scrap value, Rs. 200. The running costs are found from experience to be as follows:

Year :	1	2	3	4	5	6	7	8
Running cost (Rs.)	200	500	800	1200	1800	2500	3200	4000

when should the machine be replaced?

Solution : We are given the running cost, $R(n)$, the scrap value S = Rs. 200 and the cost of the machine C = 12,200. In order to determine the optimal time n when the machine should be replaced, first we calculate an average cost per year during the life of the machine as shown in the table.

Year of Service n	Running Cost (Rs.) $R(n)$	Cumulative running Cost (Rs.) $\sum R(n)$	Depreciation Cost (Rs.) $C-S$	Total Cost (Rs.) TC	Average Cost (Rs.) ATC_n
(1)	(2)	(3)	(4)	(5)=(3)+(4)	(6)=(5)/(1)
1	200	200	12,000	12,200	12,000
2	500	700	12,000	12,500	6,350
3	800	1,500	12,000	13,500	4,500
4	1,200	2,700	12,000	14,700	3,675

Year of Service n	Running cost (Rs.) $R(n)$	Cumulative running Cost $\sum R(n)$ (Rs.)	Depreciation Cost (Rs.) $C-S$	Total Cost Rs. TC	Average Cost (Rs.) ATC_n
(1)	(2)	(3)	(4)	(5)=(3)+(4)	(6)=(5)/(1)
5	1,800	4,500	12,000	16,500	3,300
6	2,500	7,000	12,000	19,000	3,167
7	3,200	10,200	12,000	22,200	3,171
8	4,000	14,200	12,000	26,200	3,275

From the table, it may be noted that the average cost per year, ATC_n is minimum in the sixth year (Rs. 3,167). Also the average cost in seventh year (Rs. 3,171) is more than cost in the sixty year. Hence the machine should be replaced after six years.

Example 117.2 The data collected in running a machine, the cost of which is Rs. 60,000 are given below:

Year	1	2	3	4	5
Resale value (Rs.)	42,000	30,000	20,400	14,400	9,650
Cost of spares (Rs.)	4,000	4,270	4,880	5,700	6,800
Cost of labour (Rs.)	14,000	16,000	18,000	21,000	25,000

Determine the optimum period for replacement of the machine.

Solution. The costs of spares and labour together determine running cost. Thus the running costs and the resale price of the machine in successive years are as follows:

Year	1	2	3	4	5
Resale Value (Rs.)	42,000	30,000	20,400	14,400	9,650
Running Cost (Rs.)	18,000	20,270	22,880	26,700	31,800

The calculations of average running cost per year during the life of the machine are shown in the following table.

Year of Service n	Running Cost (Rs.) $R(n)$	Cumulative running Cost (Rs.) $\sum R(n)$	Resale value (Rs.) S	Depreciation Cost (Rs.) $C-S$	Total Cost (Rs.) TC	Average Cost (Rs.) ATC_n
(1)	(2)	(3)	(4)	(5)=60,000–(4)	(6)=(3)+(5)	(7)=(6)/(1)
1	18,000	18,000	42,000	18,000	36,000	36,000.00
2	20,270	38,270	30,000	30,000	68,270	34,135.00
3	22,880	61,150	20,400	39,600	1,00,750	33,583.00
4	26,700	87,850	14,400	45,600	1,33,450	33,362.50
5	31,800	119,650	9,650	50,350	1,70,000	34,000.00

The calculations in the table show that the average cost is lowest during the fourth year. Hence the machine should be replaced after every fourth year. Otherwise the average cost per year for running the machine would start increasing.

117.5.2. Replacement policy for equipments when value of money changes with constant rate during the period

Value of money criteria

If the effect of the time value of the money is to be considered, then replecement decision analysis must be based upon an equivalent annual cost, for example, if the interest rate on Rs. 100 is 10% per year, then value of Rs. 100 to be spent after one year will be Rs. 110. This is also called value of money. Also, the value of money that decreases with constant rate is known as its depriciation ratio or discounted factor. The discounted value is the amount of money required at the time of the policy decision to build up funds at compound interest large enough to pay the required cost when due. For example, if the interest rate on Rs. 100 is r percent per year the present value (or worth) of Rs. 100 to be spent after n years will be

$$d = \left(\frac{100}{100+r}\right)^n$$

where d is called the discount rate or depreciation value. After having the idea of discounted cost of objective should be to determine the critical age at which an item should be replaced so that the sum of all discount costs is minimum.

Example 117.3. A pipeline is due for repairs. It will cost Rs. 10,000 and lasts for three years. Alternately, a new pipeline can be laid down at a cost of Rs. 30,000 and lasts for 10 years. Assuming cost of capital to be 10% and ignoring salvage value, which alternative should be chosen?

Solution : Consider the two types of pipelines for infinite replacement cycles of ten years for the new pipeline, and three years for the existing pipeline.

Since the discount rate of money per year is 18%, three the present worth of the money to be spent over a period of one year is

$$d = \frac{100}{100+10} = \frac{10}{11} = 0.9091.$$

Let D_n denote the discounted value of all future costs associated with a policy of replacing the equipment after n years. Then we shall have

$$D_n = c + c \times d^n + c \times d^{2n} + \ldots$$

$$= c\left(1 + d^n + d^{2n} + \ldots\right) = \frac{c}{1-d^n}$$

where c is the initial cost.

Now, substituting the values of c, d's and n for two types of pipelines, we get

$$D_3 = \frac{10,000}{1-(0.9091)^3} = Rs. 4021, \text{ for existing pipeline.}$$

and $$D_{10} = \frac{30,000}{1-(0.9091)^{10}} = Rs. 48,820, \text{ for new pipeline.}$$

Since the value of $D_3 < D_{10,}$ the existing pipeline should be continued.

Alternatively, the comparison may be made over 3x10=30 years.

Present worth factor criteria

In this case the optimal value of replacement age of an equipment can be determined under two different situations:

(i) The running cost of an equipment which deteriorates over a period of time increases monotonically and the value of the money decreases with a constant rate. If r is the interest rate then

$$Pwf = (1+r)^{-n}$$

is called the present worth factor (*Pwf*) or present value of one rupee spent in n years from time now onwards. But, if $n = 1$ the *Pwf* is given by

$$d = (1 + r),^{-1}$$

where d is called the discount rate or depreciation value.

(ii) The money to be spent is taken on loan for a certain period at a given rate under the condition of repayment in instalments.

The replacement of items on the basis of present worth factor (*Pwf*) includes the present worth of all future expenditure and revenues for each replacement alternatives. An item for which present worth factor is less is preferred. Let

C = purchases cost of an item

R = annual running cost

n = life of the item in years

S = scrap (or salvage) value of the item at the end of its life.

r = annual interest rate.

Then the present worth of the total cost during n years is given by

Total cost = $C + R$ (*Pwf* for r% interest rate for n years)

$- S$ (*Pwf* for r% interest rate for n years).

If the running cost of the item is different in its different operational life, then the present worth of the total cost during n years given by

Total cost = $C + R$ (*Pwf* for $r\%$ interest rate for i year)

$-S$ (*Pwf* for $r\%$ interest rate for i years)

where $i = 1, 2, ..., n$.

Example 117.4. A person is considering to purchase a machine for his our factory. Relevant data about alternative machine are as follows:

	Machine A	**Machine B**	**Machine C**
Present investement (Rs.)	10,000	12,000	15,000
Total annual cost (Rs.)	2,000	1,500	1,200
Life (years)	10	10	10
Salvage value (Rs.)	500	1,000	1,200

As an advisor to the byer, you have been asked to select the best machine, considering 12% normal rate of return.

You are given that

(i) Single payment present worth factor (*Pwf*) at 12% interest for 10 years = 0.322

(ii) Annual series present worth factor (*Pwf*) at 12% interest for 10 years = 5.650

Solution : The present value of total cost of each of the three machines for a period of ten years is given in the following table.

Machine	**Present investment**	**Present value of total annual cost**	**Present value of salvage value**	**Net Cost (Rs.)**
(1)	(2)	(3)	(4)	(5)= (2)+(3)–(4)
A	10,000	2000×5.65=11,300	500×0.322=161.00	21,139.00
B	12,000	1500×5.65=8,475	1000×0.322=322.00	**20,153.00**
C	15,000	1200×5.65=6,780	1200×0.322=386.00	21,393.60

Table shows that the present value of total cost for machine *B* is the least and hence machine *B* should be purchased.

117.6. Replacement of Items that Fail Completely

It is a very difficult task to predict that a particular equipment will fail at a particular time. This uncertainty can be avoided by deriving the probability distribution of failures. Here it is assumed that the failures occur only at the end of the period, say t. Thus, the objective becomes to find the value of t which minimizes the total cost involved for the replacement.

Mortality Tables

These tables are used to derive the probability distribution of life span of an equipment in question. Let,

$M(t)$ = number of survivíors at any time t

$M(t-1)$ = number of survivors at any time $t-1$

N = initial number of equipments.

Then the probability of failure during time period t is given by

$$P(t)=\frac{M(t-1)-M(t)}{N}.$$

The probability of survival to an age t is given by

$$P_S(t)=\frac{M(t)}{N}.$$

Mortality theorem. A Large population is subject to a given mortality law for a very long period of time. All deaths are immediately replaced by births and there are no other entries or exits. Show that the age distribution ultimately becomes stable and that the number of deaths per unit time becomes constant and is equal to the size of the total population divided by the mean age of death.

Proof. Without any loss of generality, it is assumed that death (or failure) occurs just before the age of $(k+1)$ years, where k is an integer. That is, life span of an item lies between $t=0$ and $t=k$. Let us define

$f(t)$ = number of births (replacements) at time t, and

$p(x)$ = probability of death (failure) just before the age $(x+1)$, i.e., failure at time x, and $\sum_{x=0}^{k} p(x) = 1.$

If $f(t-x)$ represents the number of births occur at time $t-x$; $t = k, k+1, k+2, \ldots$, then the age of newly borns attain age x at time t as shown in the following figure.

Hence the expected number of deaths of such newly borns before reaching the time $t+1$ (i.e., at time t) will be expected number of death $= \sum_{x=0}^{k} p(x) f(t-x),$

$t = k, k+1, \ldots.$

Since all deaths (failure) at time t are replaced immediately by births (replacements) at time $t+1$, expected number of births are :

$$f(x+1) = \sum_{x=0}^{k} p(x) f(t-x), t = k, k+1, \ldots \qquad \text{.................. (1)}$$

The solution to the equation (1) in t can be obtained by putting the value $f(x) = A\alpha^t$, where A is some constant. The equation (1), becomes

$$A\alpha^{t+1} = A\sum_{x=0}^{k} p(x)\alpha^{t-x}. \qquad \text{.................. (2)}$$

Dividing both sides of (2) by $A\alpha^{t-k}$, we get

$$\alpha^{k+1} = \sum_{x=0}^{k} p(x)\alpha^{k-x} = \alpha^k \sum_{x=0}^{k} p(x)\alpha^{-x}$$

$$= \alpha^k \left[p(0) + p(1)\alpha^{-1} + p(2)\alpha^{-2} + \ldots\right]$$

or, $$\alpha^{k+1} - \left[p(0)\alpha^k + p(1)\alpha^{k-1} + \ldots + p(k)\right] = 0. \qquad \text{.................. (3)}$$

Equation (3) is of degree $(k+1)$ and will therefore have exactly $(k+1)$ roots. Let us denote the roots of equation (3) by $\alpha_0, \alpha_1, \ldots, \alpha_k$.

For the $\alpha = 1$, *lhs* of (3) is

$$1-\{p(0)+p(1)+...+p(w)\}=1-\sum_{x=0}^{w}p(x)=0=rhs.$$

Hence one root of (3) is $\alpha = 1$. Let us denote this root by α_0. The general solution of (3) will then be of the form

$$\begin{aligned} f(t) &= A_0\alpha_0^t + A_1\alpha_1^t + ... + A_k\alpha_k^t \\ &= A_0 + A_1\alpha_1^t + A_2\alpha_2^t + ... + A_k\alpha_k^t \end{aligned} \qquad ....(3)$$

where $A_0, A_1, A_2, ..., A_k$ are constants whose values are to be calculated.

Since one of the root of (3), $\alpha_0 = 1$ is positve, then by Descarte's rule of sign all other roots will be negative and their absolute value is less than unity, i.e. $|\alpha_i| < 1, i = 1, 2, ...k$. It follows that the value of these roots tend to zero as $t \to \infty$. With the result, (4) becomes $f(t) = A_0$. This indicates that the number of deaths (as well as births) becomes constant at any time.

Now, the problem is to determine the value of the constant A_0. For this we can proceed as follows. Let us define

$g(x)$ = Probability of survivor for more than x years.

= 1 – prob. (survior will die before attaining the age x)

= $1 - \{p(0) + p(1) + ... + p(x-1)\}$.

Obviously, it can be assumed that $g(0) = 1$.

Since the number of births as well as deaths have become constant equal to A_0, expected number of survivors of age x is given by $A_0 \cdot g(x)$.

As deaths are immediately replaced by births and therefore size N of the populations remains constant. That is,

$$N = A_0\sum_{x=0}^{k}g(x) \text{ or, } A_0 = \frac{N}{\sum_{x=0}^{k}g(x)} \qquad ....(5)$$

The denominator in (5) represents the average age at death. Hence

$$A_0 = \frac{N}{\text{Average age at death}}$$

117.6.1. Individual replacement policy

Under this policy, an item (or equipment) is replaced just after its failure in the given system. This ensures smooth running of the system.

117.6.2. Group replacement policy

Sometimes just after the complete break down of a system, the immediate replacement of the item(s) may not be available. This may result in heavy losses. In such circumstances a group replacement policy can be adopted. Under this policy items are replaced (*i*) individually as and when they fail during a specified time period, (*ii*) in groups at the end of some suitable time period without waiting for their failure, with the provision that if any items then fails before the time specified, it may be replaced individually. Here, we have to see two things namely :

(*i*) rate of individual replacement during the specified time period.

(*ii*) total cost incurred for individual as well as group replacement during the specified time.

Obviously, the decision-maker will call a time period optimal for which total cost incurred is minimum. In order to calculate this optimal time period for replacement he/she has to keep the record of (*i*) probability of failure, (*ii*) loss incurred due to these failures, (*iii*) cost of individual replacement, and (*iv*) cost of group replacement.

Note : The group replacement policy is suitable for a large number of identical low cost items which are likely to fail with age and it is difficult as well as not justified to keep the record of their individual ages.

The rate of replacement and total cost involved in the group replacement and total cost involved in the group replacement is based on the following theorem.

Theorem 117.1 (Group replacement policy)

(a) Group replacement should be made at the end of the period t, if the cost of individual replacements for the period t is greater than the average cost per period through the end of period t.

(b) Group replacement is not advisable at the end of period t if the cost of individual replacement at the end of period $(t-1)$ is less than the average cost per period through the end of period t.

Proof. Let us consider the following notations :

n = total number of items in the system.

$F(t)$ = number of items failing during time t

$C(t)$ = total cost of group replacement until the end of period t

C_1 = unit cost of replacement in a group

C_2 = unit cost of individual replacement after time t, failure

L = maximum life of any item

$p(t)$ = probability of failure of any item at age t.

Rate of Replacement of Time t : The number of failures at any time t is

$$F(t)=\begin{cases} np(t)+\sum_{x=1}^{t-1} p(x)F(t-x), t\le L \\ \sum_{x=1}^{L} p(x)F(t-x), \quad t>L \end{cases} \qquad \text{..................(1)}$$

Cost of Replacement of time t : The cost of group replacement after time period t is given by

$$c(t)=nC_1+C_2\sum_{x=1}^{t-1}F(x). \qquad \text{..................(2)}$$

In equation (2) nC_1 is the cost of replacing the items as a group and $C_2\sum_{x=1}^{t-1}F(x)$, is the cost of replacing the individual failures at the end of each of $(t-1)$ periods before group is again replaced.

The average cost per unit period is then given by

$$\frac{C(t)}{t}=\frac{nC_1}{t}+\frac{C_2}{t}\sum_{x=1}^{t-1}F(x). \qquad \text{.................. (3)}$$

For optimal replacement period t, the value of average cost per unit period by (3) should be minimum. The condition for minimum of $C(t)/t$ is

$$\Delta\left\{\frac{C(t)}{t-1}\right\}<0<\Delta\left\{\frac{C(t)}{t}\right\}.$$

Now, for $\Delta\left\{\frac{C(t)}{t}>0\right\}$, we have

$$\Delta\left\{\frac{C(t)}{t}\right\}=\frac{C(t+1)}{t+1}-\frac{C(t)}{t}>0.$$

From (3), we get

$$\frac{C(t+1)}{t+1}-\frac{C(t)}{t}=nC_1\left(\frac{1}{t+1}-\frac{1}{t}\right)+C_2\sum_{x=1}^{t-1}F(x)\left[\frac{1}{t+1}-\frac{1}{t}\right]+\frac{C_2F(t)}{t+1}$$

$$=\left\{-nC_1-C_2\sum_{x=1}^{t-1}F(x)+tC_2F(t)\right\}\Big/\left[t(t+1)\right].$$

For $\frac{C(t+1)}{t+1}-\frac{C(t)}{t}>0,$ it is necessary that

$$t.C_2\,F(t)>nC_1+C_2\sum_{x=1}^{t-1}F(x)$$

or, $$C_2\,F(t)>\left[nC_1+C_2\sum_{x=1}^{t-1}F(x)\right]\Big/t. \qquad \text{.................(4)}$$

Similarly, for $\Delta\left\{\frac{C(t)}{t}\right\}=\frac{C(t-1)}{t-1}-\frac{C(t)}{t}>0,$ the following condition can be derived

$$C_2F(t-1)<\left\{nC_1+C_2\sum_{x=1}^{t-2}F(x)\right\}\Big/(t-1). \qquad \text{.............. (5)}$$

Inequalities (4) and (5) describe the necessary condtion for optimal group replacement. In (4), the expression

$$\left[nC_1+C_2\sum_{x=1}^{t-1}F(x)\right]\Big/t,$$

represent the average cost per period if all items are replaced at the end of period t. Where as expression C_2+$F(t)$ represents the cost for the tth period if group replacement is not made at the end of period t.

Example 117.5.

(a) At time zero, all items in a system are new. Each item has a probability p of failing immediately before the end of the first month of life, and a probability $q=1-q$ of failing immediately before the end of the second month (i.e., all items fail by the end of the second month). If all items are replaced as they fail, then show that the expected number of failures $f(x)$ at the end of month x is given by

$$f(x)=\frac{N}{1+q}\left[-(-q)^{x+1}\right]$$

where N is the number of items in the system.

(b) If the cost per item of individual replacement c_1, and the cost per item of group replacement c_2, find the conditions under which (*i*) a group replacement policy at the end of each month is the most profitable; (ii) no group replacement policy is better that of a pure individual replacement.

Solution : (a) Let N_i be the expected number of items to fail at the end of the *i*th month. Then

N_0 = number of items in the system in the beginning (=N)

N_1 = expected number of failures at the end of the first month.

$= N_0 p = N(1-q)$, since $p = 1-q$.

N_2 = expected number of failures at the end of the second month

$$= N_0 q + N_1 p = Nq + N_1(1-q)$$

$$= Nq + N(1-q)^2 = N(1-q+q^2).$$

N_3 = expected number of failures at the end of the third month

$$= N_0 q + N_1 q + N_2 p = N(1-q)q + N(1-q+q^2)(1-q)$$

$$= N(1-q+q^2-q^3)$$

and so on. In general

$$N_k = N\{1-q+q^2-q^3+...+(-q)^k\}.$$

Therefore,

$$N_{k+1} = N_{k+1}q + N_k p$$

$$= n\{1-q+q^2+...+(-q)^{k-1}\}q + N\{1-q+q^2+...+(-q)^k\}\times(1-q)$$

$$= N\{1-q+q^2+...+(-q)^{k+1}\}.$$

By mathematical induction, the expected number of items to fail, $f(x)$, at the end of month x is then given by

$$f(x) = N\{1-q+q^2+...+(-q)^x\}$$

$$= \frac{N\{1-(-q)^{x+1}\}}{1+q}.$$

(b) The value of $f(x)$ at the end of month x will vary for different values of $(-q)^{x+1}$ and it will reach in steady state as $x \to \infty$. Hence in the steady state the expected number of failures becomes

$$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \frac{N}{1+q}\left\{1-(-q)^{x+1}\right\}$$

$$= \frac{N}{1+q}, q<1 \text{ and } (-q)^{x+1} \to 0 \text{ as } x \to \infty$$

where $(1+q)$ represents the mean age at failure and is given by $p+2q=(1-q)+2q=1+q$.

Since c_2 is the cost of replacement per item individually, the average cost per month for an individual replacement policy will be

$$\frac{N}{1+q}\cdot c_2.$$

(i) The average cost for group replacement policy at the end of every month is given by

$Nc_1 + Npc_2 = Nc_1 + N(1-q)c_2.$

A group replacement policy at the end of each month is the most profitable, when

$$Nc_1 + Npc_2 < \frac{Nc_2}{1+q}, \text{ or, } c_1 < \frac{q^2}{1+q}c_2$$

or, $$c_2 > \frac{1+q}{q^2}c_1.$$

Then individual replacement policy is always better than any group replacement policy.

Example 117.6. A computer contains 10,000 resistors. When any resistor fails, it is replaced. The cost of replacing a resistor individual is Rs. 1 only. If all the resistors are replaced at the same time, the cost per resistor would be reduced to 35 paise. The percentage of surviving resistors say $s(t)$ at the end of month t and $p(t)$ the probability of failure during the month t are :

t	:	0	1	2	3	4	5	6
$s(t)$	:	100	97	90	70	30	15	0
$p(t)$	:	–	0.03	0.07	0.20	0.40	0.15	0.15

what is the optimal replacement plan?

Solution : Let N_i be the number of resistors replaced at the end of the ith month. Then different values of N_i can be calculated as follows:

N_0 = number of resistors in the beginning = 10,000

N_1 = number of resistors being replaced by the end of first month

$= N_0 p_1 = 10,000 \times 0.03 = 300$

N_2 = number of resistors being replaced by the end of second month

$= N_0 p_2 + N_1 p_1 = 10,000 \times 0.07 + 300 \times 0.03 = 709.$

$N_3 = N_0 p_3 + N_1 p_2 + N_2 p_1$

$= 10,000 \times 0.20 + 300 \times 0.07 + 709 \times 0.03 = 2042$

$N_4 = N_0 p_4 + N_1 p_3 + N_2 p_2 + N_3 p_1$

$= 10,000 \times 0.40 + 300 \times 0.20 + 709 \times 0.07 + 2042 \times 0.03$

$= 4171.$

$N_5 = N_0 p_5 + N_1 p_4 + N_2 p_3 + N_3 p_2 + N_4 p_1$

$= 10,000 \times 0.15 + 300 \times 0.40 + 709 \times 0.20 + 2042 \times 0.07 + 4171 \times 0.03 = 2030.$

$N_6 = N_0 p_6 + N_1 p_5 + N_2 p_4 + N_3 p_3 + N_4 p_2 + N_5 p_1$

$= 10,000 \times 0.15 + 300 \times 0.15 + 709 \times 0.40 + 2042 \times 0.20$

$+ 4171 \times 0.07 + 2030 \times 0.03 = 2590.$

From the values of N_i $(i = 0, 1, 2, \ldots, 6)$ so calculated, it can be seen thac the expected number of resistor failing each month increases upto fourth month and then starts decreasing and later increases in the sixth month. Thus N_i will oscilate till the system acquires steady state. The expected life of each resistor is given by

$$\text{Expected life} = \sum_{i=1}^{6} x_i p(x_i)$$

$= 1 \times 0.03 + 2 \times 0.07 + 3 \times 0.20 + 4 \times 0.40 + 5 \times 0.15 + 6 \times 0.15 = 4.02$ months.

Average number of failures per month is

$$\frac{N}{\text{Mean age}} = \frac{10,000}{4.02} = 2487.5 = 2488 \text{ resistors.}$$

Hence the total cost of individual replacement at the cost of Re 1 per resistor will be Rs. 2488x1=Rs. 2848.

The cost of replacement of all the reistors at the same time can be calculated as follows:

End of month	Total cost of group replacement (Rs.)	Average cost per month (Rs.)
1.	300x1+10,000x0.35 = 3,800	3800.00
2.	(300+709)x1+10,000x0.35 = 4,509	2254.50
3.	(300+709+2042)x1+10,000x0.35=6,551	**2183.66**
4.	(300+709+2042+4171)+10,000x0.35=10,722	2680.50
5.	(300+709+2042+4171+2030)x1+10,000x0.35 = 12,752	2550.40
6.	(300+709+2042+4171+2030+2590)x1+10,000x0.35 = 15,442	2557.00

Since the average cost per month of Rs. 2183.66 is obtained in the third month, it is optimal to have a group replacement after every third month.

Example 117.7. The following mortality rates have been observed for a certain type of fuse :

Week	1	2	3	4	5
% failing by the end of week	5	15	35	57	100

There are 100 fuses in use and it costs Rs. 5 to replace an individual fuse. If all fuses were replaced simultaneously it would cost Rs. 1.25 per fuse. It is proposed to replace all fuses at fixed intervals of time, whether or not they have burnt out, and to continue replacing burnt out fuses as they fail. At what intervals the group replacement should be made? Also, prove that this optimal policy is superior to the straight forward policy of replacing each fuse only when it fails.

Solution. Let p_i be the probability that a fuse which was new when placed in position for use, fails during the ith week of its life. Then the following probability distribution is obtained assuming to replace burnt out fuses as and when they fails.

Solution. Let p_i be the probability that a fuse which was new when placed in position for use fails during the ith week of its life. Then the following probability distribution is obtained assuming to replace burnt out fuses as and when they fail.

Week	1	2	3	4	5
Probability	5/100	(15–5)/100	(35–15)/100	(75–35)/100	(100–75)/100
of failure :	= 0.05	=0.10	=0.20	=0.40	=0.25

Furthermore, assume that

(i) fuses that fail during a week are replaced just before the end of the week, and

(ii) the actual percentage of failures during a week for a sub-pupulation of fuses with the same age is the same as the expected percentage of failures during the week for that sub-populations.

Let N_i be number of replacements made at the end of the ith week, when all N_0=1000 fuses are new intitially. Thus expected number of failures at different weeks can be calculated as shown in the following table.

End of week	Expected number of failures (Replacements)
1.	$N_1 = N_0 p_1$=100 x 0.05 = 500
2.	$N_2 = N_0 p_2 + N_1 p_1$=1000x0.10+50x0.05=102
3.	$N_3 = N_0 p_3 + N_1 p_2 + N_2 p_1$=1000x0.02+50x0.10+102x0.05=210
4.	$N_4 = N_0 p_4 + N_1 p_3 + N_2 p_2 + N_3 p_1$
	= 100x0.4+50x0.2+102x0.10+210x0.05=430
5.	$N_5 = N_0 p_5 + N_1 p_4 + N_2 p_3 + N_3 p_2 + N_4 p_1$
	= 1000x0.25+50x0.40+120x0.20+210x0.10+430x0.05=333

From this table, we observed that the expected number of fuses failing each week increases till the fourth week and then start decreasing. Thus the value of N_i will oscillate till the system acquires a steady state in which the proportion of bulbs failing each month is reciprocal of their average life. The expected life of a fuse can be calculated as follows:

Expected life = $1 \times p_1 + 2 \times p_2 + 3 \times p_3 + 4 \times p_4 + 5 \times p_5$

= 1 x 0.05+2 x 0.1+3 x 0.2+4 x 0.4+5 x 0.25

= 3.70 weeks

Expected number of failures during a week in steady state condition becomes 1000/3.70=270 fuses.

Under individual replacement policy the cost of replacement only on failure is

expected number of failures x cost per fuse

= 270 x 5 = Rs. 1350

Under the group replacement policy along with individual replacement, the cost of replacement is show in the following table:

End of week	Cost of individual replacement	Total cost of replacement Individual + Group	Average cost per week
1	50x5=250	50x5+1000x1.25=1500	1500
2	102x5=510	102x5+1250=1700	880
3	210x5=1050	210x5+1250=2300	766.66

The above table shows that the cost of individual replacement in the third week (Rs. 1050) is more than the average cost for two weeks. Hence, it is economical to replace all the fuses after every two weeks, otherwise the average cost will start increasing.

117.7. Other Replacement Problems

117.7.1 Staffing Problem

So far we have discussed replacement problems which are not related with human beings working in an organization. The principles of replacement may be applied to formulate some useful recruitment and promotion policies for the staff working in the organization.

For this we assume that life distribution for the service of staff in the organization is already known.

Example 117.8. An airline requires 200 assistant hostesses, 300 hostesses, and 50 supervisors. Women are recuited at the age of 21, and if still in service retire at 60. Given the following life table, determine

(i) How many women should be recruited in each year?

(ii) At what age should promotion take place?

Airline Hostesses' Life Record

Age	21	22	23	24	25	26	27	28	29	30
No. in Service	1000	600	480	384	307	261	228	206	190	181
Age	31	32	33	34	35	36	37	38	39	40
No. in Service	173	167	161	155	150	146	141	136	131	125
Age	41	42	43	44	45	46	47	48	49	50
No. of Service	119	113	106	99	93	87	80	73	66	59
Age	51	52	53	54	55	56	57	58	59	
No. in Service	53	46	39	33	27	22	18	14	11	

Solution. If 1000 women had been recuited each year for the past 39 years, then the total number of them recruited at the age of 21 and those serving upto the age of 59 is 6480. Total number of women recruited in the airline are 200+300+50=550.

(i) Number of women to be recruited every year in order to maintain a strength of 550 hostesses

=550x(1000/6480) = 85 approx.

(ii) If the assistant hostesses are promoted at the age of x, then upto age $(x-1)$, 200 assistant hostesses will be required. Among 550 women, 200 are assistant hostesses. Therefore, out of a strength of 1000 there will be 200x(1000/550)=364 assistant hostesses. But, from the life table given in the question, this number is available upto the age of 24 years. Thus, the promotion of assistant hostesses is due in the 25th year.

Since out of 550 recruitments only 300 hostesses are needed, if 1000 girls are recruited, then only 1000x(300/550)=545 (approx.) will be hostesses.

Hence total number of hostesses and assistant hostesses in a recruitment of 100 will be 545+364=909. This means, only 1000–909=91 supervisors are required. But, from life table this number is available upto the age of 46 years. Thus promotion of hostesses to supervisors will be due in 47th year.

117.7.2. Equipment Renewal Problem

The term renewal here refers to either replacing old equipment by new or to repairing it so that the probability density function of its future life time is equivalent to that of a new piece of equipment. The future life time of the equipment is considered to be a random variable.

The probability that an equipment will need a renewal in the interval t to $t+dt$ is called the renewal rate at time t, provided it is in running order at age t. It is given by $r(t)\,dt$ (also called renewal density function).

Example 117.9. A certain piece of equipment is extremely difficult to adjust. During a period when no adjustment is made, the running cost increases linearly with time, at a rate of b rupees per hour. The running cost immediately after an adjustment is not known precisely until the adjustment has been made. Before the adjustement, the resulting running cost x is deemed to be a random variable x with density function $f(x)$. If each adjustment costs k rupees, when should the replacement be made?

Solution: The running cost Rs. x is a random variable with density function $f(x)$. Suppose that the maximum of x be X.

If the adjustment is made when running cost equals Z, then there can be two possibilities

(i) $Z > X$ and (ii) $Z > X$.

Case I. $Z > X$. Let Rs. x be the running cost at time $t = 0$.

If adjustment is made after time t, then the running cost at time t will be Rs. $(x+bt)$, because running cost increases at the rate of Rs. b per hour. Obviously,

$$Z = x + bt \text{ or } t = \frac{Z - x}{b}.$$

If $c(Z)$ is the total cost incurred between the period of one adjustment to another adjustment, then

$c(Z)$ = Cost of one adjustment + total running cost from $t = 0$ to $t = (Z-x)/b$

$$= k + \int_0^{(Z-x)/b} (x + bt)\,dt$$

$$= k + \left[\frac{(x+bt)^2}{2b}\right]_0^{(Z-x)/b} = k + \frac{1}{2b}\left(Z^2 - x^2\right).$$

Therefore, the average cost per hour is given by average cost poer hour $= \frac{c(Z)}{t} = \frac{kb}{Z-x} + \frac{Z+x}{2}$.

Since the running cost x is random variable with density function therefore expected cost per hour is given by

$$E\{c(Z)\} = \int_0^x \left(\frac{kb}{Z-x} + \frac{Z+x}{2}\right) f(x)\,dx.$$

The value of $E\{c(Z)\}$ will be minimum for some value of Z, for which

$$\frac{d}{dZ}\{E(c(Z))\} = 0 \text{ and } \frac{d^2}{dZ^2}\{E(c(Z))\} > 0.$$

Now, $\frac{d}{dZ}\left[E\{c(Z)\}\right] = \frac{d}{dZ}\int_0^x \left(\frac{kb}{Z-x} + \frac{Z+x}{2}\right) f(x)\,dx$

$$= \int_0^x \frac{d}{dZ}\left(\frac{kb}{Z-x} + \frac{Z+x}{2}\right) f(x)\,dx$$

$$= \int_0^x \left(-\frac{kb}{(Z-x)^2} + \frac{1}{2}\right) f(x)\,dx$$

$$= \frac{1}{2} - kb\int_0^x \frac{f(x)}{(Z-x)^2}\,dx \qquad \left[\because \int_0^x f(x)\,dx = 1.\right]$$

For $E\{c(Z)\}$ to be minimum, we must have $\frac{d}{dZ}\left[E\{c(Z)\}\right] = 0$, which gives $\frac{1}{2} - kb\int_0^x \frac{f(x)}{(Z-x)^2}\,dx = 0$

i.e. $\int_0^x \frac{f(x)}{(Z-x)^2}\,dx = \frac{1}{2kb}$.

Hence the value of Z can be determined from the above equation.

Case II. $Z<X$. In this case, it can be shown that the minimum of Z can not occur and therefore, its optimal value can only be determined in Case I.

117.8. Module Summary

In this module, we have discussed about different kinds of failure, replacement of items when deteriorates with time in case of constant money value and variable money value. We also discussed about the replacement of items that fail completely based on individual replacement policy and group replacement policy. Staffing problem and equipment renewal problem are also discussed here. Several examples are given in different types of problems.

117.9. Self Assessment Questions

1. What is replacement? Describe some important replacement situations.
2. Describe the problem of replacement of items whose maintainance cost increase with time. Assume that the value of money remains constant.
3. Machine *A* costs Rs. 45,000 and operating costs are estimated at Rs. 1000 for the first year, increasing by Rs. 10,000 per year in the second and subsequent years. Machine *B* costs Rs. 50,000 and operating costs are Rs. 2,000 for the first year, increasing by Rs. 4,000 in the second and subsequent years. If we now have a machine of type *A*, should we replace it with *B*? If so when? Assume that both machines have no resale value and future costs are not discounted.
4. The data on the running costs per year and resale price of equipment *A* whose purchase price is Rs. 2,00,000 are as follows:

Year	1	2	3	4	5	6	7
Running Cost (Rs.):	30,000	38,000	46,000	58,000	72,000	90,000	1,10,000
Resale value (Rs.):	1,00,000	50,000	25,000	12,000	8,000	8,000	8,000

   (i) What is the optimum period of replacement?

   (ii) What equipment *A* is two years old, equipment *B* which is a new model for the same usage is available. The optimum period for replacement is 4 years with an average cost of Rs. 72,000.00. Should equipment *A* changed with equipment *B*? If so, when?

5. The cost of a machine is Rs. 6,100 and its scrap value is Rs. 100. The maintenance costs found from experience are as follows:

Year	1	2	3	4	5	6	7	8
Maintenance cost (Rs.):	100	250	400	600	900	1200	1600	2000

   When should the machine be replaced?

6. Assume that present value of one rupee to be spent in a year's time is Rs. 0.9 and C = Rs. 3,000, capital cost of equipment and the running cost are given in the following table.

Year	1	2	3	4	5	6	7
Running Cost (Rs.) :	500	600	800	100	1300	1600	2000

When should the machine be replaced?

7. State some of the simple replacement policies and give the average cost functions for the same explaing your notations.

8. The cost of maintenance of a machine is given as a function that the average annual cost will be minimized by replacing the machine when the average cost to date becomes equal to the current maintenance cost.

9. Explain how the theory of replacement is used in the following problems:

   (i) Replacement of items whose maintenance cost varies with time.

   (ii) Replacemnt of items that fail completely.

10. The following failure rates have been observed for a certain type of light bulbs:

Year	1	2	3	4	5	6	7	8
Prob. of failure to date :	0.05	0.13	0.25	0.43	0.68	0.88	0.96	1.00

The cost of replacing an individual bulb is Rs. 2.25, the decision is made to replace all bulbs simultaneously at fixed intervals, and also to replace individual bulbs as they fail in service. If the cost of group replacement is 60 paise per bulb and the total number of bulbs is 1000, what is the best interval between group replacements?

11. The following mortality rates have been observed for a special type of light bulbs :

Month	1	2	3	4	5
Percent failing at the end of month :	10	25	50	80	100

In an individual unit there are 1000 special type of bulbs in use, and it costs Rs. 10 to replace in individual bulb which has burnt out. If all bulbs were replaced simultaneously it would cost Rs. 2.50 per bulb. It is proposed to replace all bulbs at fixed intervals, whether or not they have burnt out, and to continue replacing burnt out bulbs as they fail. At what intervals of time should the manager replace all the bulbs?

12. A computer has 20,000 resistors. When any resistor fails, it is replaced. The cost of replacing a resister individually is Re 1. If all the resistors are replaced at the same time the cost per resistor is reduced to Re

0.40. The percent surviving at the end of month *t*, and the probability of failure during the month *t* are given below:

Year	0	1	2	3	4	5	6
Percent surviving at the end of *t*	100	96	90	65	35	20	0
Prob. of failure during the month *t*	-	0.04	0.06	0.25	0.30	0.15	0.20

What is the optimum replacement plan?

13. Calculate the probability of a staff resignation in each year from the following survival table.

Year	0	1	2	3	4	5	6	7	8	9	10
No. of original staff in service at the end of year	1000	940	820	580	400	280	190	130	70	30	0

14. An airline, whose staff are subject to the same survival rates as in the previous problem, currently has a staff whose ages are distributed in the following table. It is estimated that for the next two years staff requirements will increse by 10% per year. If women are to be recruited at the age of 21, how many should be recruited for the next year and at what age will promotions take place? How many should be recruited for the following year and at what age will promotions take place?

Assistant

Age :	21	22	23	24	25
Number :	90	50	30	20	10 (Total 200)

Hostesses

Age :	26	27	28	29	30	31	32	33	34
Number :	40	35	35	30	28	26	20	18	16
Age :	35	36	37	38	39	40	41		
Number :	12	10	8	-	8	6	(Total 300)		

Supervisors

Age :	42	43	44	45	46	47	48	49	50	
Number :	5	4	5	3	3	3	6	2	-	
Age :	51	52	53	54	55	56	57	58	59	
Number :		4	3	5	-	3	2	-	2	(Total 50)

14. Suppose the life X of electric light bulbs follows the gamma distribution

$$p(x \le X < x + dx) = \frac{a^p}{\Gamma(p)} e^{-a\lambda} x^{p-1} dx, a > 0, 0 \le x < \infty.$$

Determine the renewal rate for one point at the end of time period $(0, t)$.

117.10. References

1. J.K. Sharma, Operations Research : Theorey and Applications, Macmilan.
2. S.D. Sharma, Operations Research, Kedar Nath Ram Nath & Co.
3. Prem Kumar Gupta and D.S. Hira, Operations Research, S. Chand.

---- 0 ----

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X Special Paper : Operations Research

Module No. - 118

Advanced Optimization and Operations Research-II

(Simulation)

Module Structure :

118.1 Introduction

Simulation is a method of solving decision making problems by desiring, constructing or manupulating a model of a real system. It is defined to be the action of performing experiment on a model of a given system. It duplicates the essence of a system or activity without actually obtaining the reality.

Simulation involves the construction of symbolic model that describes the systems operation in terms of individual events and components dividing the system into smaller components and combining them in their logical order, analyses the effect of their interactions on one another studing the various alternatives with respect to the performance of the model and choosing the best one.

Simulation must be treated as a statistical experiment unlike the mathematical models where the output of the model represents a long run steady state behaviour. A simulation experiment differs from regular laboratory experiment in the sense that it can be conducted totally on the computer by expressing the interaction among the components of the system as mathematical relationship.

118.2. Objectives

Go through this module, you will learn

* The definition of simulation.
* Types of simulation
* Merits and demerits of simulation
* Application of simulation
* Generation of random numbers
* Monte Carlo method of simulation
* Problems on simulation

118.3 Key words

Simulation, Stochastic simulation, Random numbers, Monte Carlo method.

118.4. When to use simulation :

Techniques like linear programming, dynamic programming, queuing theory, network models, etc. are not sufficient to tackle all the important managerial problems requiring data analysis. East technique has its own limitations.

Linear programming models assume that the data do not alter over the planning horizon. It is one time decision process and assumes average values for the decision variables. If the planning horizon is long, the multiperiod linear programming model may deal with the yearly average data, but will not take into account the variations over the months and weeks with the results that month to month and week to week operations are left implicit.

Dynamic programming models can be used to tackle very simple situations involving only a few variables. If the number of state variables is a bit larger, the computation task becomes quite complex and involved.

In general, the simulation technique is a dependable tool in situations where mathematicsl analysis is either too complex or too costly.

118.5 What is simulation?

Simulation is an immitation of reality. In the laboratories a number of experiments are performed on simulated models to determine the behaviour of the real system in true environments. A simple illustration is the testing of an aircraft mode in a wind tunnel from which we determine the performance of the actual aircraft under real operating conditions. Planetarium represents a beautiful simulation of the planet system.

118.6. Types of simulation

Simulation is mainly of two types :

(i) Analogue (or environmental) simulation. The simple examples cited in section 118.5 are of simulating the reality in physical form, which we may refer as analogue (or environmental) simulation.

(ii) Computer (or system) simulation. For the complex and intricate problems of managerial decision making, the analogue simulation may not be applicable. A mathematical model is formulated for which a computer programme is developed, and then the problem is solved by using high speed electronic computer. Such system of simulation is called a computer simulation or system simulation.

118.7. Advantages of the simulation technique

1. This approach is suitable to analyse large and complex real life problems which cannot be solved by usual quantitative methods.
2. Simulation eliminates the need of costly trial and error methods of trying out the new concept on real methods and equipment.
3. Simulation has the advantage of being relatively free from mathematics and can be easily understood by the operating personnel.
4. Simulation models are compartively flexible and can be modified to accommodate the changing environments of the real situation.
5. Computer simulation can compress the performance of a system over several years and involving large calculations into a few minutes.
6. The simulation technique is easier to use than mathematical model and is quite superior to the mathematical analysis.

118.8 Limitations of the simulation technique

1. It is the trial and error approach that produce different solutions in repeated runs. That is, it does not generate optimal solutions to problem.
2. The simulation model does not produce solution by itself. The user has to provide all the constraints for the solutions which he wants to examine.
3. Sometimes simulation models are expensive and take a long time to develop it.
4. Each application of simulation is adhoc to a great extent.

118.9 Applications of simulation

1. In the field of basic sciences, it has been used to evaluate the area under a curve, to estimate the value of π, in matrix inversion and study of particle diffusion.
2. In industrial problems including the design of queuing systems, inventory control, communication networks, chemical processes, nuclear reactors and scheduling of production processes.
3. In business and economic problems, including customer behaviour, price determination, economic forecasting and capital budgeting.
4. In social problems, including population growth, effect of environment on health and group behaviour.

5. In biomedical systems, including fluid balance, distribution of electrolyte in human body and brain activities.
6. In the design of weapon systems, war strategies tactics.

118.10 Steps of simulation process

Step 1 : Identify the problem

Step 2 : (a) Identify the decision variables

(b) Decide the performance criterion and decision rules.

Step 3 : Construct a numerical model so that it can be analysed on the computer.

Step 4 : Validate the model to ensure whether it is truely respresenting the system being analysed and the results will be reliable.

Step 5 : Design the experiments to the conducted with the simulation model.

Step 6 : Run the sumulation model on the computer to get the results.

Step 7 : Examine the results in terms of problem solution as well as their reliability and correctness.

118.11 Stochastic simulation and random numbers

When a system contains certain decision variables that can be represented by a probability distribution, the simulation model used to study this type of system is called the stochastic simulation.

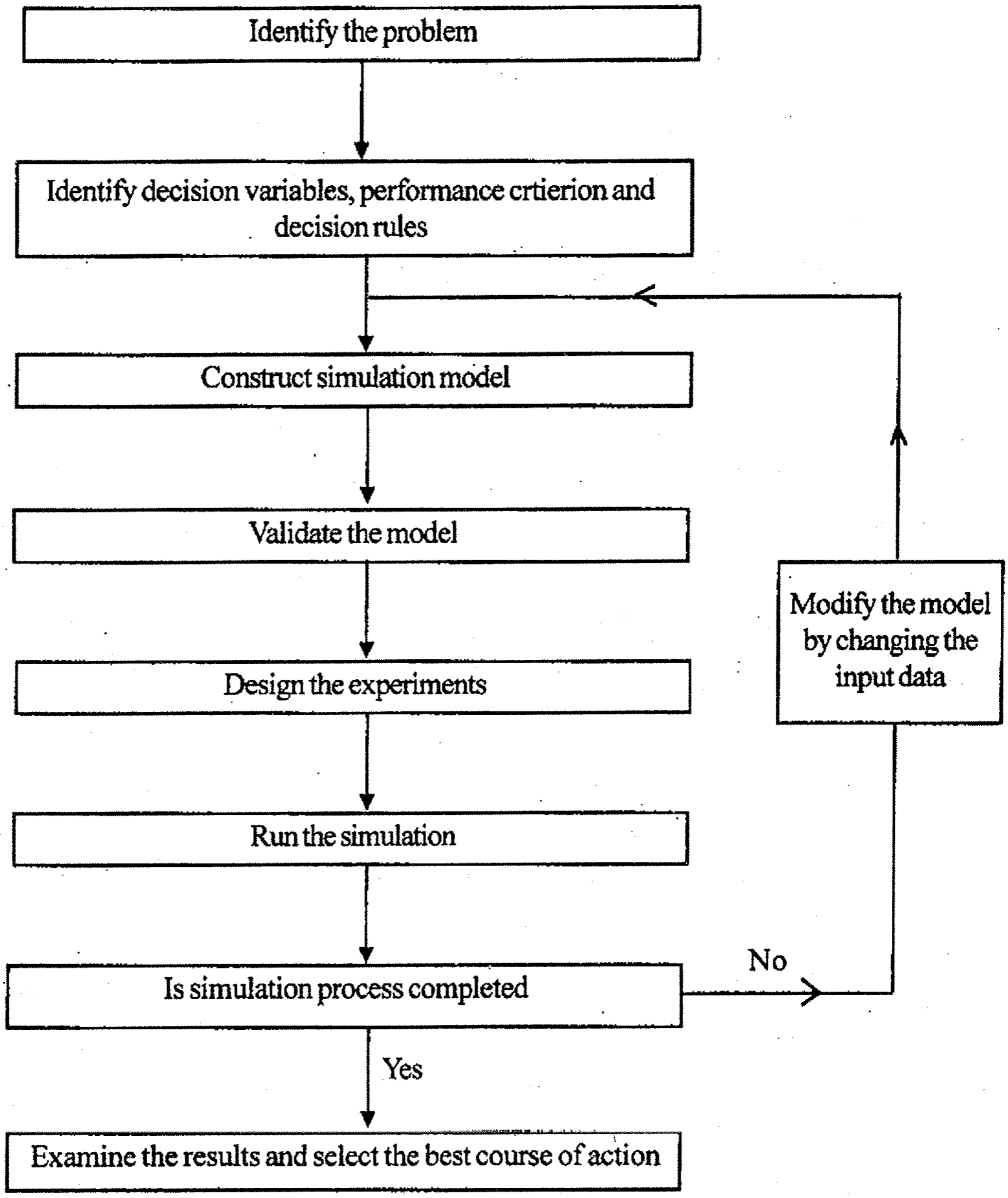
Stochastic simulation models use random numbers to generate in certain events.

118.11.1 Monte Carlo simulation

The Monte Carlo method of simulation was developed by the two mathematicians John Von Neumann and Stainslaw Ulam, during World War II to study how far neutrons would travel through different materials. The technique provided an approximate but quite workable solution to the problem.

The steps involved in carrying out Monte Carlo simulation are :

1. Select the objective function of the problem. It is either to be maximized or minimized.
2. Identify the variables that effect the measure of effectiveness significantly.
3. Determine the cumulative probability distribution of each variable selected in step 2. Plot these distributions with values of the variables along X-axis and cumulative probability values along the *Y*-axis.


Identify the problem
Identify decision variables, performance crtierion and decision rules
Construct simulation model
Validate the model
Design the experiments
Run the simulation
Is simulation process completed
No
Modify the model by changing the input data
Yes
Examine the results and select the best course of action

4. Get a set of random numbers.
5. Consider each random number as a decimal value of the cumulative probability distribution. Each cumulative distribution plot along the Y-axis. Project this point horizontally till it meets the distribution curve. Then project the point of intersection drawn on the X-axis.
6. Record the value generated in step 5. Substitute in the formula chosen for measure of effectiveness and find its simulated value.
7. Repeat steps 5 and 6 until sample is large enough to the satisfaction of the decision maker.

118.11.2 Generation of random numbers

The random numbers are generated by a random process and these numbers are the values of a random variable. Several methods are available to generate random numbers. Practically, these methods do not generate ideal random numbers, because these methods follow some algorithms. So the available methods generate pseudo random numbers.

The commonly used simplest method to generate random numbers is the power residue method. A sequence of non negative integers $x_1, x_2, \ldots$ is generated from the following relation

$$x_{n+1}=(ax_n) \pmod m$$

where x_0 is a starting value called the seed, a and m are two positive integers ($a<m$). The expression $(ax_n) \pmod m$ gives the remainder when ax_n is divided by m. The possible values of x_{n+1} are 0, 1, 2, ..., $m-1$, i.e. the number of different random numbers is m. The period of random number depends on the values of a, x_0 and m. Appropriate choice of a, x_0 and m gives a long period random numbers.

Suppose the computer which will generate the random numbers have a word length of b bits. Let $m = 2^{b-1}-1$, a = an odd integer of the from 8k±3 and closed to $2^{b/2}$, x_0 = an odd integer between 0 and m. Now ax_0 is a $2b$ bits integer. The least b significant bits form the random numbers x_1. The process is repeated for a desired number of times.

For a 32-bit computer,

$$m = 2^{31}-1 = 2147483647, a = 2^{16}+3 = 65539, x_0 = 1267835015, (0 < x_0 < m).$$

To obtain the random numbers between [0, 1], all numbers are divided by $m-1$. These numbers are uniformly distributed over [0,1]. The function RANDO1() generate a random number between 0 and 1. The following program will generate 50 random numbers between 0 and 1.

```
C       FUNCTION TO GENERATE RANDOM NUMBERS BETWEEN 0 AND 1
        FUNCTION RANDO1( )
        INTEGER * 4 A, X0
        REAL M
        PARAMETER (M = 2147483647.0)
        DATA A, X0/65539, 1267835015/
        XO = IABS (A* X0)
        RANDO1 = X0/M
        RETURN
        END
C       MAIN PROGRAM TO GENERATE 50 RANDOM NUMBERS
C       BETWEEN 0 AND 1
        DIMENSION X (50)
        WRITE (*, *)' 50 RANDOM NUMBERS BETWEEN 0 AND 1'
        DO 3 I = 1, 50
3       X(I) = RANDO1( )
        WRITE (*, 4) (X(J), J=1,50)
4       FORMAT (1X, 10(F6.4, 1X))
        END
```

118.12 Applications

118.12.1 Evaluation of π by Monte Carlo method. Let $x^2+y^2=1$ be the equation of a circle whose centre at (0, 0) and radius 1. An approximate area of the circle can be found by counting the random points in a square of side 1 unit into which a circle of radius 1 is inscribed. Consider the first quadrant of the circle.

Now generate a pair of random numbers x and y between 0 and 1. If these random numbers satisfying the inequation $x^2 + y^2 \leq 1$ lie within the first quadrant of the circle. The area of the square is 1 and that of the circle in first quadrant is $\pi/4$. Now generate a large number of random pairs and count the number of points within the first quadrant of the circle. If n_1 be the number of points within the first quadrant of the circle and n be the total number of points generated then the reqired value of $\frac{\pi}{4}$ is $\frac{n_1}{n}$.

The following program computes the value of π using Monte Carlo method.

```
C     EVALUATION OF PI BY MONTE CARLO METHOD
      PRINT *, 'ENTER NUMBER OF POINTS TO BE GENERATED'
      READ *, N
      N1 = 0,
      DO11 I = 1, N
       X = RANDO1( )
       Y = RANDO1( )
       IF (X * X + Y * Y. LE. 1.) N1=N1+1
11    CONTINUE
      PI = 4.0 * REAL (N1)/N
      WRITE (*, 22) N, PI
22    FORMAT (1X, 'N=', I5, 1H, 2X, 'VALUE OF PI IS:', F9.5)
      END
C     FUNCTION TO GENERATE RANDOM NUMBER BETWEEN 0 AND 1
      FUNCTION RANDO1( )
      INTEGER * 4 A, X0
      REAL M
      M = 2147483647.0
      DATA A, X0/65539, 1267835015/
      X0=IABS (A*X0)
      RANDO1 = X0/M
      RETURN
      END
```

118.12.2 Simulation of coin tossing

By generating the random numbers between 0 and 1, the probability of getting 'head' or 'tail' in a single toss can be determined. Suppose 0 and 1 represent respectively the head and tail.

Generate a large number of random numbers, say n, and count the occurence of 0, let it be n_1, then the bability of occurence of head in a single trial is $\frac{n_1}{n}$ and similarly the probability of tail is $\frac{n-n_1}{n}$. The following gram finds the probability of occurence of head in a single trial.

```
C      SIMULATION OF COIN TOSSING, 0 AND 1 RESPECTIVELY
C      REPRESENT HEAD AND TAIL.
       INTEGER RANDON
       READ *, N
       N1 = 0
       DO 2 I = 1, N
       M = RANDON(2)
       IF (MOD (M, 2). EQ.0) N1= N1+ 1
2      CONTINUE
       P = REAL (N1)/N
       WRITE (*, 9) N1, N, P
9      FORMAT (1X, 'FAVOURABLE POINTS=', I5, 1X, 'TOTAL
+      POINTS=', I5, 2X, 'PROBABILITY=', F7.4)
       END
C      FUNCTION TO GENERATE A RANDOM NUMBER BETWEEN 0 AND N
       INTEGER FUNCTION RANDON (N)
       REAL M
       INTEGER * 4 A, X0
       M = 2147483647.0
       DATA A, X0/65539, 1267835015/
       X0=IABS (A* X0)
       RANDON = N* (X0/M)
       RETURN
       END
```

118.13 Simulation of Inventory Problems

Example 118.13.1 Using random numbers to stimulate a sample, find the probability that a packet of 6 products does not contain any defective product, when the production line produces 10% defective products. Compare the answer with the expected probability.

Solution: Given that 10% of the total production is defective and 90% is non-defective. If we have 100 random numbers (0 to 99), then 90 of them (or 90%) represent non-defective products and remaining 10 of them (or 10%) represent defective products. Thus, the random numbers 00 to 89 are assigned to variables representing non-defective products and 90 to 100 are assigned to variables representing defective products.

If we choose a set of 2-digit random numbers in the range 00 to 99 to represent a packet of 6 products as shown below, then we would expect that 90% of the time they would fall in the range 00 to 89.

Sample number	Random Number					
A	86	02	22	57	51	68
B	39	77	32	77	09	79
C	28	06	24	25	93	22
D	97	66	63	99	61	80
E	69	30	16	09	05	53
F	33	63	99	19	87	26
G	87	14	77	43	96	43
H	99	53	93	61	28	52
I	93	86	52	77	65	15
J	18	46	23	34	25	85

Here it may be noted that out of ten simulated samples 6 contains one or more defectives and 4 contain no defectives. Thus the expected percentage of non-defective products is 40%. However, theoretically the probability that a packet of 6 products containing no defective product is $(0.9)^6 = 0.53144 = 53.14\%$.

Example 118.13.2. A bakery keeps stock of a popular brand of cake. Previous experience shows the daily demand pattern for the item with associated probabilities, as given below:

Daily demand (number)	: 0	10	20	30	40	50
Probability	: 0.01	0.02	0.15	0.50	0.12	0.02

use the following sequence of random numbers to simulate the demand for next 10 days.

Random numbers : 25, 39, 65, 76, 12, 05, 73, 89, 19, 49.

Also estimate the daily average demand for the cakes.

Solution : Using the daily demand distribution, we obtain a probability as shown in the table

Daily Demand	Probability	Cumulative probability	Random number interval
0	0.01	0.01	00
10	0.02	0.01 + 0.20 = 0.21	01 – 20
20	0.15	0.21 + 0.15 = 0.36	21 – 35
30	0.50	0.36 + 0.50 = 0.86	36 – 85
40	0.12	0.86 + 0.12 = 0.98	86 – 97
50	0.02	0.98 + 0.02 = 1.00	98 – 99

Conduct the simulation experiment for demand by taking a sample of 10 random numbers, which represent the sequence of 10 samples. Each random sample number here is a sample of demand.

The simulation calculations for a period of 10 days are given in the following table.

Days	Random number	Demand	
1	40	30	because 0.36<0.40<0.85
2	19	10	because 0.01<0.19<0.02
3	87	40	and so on.
4	83	30	
5	73	30	
6	84	30	
7	29	20	
8	09	10	
9	02	10	
10	20	10	
		Total = 220	

Expected demand = 220/10 = 22 units per day.

118.14 Simulation of Queuing Problems

Example 118.14.1 A firm has a single channel service station with the following arrival and service time probability distributions:

Inter-arrival time (Minutes)	Probability	Service time (Minutes)	Probability
10	0.10	5	0.08
15	0.25	10	0.14
20	0.30	15	0.18
25	0.25	20	0.24
30	0.10	25	0.22
		30	0.14

The customer's arrival at the service station is a random phenomenon and the time between the arrival varies from 10 minutes to 30 minutes. The service time varies from 5 minutes to 30 minutes. The queuing process begins at 10 A.M. and proceeds for nearly 8 hours. An arrival goes to the service facility immediately, if it is free. Otherwise it will wait in a queue. The queue discipline is first come-first-served. If the attendant's wages are Rs. 10 per hour and the customer's waiting time costs Rs. 15 per hour, then would it be an economical proposition to engage a second atendant? Answer using Monte Carlo simulation technique.

Solution : The cumulative probability distributions and random number interval both for inter-arrival time and service time are shown in the following table 1 and table 2.

Table –1

Inter-arrival time (Minutes)	Probability	Cumulative probability	Random number interval
10	0.10	0.10	00 – 09
15	0.25	0.35	10 – 34
20	0.30	0.65	35 – 64
25	0.25	0.90	65 – 89
30	0.10	1.00	90 – 99

Table – 2

Inter-arrival time (Minutes)	Probability	Cumulative probability	Random number interval
5	0.08	0.08	00 – 07
10	0.14	0.22	08 – 21
15	0.18	0.40	22 – 39
20	0.24	0.64	40 – 63
25	0.22	0.86	64 – 85
30	0.14	1.00	86 – 99

The simulation work sheet developed to the given problem is shown in Table 3.

Table – 3

Arrival number (1)	Random number (2)	Arrival Interval (3)	Arrival time (4)	Service time (5) begin	Waiting time (6)	Random number (7)	Service time (8) end	Exist time (9)	Time in system (10)=(6)+(8)
1	20	15	15	15	0	26	15	30	15
2	73	25	40	40	0	43	20	60	20
3	30	15	55	60	5	98	30	90	35
4	99	30	85	90	5	87	30	120	35
5	66	25	110	120	10	58	20	140	30
6	83	25	135	140	5	90	30	170	35
7	32	15	150	170	20	84	25	195	45
8	75	25	175	195	20	60	20	215	40
9	04	10	185	215	30	08	10	225	40
10	15	15	200	225	25	50	20	245	45
11	29	15	215	245	30	37	15	260	45
12	62	20	235	260	25	42	20	280	45
13	37	20	255	280	25	28	15	295	40
14	68	25	280	295	15	84	25	320	40
15	94	30	310	320	10	65	25	345	35

From the 15 samples of waiting time, 225 minutes and the time spent 545 minutes by the customer in the system, we compute all average waiting time in the system and average service time as follows:

$$\text{Average waiting time} = \frac{225}{15} = 15 \text{ minutes}$$

$$\text{Average service time} = \frac{545}{14} = 36.33 \text{ minutes}$$

Thus the average cost of waiting and service per hour is given by

$$\text{Cost of waiting} = 15 \times \frac{15}{60} = \text{Rs. } 3.75$$

$$\text{Cost of service} = 10 \times \frac{36.33}{60} = \text{Rs. } 6.05$$

Since average cost of service per hour is more than the average cost of waiting per hour, therefore it would not be an economical proposition to engage a second attendant.

118.15 Role of Computers in Simulation

The role of computers in simulation is vital. They are used to generate random numbers, simulate the given problem with varying values of variables in few minutes and help the decision-maker to prepare reports which enable him to make decisions quickly.

Computer languages available to help the simulation process can be divided into two categories:

118.15.1 General purpose programming languages

The general purpose programming languages includes FORTRAN, BASIC, COBOL, PL/1, C, Pascal, etc.

To use these languages for simulation process an extensive programming experience is required.

118.15.2. Special purpose simulation languages

Special simulation languages have few advantages such as :

(i) They reduce programme preparation time and cost with features specially designed for simulation model.

(ii) They have the capability to readily generate different types of random variates automatic generation of certain types of statistical table and various other features.

(iii) They require little or no prior programming knowledge for use.

118.16 Module Summary

In this module we have discussed – what is simulation, when to use simulation, different types of simulation. Advantages and disadvantages of simulation techniques are also discussed. We have also discussed the Monte Carlo simulation technique and generation of random numbers. We have solved problems of inventory and queuing theory using simulation.

118.17 Selft assessment questions

1. What is simulation? Describe its advantages in solving the problems. Give its main limitations with suitable examples.
2. What is simulation? Describe simulation process. What are the reasons for using simulation.
3. What is the need of simulation?
4. Explain how do you apply Monte Carlo method for queuing problems.
5. Explain in brief the advantages and disadvantages of Monte Carlo methods.
6. Explain the different mathematical steps in a Monte Carlo method.
7. Describe a method for generation of random numbers. Generate 10 random numbers by using the method suggested.
8. The management of a bank is thinking of opening a drive in facility for its branch office in a commercial area. The inter-arrial times of the customers at the branch are as follows:

Inter-arrival time (minutes)	**Probability**
3	0.17
4	0.25
5	0.25
6	0.20
7	0.13

It is planned to have one cashier who can serve the customers at the following rate :

Service time (minutes)	Probability
3	0.10
4	0.30
5	0.40
6	0.15
7	0.05

Determine the number of spaces to be planned for the waiting cars. Simulate the operation of the facility for arriving sample of 25 cars. If the location has space for not more than two waiting cars how many customers would be turned away due to lack of space? What is the average waiting time of a customer?

9. The demand for a particular item has the probability distribution shown below:

Daily demand (units) :	4	5	6	7	8	9	10	11	12
Probability :	0.06	0.14	0.18	0.17	0.16	0.12	0.08	0.06	0.03

If the lead time is 5 days, using simulation study the implications of inventory policy of ordering 50 units whenever the inventory at the end of the day is 40 units. Assume the initial stock level of 75 units and run the simulation for 25 days.

10. Six truck loads of material are delivered everyday at a factory at regular intervals of 45 minutes. The trucks carry 2, 4, 2, 4, 3, 3 tons of material respectively. Unloading has to be carried out by teams of two men, each team capable of handling 800kg/hr, upto a maximum of 6 items simultaneously. Assuming that penalty for detaining a lorry for more than 45 minutes (including unloading time) is Rs. 10 per hour and the cost of each labourer is Rs. 8 per day, determine the least costly number of teams.

11. The following table gives the arrival, pattern at a Coffee Counter for 'one minute' intervals. The service is taken as 2 person in one minute in one counter.

No. of persons arriving :	0	1	2	3	4	5	6	7
Probability percentage:	5	10	15	30	20	10	5	5

Using Monte Carlo simulation technique and the following random numbers, generate the pattern of arrivals and queue formed when the following 20 random numbers are given :

5, 25, 16, 80, 35, 48, 67, 79, 90, 92, 9, 14, 1, 55, 20, 71, 30, 42, 60 and 85.

Find the queue length if two counters are used, i.e., 4 persons in one minute.

118.18 References

1. S.D. Sharma, Operations Research, Kedar Nath Ram Nath & Co.
2. Prem Kumar Gupta and D.S. Hira, Operations Research, S. Chand
3. J.K. Sharma, Operations Research, Macmillan.

------ 0 ------

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X **Special Paper – OR**

Module No. - 119

Advanced Optimization and Operations Research-II

(Information Theory – I)

STRUCTURE :

119.1. Introduction

The word information is very common in daily life problems. Information transmission usually occurs through human voice (as in telephone, radio, television, etc.), books, newspapers, letters, etc. In all these cases a piece of information is transmitted from one place to another. However, one might like to quantitatively assess the quantity of information contained in a piece of information. Some examples are listed below:

1. Suppose we are listening to the local weather forecast on the radio on first July and we hear the weather man, say, "Moonsoon will come during this week". Since it has probably not rained during first week of July for a decade or more, we would consider that such a statement by the weather man to be very reliable and we heard him say, "Moonsoon will not come during this week", and we would consider that such a statement by the weather man give us a very little information. Thus, according to our usual way of looking at information, if something is very likely to occur, the statement that it will occur does not give much information. On the other hand, if something is unlikely to occur, the statement that it will occur gives a good deal of information.
2. Suppose a man goes to a new community to rent a house and asks an unreliable agent. "Is this house cool in the summer season?" If the agent answers, "yes", the man has received very little information because more than likely the agent would have answered, "yes" regardless of the faits. If on the other hand, the man has a friend who lives in the neighbouring house, he can get more information by asking his friend the same question because the answer will be more reliable.

In general, the amount of information in the message be measured by the extent of the change in probability produced by the message.

In this module, we introduce the basic idea about information theorem. In the next module, we shall discuss about the measurement technique of information using entropy.

119.2. Objectives

After completion of this module you will able to:

- Answer the question "which is information?"
- Explain the communication system
- Explain different components of a communication system
- Explain memoryless channel, channel matrix, probability relation in a channel, noiseless channels.

119.3. Key words

Information, information channel, Receiver, Encoder, Decoder, Memoryless Channel, Channel matrix, Noiseless Channels.

119.4. Communication Processes

The communication process may be defined as the procedure by which one mind affects the another. This may be any means by which the information in carried from source to the receiver.

The essential terms of a communication system is explain below.

(a) **Source (or Transmitter):**

It is the source of message (either person or machine) which produces the information to be communicated or transmitted.

(b) **Communication Channel :** It is the transmission network or media which carries the message from the source to receiver, e.g., human voice, newspapers, books, etc. A Communication channel can be with or without noise.

(c) **Receiver :** It is the destination to which the message is conveyed from source or transmitter through a Communication channel.

The above three essential parts of communication system are related to each other as shown in Figure 119.1.

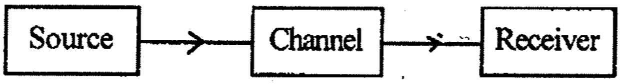


Figure 119.1 Components of a communication system.

Other parts of the communication system are as follows:

(a) **Encoder :** It is an equipment which is used to improve the efficiency of the transmission channel through which a message is transmitted to the receiver.

(b) **Noise :** It is the general term which creates interruptions or disturbances in the transmission of message or information from transmitter or receiver. For example, noise or disturbance in radio or television during the relay of a programme; error in newspaper printing, etc.

(c) **Decoder :** It is used to transform encoded message into the original form at the receivers end.

The general structure of a communication system with six parts is shown in Figure 119.2.

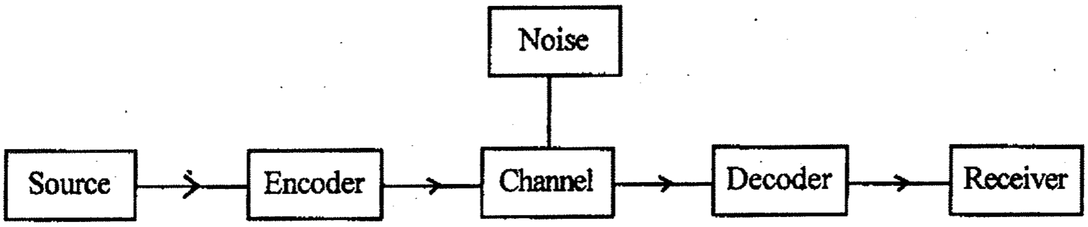


Figure 119.2. The general structure of a communication system.

Fundamental theorem of information theory

The information theory is essentially a study of so-called "fundamental theorem of information theory", which stated below.

"It is possible to transmit information through a noisy channel at any rate less than the channel capacity with an arbitrary small probability of error".

The communication system described in Figure 119.2 is statistical in nature because the source selects and transmits sequence of symbols from a given alphabet to the channel based on some statistical rule. The channel transmits this symbolic information to the receiver under some statistical rule also.

119.4. Memory less channel

A memory less channel is described by an input alphabet $X = \{x_1, x_2, ..., x_m\}$; an output alphabet $Y = \{y_1, y_2, ..., y_n\}$, and a set of conditional probabilities $P(y_j / x_i)$ for $i, j = 1,2$, then it is called a binary memory less channel.

A binary memory less channel is always symmetric because

$$p(y_1/x_1) = p(y_2/x_1) = q$$

$$p(y_1/x_2) = p(y_2/x_2) = p$$

where $q = 1-p, p$ being the probability of error in transmission.

119.4.2. Channel matrix

The input to the channel, the output from the channel and conditional probabilities for a pair of input symbol and output symbol can be expressed in the form of a matrix called channel matrix.

$$\text{Input } X \begin{array}{c} \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{array} \overset{\text{Output } Y}{\begin{array}{c} \begin{array}{cccc} y_1 & y_2 & \cdots & y_n \end{array} \\ \begin{bmatrix} p_{1/1} & p_{1/2} & \cdots & p_{1/n} \\ p_{2/1} & p_{2/2} & \cdots & p_{2/n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m/1} & p_{m/2} & \cdots & p_{m/n} \end{bmatrix} \end{array}}$$

where $p_{i/j} = P(y_j/x_i)$; $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$.

It may be remembered that sum of conditional probabilities in each row must be equal to one. That is,

$$\sum_{j=1}^{n} p_{j/i} = 1, \quad i = 1, 2,...., m.$$

For example, the channel matrix of the binary symmetric channel is given by $\begin{pmatrix} q & p \\ p & q \end{pmatrix}$.

119.4.3. Probability relation in a channel

Let us consider a channel with m input symbols $x_1, x_2, ... x_m$ and a output symbols $y_1, y_2, ..., y_m$; and the channel matrix

$$\begin{bmatrix} p_{1/1} & p_{2/1} & \cdots & p_{n/1} \\ p_{1/2} & p_{2/2} & \cdots & p_{n/2} \\ \vdots & \vdots & \cdots & \vdots \\ p_{1/m} & p_{2/m} & \cdots & p_{n/m} \end{bmatrix}$$

If $p_{io} = P(x_i), i = 1, 2, ..., m$ denotes the probability that symbol x_i will be selected for transmission through the channel and $p_{oj} = P(y_j), j = 1, 2, ..., n$ denote the probability that output symbol y_j will be received as channel output. Then the relation between the probabilities of various input symbols and output symbols may be obtained. The following relations may be easily obtained

$$\sum_{i=1}^{m} p_{io} p_{j/i} = p_{oj} \text{ for } j = 1, 2, ..., n \quad \text{......} (1)$$

If we are given the input probabilities p_{io} and the channel probabilities, then the output probabilities p_{oj} can be easily obtained from above relation.

Furthermore, the following probability relations also hold :

$$P(x_i, y_j) = p_{j/i}\, p_{io} \text{ for all } i,j \qquad \text{(2)}$$

and $P(x_i / y_j) = p_{j/i}(p_{io} / p_{oj})$ for all i,j. .... (3)

The relations (2) give the joint probabilities of sending a symbol x_i and receiving the symbol y_j, while relations (3) give the backward channel probabilities given that an output y_j has been received.

Example 1. Consider a binary channel with input symbols $A = \{0,1\}$, output symbols $B = \{0,1\}$ and the channel matrix $\begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{10} & \frac{9}{10} \end{pmatrix}$.

Let us assume the input probabilities $p_{10} = 3/4, p_{20} = 1/4$. Using (1), the following output probabilities are obtained:

$$p_{01} = \frac{3}{4}\cdot\frac{2}{3} + \frac{1}{4}\cdot\frac{1}{10} = \frac{21}{40},\ p_{02} = \frac{3}{4}\cdot\frac{1}{3} + \frac{1}{4}\cdot\frac{9}{10} = \frac{19}{40}.$$

The conditional backward input probabilities are obtained by using (3),

$$P(0/0) = \frac{\frac{2}{3}\cdot\frac{3}{4}}{\frac{21}{40}} = \frac{20}{21}$$

$$P(0/1) = \frac{\frac{1}{3}\cdot\frac{3}{4}}{\frac{19}{40}} = \frac{10}{19}$$

$$P(1/0) = \frac{\frac{1}{10}\cdot\frac{1}{4}}{\frac{21}{40}} = \frac{1}{21}$$

$$P(1/1) = \frac{\frac{9}{10}\cdot\frac{1}{4}}{\frac{19}{40}} = \frac{9}{19}.$$

The joint probabilities are obtained by relation (2) as

$$P(0,0)=\frac{2}{3}\cdot\frac{3}{4}=\frac{1}{2};P(0,1)=\frac{1}{3}\cdot\frac{3}{4}=\frac{1}{4};$$

$$P(1,0)=\frac{1}{10}\cdot\frac{1}{4}=\frac{1}{40};P(1,1)=\frac{9}{10}\cdot\frac{1}{4}=\frac{9}{40}.$$

119.4.4. Noiseless channels

A channel described by a channel matrix with one and only one non-zerox element in each column is called a noiseless channel.

A binary symmetric channel with $p=0$ or 1 is a noiseless channel. The channel represented by the following channel matrix is a noiseless channel,

$$\begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 3/5 & 2/5 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

119.5. Module Summary

In this module we define information and introduced the different components associated with communication system such as transmiter, encoder, decoder, channel, receiver, etc. The fundamental theorem of information is stated. Some basic terms such as memory less channel, channel matrix, probability relation in a channel and noiseless channel are defined. The module is ended with an exercise and references.

119.6. Self Assessment Questions

1. What do you mean by information? Explain with an example.
2. Write an essay on information theory emphasizing the basic concepts?
3. Draw a general structure of a communication system and explain it.
4. Explain the following terms:

   source, channel, encoder, decoder and receiver.
5. What do you mean by memoryless channel and channel matrix.

119.7. Suggested Further Readings

1. S.D. Sharma, Operations Research, Kedar Nath Ram Nath & Co., Meerut.
2. J.K. Sharma, Operations Research, Macmillan.
3. M.P. Gupta and J.K. Sharma, Operations Research for Management (2nd Ed.), National Publishing House, Delhi.
4. K. Swarup, P.K. Gupta and M. Mohan, Operations Research, Sultan Chand, Delhi.

---- 0 ----

M.Sc. Course
in
Applied Mathematics with Oceanology
and
Computer Programming

PART-II

Paper-X **Special Paper – OR**

Module No. - 120

Advanced Optimization and Operations Research-II

(Information Theory – II)

STRUCTURE :

120.1. Introduction

This is the continuation of the previous module (No. 119). Here we will discuss about the measurement of information and uncertainty using entropy. Several properties of entropy function are also studied here. Encoding procedure is also presented.

120.2. Objectives

After completion of this module you will able to:

- Explain the measurement of information
- Explain the measurement of uncertainty
- Explain entropy and its properties
- Explain encoding
- Explain encoding including Shannon-Fano encoding procedure.

120.3. Keywords

Measure of information, Measure of uncertainty, Entropy, Encoding.

120.4. A Measure of Information

An amount of information is virtually a search for statistical parameter associated with a probability scheme. This parameter should indicate a relative message of uncertainty applicable to the occurrence of each particular message in the set of messages.

Now, to obtain a formula for the amount of information, suppose there are n distinct models

$$m_1, m_2, \ldots\ldots, m_n$$

of a particular machine. The problem is to select a machine from this list.

The desired amount of information $I(m_k)$ associated with the selection of a particular model m_k must be a function of the probability of selecting m_k, i.e.,

$$I(m_k) = f(P(m_k)). \qquad \ldots\ldots (1)$$

Now, assuming for simplicity, that each of the model $m_1, m_2, \ldots\ldots, m_n$, is selected with an equal probability

$$P(m_1) = P(m_2) = P(m_3) = ... = P(m_n) = \frac{1}{n}. \qquad (2)$$

The equation (1) becomes

$$I_1(m_k) = f(1/n), \qquad (3)$$

means that the amount of information is the function of *n*.

Further, assume that each piece of the machine can be ordered in one of *m* distinct colours. Again, for simplicity, if selection of colours is also assumed to have equal probabilities, then the amount of information associated with the selection of a colour c_j among $\{c_1, c_2, c_3, ..., c_m\}$ is

$$I_2(c_j) = f(P(c_j)) = f(1/m) \qquad (4)$$

where the function *f* must be the same as used in equation (3).

Thus, the selection is done in two ways:

1. First select the machine and colour, the two selections being independent of each other. In this case,

$$I_1(m_k \text{ and } c_j) = I_1(m_k) + I_2(c_j)$$

or, $$I_1(m_k \text{ and } c_j) = f(1/n) + f(1/m). \qquad (5)$$

2. Alternatively, select the machine and its colour simultaneously as one selection from *mn* possible number of selections with equal probability.

Hence $$I(m_k \text{ and } c_j) = f\left(\frac{1}{mn}\right) \qquad (6)$$

The equations (5) and (6) give

$$f(1/n) + f(1/m) = f(1/mn) \qquad (7)$$

which is a functional equation.

The functional equation has one of the solutions given by

$$f(x) = \log(1/x) \text{ or } f(x) = -\log x. \qquad (8)$$

Substitute in equation (7) the values

$$f(1/n) = \log n, f(1/m) = \log m \text{ and } f(1/mn) = \log mn$$

to obtain the result

$$log\, n + log\, m = log\, mn \quad \text{.......} \quad (9)$$

which is always true.

Example 1. Suppose a baby has just been born at a neighbour's house and the question is asked "whether the baby is a boy or a girl?" The answer, "It is a boy", then gives a specific amount of information according to equation (8).

We assume that the baby was equally likely to have a boy or a girl, so the probability of its being a boyd is. $1/2$. Hence,

$$\text{amount of information} = -\log\frac{1}{2} = \log 2.$$

The numerical value of the amount of information in the above equation depends upon what logarithmic base is used. If 2 is used as the base of logarithm, then

$$\text{amount of information} = \log_2 2 = 1 \quad \text{.......} \quad (10)$$

which may be called a binary digit.

If 10 is used as the base of logarithm, a unit information may be called a decimal digit. In short this unit is called as a decit.

Example 2. In a certain community 25% of all girls are blondes and 75% of all blondes have blue eyes. Also, 50% of all girls in the community have blue eyes. If you know that a girl has blue eyes, how much additional information do you get by being informed that she is blonde?

Solution. Let

$p_1 = P$ (blonde) = probability that a girl is blonde in absence of knowledge of the colour in her eyes = 0.25

$p_2 = P$ (blue eyes/blonde) = probability that a blonde has blue eyes = 0.75

$p_3 = P$ (blue eyes) = probability that a girl has blue eyes in the absence of knowledge of the colour of her hair = 0.50

$p_4 = P$ (blonde, blue eyes) = probability that a girl is blonde and has blue eyes = $p_1 p_2$

$p_5 = P$ (blonde/ blue eyes) = probability that a blue-eyed girl is blude and has blue eyes.

Then $p_4 = p_1 p_2 = p_3 p_5$

or, $p_5 = \dfrac{p_1 p_2}{p_3}$.

If a girl has blue eyes, the additional information obtained by being informed that she is blonde, is

$$\log\frac{1}{p_5} = \log\left(\frac{p_3}{p_1 p_2}\right) = \log p_3 - \log p_1 - \log p_2$$

$$= -\log 2 + \log 4 - \log(3/4)$$

$= 1.42$ bits.

Example 3. An alphabet consists of 8 consonants and 8 vowels. Suppose all letters of the alphabet are equally probable and there is no inter-symbol influence. If consonants are always understood correctly, but vowels are understood correctly only half of the time being mistaken for other vowels the other half of the time all vowels being involved in errors the same percentage of the time. What is the average rate of information tramsmission.

Solution. According to conditions of the problem, 50% of the letters received will be correct consonants, 25% will be correct vowels, and 25% incorrect vowels. Now, first calculate the amount of information received in each of these cases.

Case I. Correct consonants

Here

p_A = probability before reception $= \frac{1}{16}$

p_B = probability after reception $= 1$.

Therefore, information/letter $= \log\left(\frac{1}{\frac{1}{16}}\right) = \log 16 = 4$ bits/ letter.

Case II. Correct Vowels

Here

p_A = probability before reception $= \frac{1}{16}$

p_B = probability after reception $= \frac{1}{2}$.

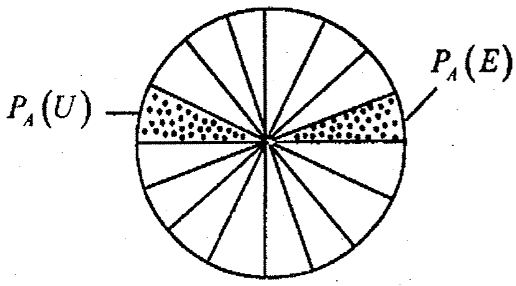
Therefore,

$$\text{information /letter} = \log\left(\frac{\frac{1}{2}}{\frac{1}{16}}\right) = 3 \text{ bits/ letter.}$$

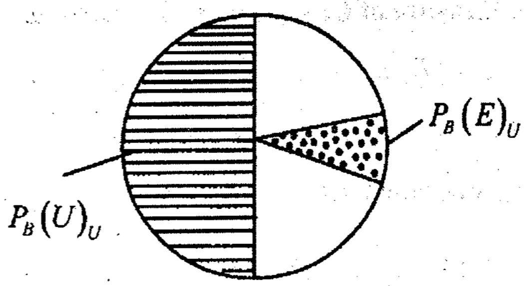
Case III. Incorrect Vowels

In this case,

$$p_A = \text{probability before reception} = \frac{1}{16}.$$



Probability before reception



Probability after reception

Now, find the probability after reception, suppose as E is sent and a U is received. The problem is to find the probability that an E was sent as a consequence of receiving a U. Before, reception the probability of E was $\frac{1}{16}$, since 16 letters have equal probabilities. After receiving a U, a vowel was sent, and the probability of a U is $\frac{1}{2}$. The other half of the probability is divided equally among other 7 vowels, one of which is E. Therefore, the probability that E was sent is $\frac{1}{7} \times \frac{1}{2} = \frac{1}{14}$.

Therefore,

$$p_B = P_B(E)_U = \text{probability offer reception} = \frac{1}{14}.$$

$$\text{Thus, information / letter} = \log\left(\frac{\frac{1}{14}}{\frac{1}{16}}\right)$$

$$= \log 16 - \log 14 = 4 - 3.8$$

$= 0.2$ bits/ letter.

It is interesting to note that the information received with an incorrect vowel is positive in this case. Since the reception of the incorrect vowel actually increases the probability of trnasmission of the correct vowel.

Finally, for the average information per symbol, take 50% of Case I and 25% each of case II and III. Therefore,

average information/ symbol $= 0.5 \times 4 + 0.25 \times 3 + 0.25 \times 0.2$

$= 2.8$ bits/ symbol.

120.5. A Measure of Uncertainty : Entropy

Let $S = \{E_1, E_2, ..., E_n\}$ be the sample space, where $E_1, E_2, ..., E_n$ are mutually exclusive events with probabilities $p_1, p_2, ..., p_n$ respectively. We assume that $p_1, p_2, ..., p_n$ are known.

It may be noted that

$$\bigcup_{k=1}^{n} E_k = S \text{ and } \sum_{k=1}^{n} p_k = 1. \qquad (11)$$

The fundamental problem of interest is to associate a measure of uncertainty $H(p_1, p_2, ..., p_n)$, with such probability schemes.

Shannon and Wiener have suggested the following measure of uncertainty associated with the sample space of a complete finite scheme,

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i, \qquad (12)$$

where the random variable X is defined over the sample space of events S, and events satisfy the equation (11).

The average amount of information or entropy of a finite complete probability scheme is defined by

$$H(X) = H(\bar{E}_k) = -\sum_{k=1}^{n} p_k \log p_k. \qquad (13)$$

120.6. Properties of Entropy Function (H)

The entropy function $H(p_1, p_2, ..., p_n)$, follows some basic properties which are presented here.

1. **(Continuity property):**

The entropy function $H(p_1, p_2, ..., p_n)$, is continuous for each $p_k, 0 \le p_k \le 1$.

Proof. We know,

$$-H(p_1, p_2, ..., p_n) = p_1 \log p_1 + p_2 \log p_2 + ... + p_n \log p_n$$

$$= p_1 \log p_1 + p_2 \log p_2 + ... + p_{n-1} \log p_{n-1}$$

$$+ (1 - p_1 - p_2 - ..... - p_{n-1}) \log(1 - p_1 - p_2 - ..... - p_{n-1})$$

as $\sum_{k=1}^{n} p_k = 1.$

It is seen that all independent variables $p_1, p_2, ..., p_{n-1}$ and $(1 - p_1 - p_2 - ..... - p_{n-1})$ are continuous in (0,1) and that the logarithm of a continuous function is continuous. Hence H is continuous.

2. **(Symmetric property):**

The entropy function H is symmetric, i.e.,

$H(p_k, 1 - p_k) = H(1 - p_k, p_k)$, where $k = 1, 2, ..., n$.

Proof. It follows from definition.

3. **(Extremal property):**

The maximum value of $H(p_1, p_2, ..., p_n)$ is $H\left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right)$.

Proof. The complete proof is give in Module 109 (dynamic programming).

In that module we have seen that $H = -\sum p_i \log p_i$ is maximum when $p_1 = p_2 = ... = p_n = 1/n$.

Hence max $H(p_1, p_2, ..., p_n) = H\left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right)$.

4. **(Additive property)**

Let x_n be a particular event with probability p_n is divided into m mutually exclusive subsets, say, $E_1, E_2, ..., E_m$ with probabilities $q_1, q_2, ..., q_m$ respectively such that $p_n = q_1 + q_2 + ... + q_m$, then

$$H(p_1, p_2, ..., p_{n-1}, q_1, q_2, ..., q_m)$$

$$= H(p_1, p_2, ..., p_{n-1}, p_n) + p_n H(q_1/p_n, q_2/p_n, ..., q_m/p_n).$$

Proof. By definition of H,

$$H(p_1, p_2, ..., p_{n-1}, q_1, q_2, ..., q_m)$$

$$= -\sum_{i=1}^{n-1} p_i \log p_i - \sum_{i=1}^{m} p_i \log q_i.$$

$$= -\left\{\sum_{i=1}^{n} p_i \log p_i - p_n \log p_n\right\} - \sum_{i=1}^{m} q_i \log q_i$$

$$= H(p_1, p_2, \ldots, p_n) + \left\{p_n \log p_n - \sum_{i=1}^{m} q_i \log q_i\right\}$$

since $H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i$.

But, $p_n \log p_n - \sum_{i=1}^{m} q_i \log q_i$

$$= p_n\left\{\frac{p_n}{p_n} \log p_n\right\} - p_n \sum_{i=1}^{m}\left\{\frac{q_i}{p_n} \log q_i\right\}$$

$$= \sum_{i=1}^{m} q_i \left\{\frac{p_n}{p_n} \log p_n\right\} - p_n \sum_{i=1}^{m}\left\{\frac{q_i}{p_n} \log q_i\right\}$$

$$= p_n \sum_{i=1}^{m} \frac{q_i}{p_n} \log p_n - p_n \sum_{i=1}^{m} \frac{q_i}{p_n} \log q_i$$

$$= -p_n \sum_{i=1}^{m} \frac{q_i}{p_n} (\log q_i - \log p_n)$$

$$= -p_n \sum_{i=1}^{m} \frac{q_i}{p_n} \log\left(\frac{q_i}{p_n}\right) \qquad \left[\because p_n = \sum_{i=1}^{m} q_i\right].$$

Thus,

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m)$$

$$= H(p_1, p_2, \ldots, p_n) + \left\{p_n \log p_n - \sum_{i=1}^{m} q_i \log q_i\right\}$$

$$= H(p_1, p_2, \ldots, p_n) - p_n \sum_{i=1}^{n} \frac{q_i}{p_n} \log\left(\frac{q_i}{p_n}\right)$$

$$= H(p_1, p_2, \ldots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_m}{p_n}\right).$$

Example 1. Evaluate the average uncertainty associated with the sample space of events A, B and C which are mutually exclusive with probability distribution

Event:	A	B	C
Probability:	1/5	4/15	8/15

Solution. From the date of the problem, we have

$p_1 = 1/5, p_2 = 4/15$ and $p_3 = 8/15$.

The entropy function H is defined as

$$= H(p_1, p_2, ..., p_n) = -\sum_{i=1}^{n} p_i \log p_i$$

where p_i's are the probabilities associated with the given probability distribution. For this problem n=3.

$$H(1/5, 4/15, 8/15) = p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3$$

$$= -\frac{1}{5}\log\left(\frac{1}{5}\right) - \frac{4}{15}\log\left(\frac{4}{15}\right) - \frac{8}{15}\log\left(\frac{8}{15}\right)$$

$$= -\frac{1}{5}\left[3\log\frac{1}{5} + 4\log\left(\frac{4}{15}\right) + 8\log\left(\frac{8}{15}\right)\right]$$

$$= -\frac{1}{5}\left[-3\log 5 + 4(\log 4 - \log 15) + 8(\log 8 - \log 15)\right]$$

$$= -\frac{1}{15}\left[-3\log 5 - 4\log(3\times 5) - 8\log(3\times 5) + 4\log 2^2 + 8\log 2^3\right]$$

$$= \frac{1}{15}[15\log 5 + 12\log 3 - 32]\ [\because \log 2 = 1]$$

$$= \log 5 + \frac{4}{5}\log 3 - \frac{32}{15}.$$

Example 2. Show that the entropy of the following probability distribution is

Event:	x_1	x_2 ...... x_i .... x_{n-1}	x_n
Probability:	$\frac{1}{2}$	$\frac{1}{2^2}$ .... $\frac{1}{2^i}$ ...... $\frac{1}{2^{n-1}}$	$\frac{1}{2^{n-1}}$

Solution. We have $p_i = \frac{1}{2^i}, i = 1, 2, ..., n-1$ and $p_n = \frac{1}{2^{n-1}}$

such that $\sum_{i=1}^{n} p_i = 1$.

The entropy function H is defined as

$$H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i = -\sum_{i=1}^{n-1} p_i \log p_i - p_n \log p_n$$

$$= -\sum_{i=1}^{n-1} \left(\frac{1}{2^i}\right) \log\left(\frac{1}{2^i}\right) - \left(\frac{1}{2^{n-1}}\right) \log\left(\frac{1}{2^{n-1}}\right)$$

$$= \sum_{i=1}^{n-1} \left(\frac{1}{2^i}\right) \log 2^i + \left(\frac{1}{2^{n-1}}\right) \log\left(2^{n-1}\right)$$

$$= \sum_{i=1}^{n-1} i\left(\frac{1}{2^i}\right) + (n-1)\frac{1}{2^{n-1}} \quad \left[\because \log^2 = 1\right]$$

$$= \left\{\frac{1}{2} + \frac{2}{2^2} + \frac{3}{3^2} + \ldots + \frac{n-1}{2^{n-1}}\right\} + \frac{n-1}{2^{n-1}} \quad \ldots\ldots (i)$$

or, $$\frac{1}{2} H(p_1, p_2, \ldots, p_n) = \left\{\frac{1}{2^2} + \frac{2}{2^3} + \frac{3}{2^4} + \ldots + \frac{n-1}{2^n}\right\} + \frac{n-1}{2^n} \quad \ldots\ldots (ii)$$

Subtracting (*ii*) from (*i*), we get

$$H(p_1, p_2, \ldots, p_n) - \frac{1}{2} H(p_1, p_2, \ldots, p_n)$$

$$= \left\{\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots + \frac{1}{2^{n-1}}\right\} + \left\{\frac{n-1}{2^{n-1}} - \frac{2(n-1)}{2^n}\right\}$$

or, $$\frac{1}{2} H(p_1, p_2, \ldots, p_n) = \left\{\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots + \frac{1}{2^{n-1}}\right\} = 1 - \left(\frac{1}{2}\right)^{n-1}$$

or, $$H(p_1, p_2, \ldots, p_n) = 2 - \left(\frac{1}{2}\right)^{n-2}$$

Theorem 1. Let $p_1, p_2, \ldots, p_m$ and $q_1, q_2, \ldots, q_m$ be arbitrary non-negative numbers with $\sum_{i=1}^{m} p_i = \sum_{j=1}^{m} q_j$. Then

$$-\sum_{i=1}^{m} p_i \log p_i \le -\sum_{i=1}^{m} p_i \log q_i,$$

with equality if and only if $p_i = q_i$ for all i.

Proof. Since the logaritam is a convex function, we have the inequality $\log x \le x - 1$ with equality iff $x=1$.

If we take $x = q_i/p_i$, then

$$\log(q_i/p_i) \le (q_i/p_i) - 1 \quad \text{.......... } (i)$$

with equality if and only if $q_i = p_i$.

Multiplying (*i*) by p_i and summing over *i*, we get

$$\sum_{i=1}^{m} p_i \log(q_i/p_i) \le \sum_{i=1}^{m} (q_i - p_i) = 1 - 1 = 0,$$

with equality iff $q_i = p_i$ for all *i*.

This proves that

$$\sum_{i=1}^{m} p_i \log q_i \le \sum_{i=1}^{m} p_i \log p_i$$

or, $-\sum_{i=1}^{m} p_i \log p_i \le -\sum_{i=1}^{m} p_i \log q_i$ with equality iff $p_i = q_i$ for all *i*.

120.7. Joint and Conditional Entropies

Let us consider two sets of messages

$$X = \{x_1, x_2, ..., x_m\}, Y = \{y_1, y_2, ..., y_m\},$$

where x_i 's are the messages sent (channel input) and y_j 's are the messages received (channel output).

Let $p_{ij} = P(X = x_i, Y = y_j)$, $i = 1, 2, ..., m; j = 1, 2, ..., n$; denote the probability of the joint event that message x_i is sent and message y_j is received.

The marginal probability distributins of *X* and *Y* are given by

$$p_{io} = \sum_{j=1}^{n} p_{ij} \text{ and } p_{oj} = \sum_{i=1}^{m} p_{ij} \text{ for all } i, j.$$

Then, obviously the <u>marginal entropies</u> of the two marginal distributions are given by

$$H(X) = -\sum_{i=1}^{m} p_{io} \log p_{io} \text{ and } H(Y) = -\sum_{j=1}^{n} p_{oj} \log p_{oj}.$$

The entropy $H(X)$ measures the uncertainty of the messages sent and $H(Y)$ performs the same role for the message received.

The joint entropy is the entropy of the joint distribution of the messages sent and received, and is therefor given by

$$H(X,Y) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log p_{ij}.$$

It may be observed that

$$\max H(X,Y) = \log(mn) = \log m + \log n$$
$$= \max H(X) + \max H(Y).$$

Theorem 1. $H(X,Y) \le H(X) + H(Y)$, with equality iff *X* and *Y* are independent.

Solution.

By definition,

$$H(X) + H(Y) = -\sum_{i=1}^{m} p_{io} \log p_{io} - \sum_{j=1}^{n} p_{oj} \log p_{oj}$$

$$= -\sum_{i=1}^{m}\left(\sum_{j=1}^{n} p_{ij}\right)\log p_{io} - \sum_{j=1}^{n}\left(\sum_{j=1}^{m} p_{ij}\right)\log p_{oj}$$

$$= -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log\left(p_{io}p_{oj}\right)$$

$$= -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log q_{ij}$$

where $q_{ij} = p_{io}p_{oj}$

Again, by definition

$$H(X,Y) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log p_{ij}.$$

Also, we know

$$-\sum_{i=1}^{n} p_i \log p_i \le -\sum_{i=1}^{n} p_i \log q_i.$$

Hence $H(X,Y) \le H(X) + H(Y)$.

The equality holds when $p_{ij} = q_{ij}$ for all *i*,*j*. The condition for equality reduces to $p_{io}p_{oj} = p_{ij}$ implies *X* and *Y* are independent.

Conditional Entropies

Consider two finite discrete sample spaces S_1 and S_2 and their product space $S(=S_1 \times S_2)$. Also, let $E_1, E_2, ..., E_n$ and $F_1, F_2, ..., F_m$ constitute the spaces S_1 and S_2.

Thus, the complete sets of events in the space $S_1 \times S_2$ is given by

$$EF = \begin{bmatrix} E_1F_1 & E_1F_2 & E_1F_3 & \cdots & E_1F_m \\ E_2F_1 & E_2F_2 & E_2F_3 & \cdots & E_2F_m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ E_nF_1 & E_nF_2 & E_nF_3 & \cdots & E_nF_m \end{bmatrix}$$

For example, an event F_r may occur in conjunction with $E_1, E_2, ...,$ or E_n.

$$F_r = \sum_{k=1}^{n} E_k E_r$$

$$P(X = x_k / Y = y_k) = \frac{P(X = x_r \cap Y = y_r)}{P(Y = y_r)}$$

or, $$P(x_k / y_r) = \frac{P(k,r)}{P(y_r)}$$

Now, we consider the following probability scheme

$$(E/F_r) = [E_1/E_r, E_2/E_r, E_3/E_r, ..., E_r/E_r]$$

$$P(E/F_r) = \left[\frac{P(1,r)}{P(y_r)}, \frac{P(2,r)}{P(y_r)}, \frac{P(3,r)}{P(y_r)}, ..., \frac{P(n,r)}{P(y_r)}\right]$$

The probability scheme is not only finite but also complete because the sum of elements of this matrix is unity. Therefore, an entropy will be given by

$$H(X/y_r) = -\sum_{k=1}^{n} \frac{P(k,r)}{P(y_r)} \log \frac{P(k,r)}{P(y_r)}$$

$$= -\sum_{k=1}^{n} P(x_k/y_r) \log P(x_k/y_r).$$

Now, take the average of this conditional entropy for all admissible values of y_r, so that a measure of average conditional entropy of the system can be obtained.

$$H(X/Y) = \overline{H(X/y_r)} = \sum_{r=1}^{m} P(y_r) \cdot H(X/y_r)$$

$$= -\sum_{r=1}^{m} P(y_r) \sum_{k=1}^{m} P(x_k/y_r) \log P(x_k/y_r)$$

$$H(X/Y) = -\sum_{r=1}^{m} \sum_{k=1}^{n} P(y_r) P(x_k/y_r) \log (x_k/y_r).$$

Similarly, it is possible to obtain the expression for the average conditional entropy $H(Y/X)$, i.e.

$$H(X/Y) = -\sum_{k=1}^{n} \sum_{r=1}^{m} P(x_k) P(y_r/x_k) \log P(y_r/x_k).$$

Thus, two conditional entropies may be expressed as

$$H(X/Y) = -\sum_{r=1}^{m} \sum_{k=1}^{n} P(x_k/y_r) \log P(x_k/y_r)$$

and $H(Y/X) = -\sum_{k=1}^{n} \sum_{r=1}^{m} P(x_k/y_r) \log P(y_r/x_k).$

Theorem 2. Prove that

$$H(X,Y) = H(X/Y) + H(Y) = H(Y/X) + H(X),$$

where $H(X) \geq H(X/Y)$

Proof. We know,

$$H(X) = -\sum_{k=1}^{n} P(x_k) \log P(x_k) \quad \text{...(i)}$$

$$H(Y) = -\sum_{r=1}^{m} P(y_r) \log P(y_r) \quad \text{...(ii)}$$

$$H(X,Y) = -\sum_{k=1}^{n} \sum_{r=1}^{m} P(k,r) \log P(k,r) \quad \text{...(iii)}$$

$$H(X/Y) = -\sum_{r=1}^{m} \sum_{k=1}^{n} P(x_k, y_r) \log P(x_k/y_r) \quad \text{...(iv)}$$

$$H(Y/X) = -\sum_{k=1}^{n} \sum_{r=1}^{m} P(x_k, y_r) \log P(y_r/x_k). \quad \text{...(v)}$$

Again, $P(k,r) = P(x_k, y_r) = P(x_k / y_r) P(y_r) = P(y_r / x_k) P(x_k)$ ...(vi)

and $\log P(x_k, y_r) = \log P(x_k / y_r) + \log P(y_r)$ ...(vii)

$= \log P(y_r / x_k) + \log P(x_k).$ ...(viii)

Now, $H(X/Y) = -\sum_{r=1}^{m}\sum_{k=1}^{n} P(x_k, y_r)\log P(x_k/y_r)$

$$= -\sum_{r=1}^{m}\sum_{k=1}^{n} P(x_k, y_r)\log\frac{P(x_k, y_r)}{P(y_r)}$$

$$= -\sum_{r=1}^{m}\sum_{k=1}^{n} P(x_k, y_r)\log P(x_k, y_r) + \sum_{r=1}^{m}\sum_{k=1}^{n} P(x_k, y_r)\log P(y_r)$$

$$= H(X,Y) + \sum_{r=1}^{m} P(y_r)\log P(y_r)$$

$$\left[\because \sum_{r=1}^{m}\sum_{k=1}^{n} P(x_k, y_r) = \sum_{r=1}^{m}\sum_{k=1}^{n} p_{kr} = \sum_{r=1}^{m}\left(\sum_{k=1}^{n} p_{kr}\right) = \sum_{r=1}^{m} p_{or} = \sum_{r=1}^{m} p_r = \sum_{r=1}^{m} P(y_r).\right]$$

$$= H(X,Y) - H(Y).$$

Hence $H(X/Y) + H(Y) = H(X,Y)$.

Similarly, $H(Y/X) + H(X) = H(X,Y)$.

<u>To prove $H(X/Y) + H(X) \le 0$.</u>

$$H(X/Y) - H(X) = \sum_{r=1}^{m}\sum_{k=1}^{n} P(x_k, y_r)\log\frac{P(x_k)}{P(x_k/y_r)}$$

$$\le \sum_{r=1}^{m}\sum_{k=1}^{n} P(x_k, y_r)\left[\frac{P(x_k)}{P(x_k/y_r)} - 1\right]\log e$$

[since $\log x \le (x-1)\log e$]

But,

$$\sum_{r=1}^{m}\sum_{k=1}^{n}\{P(x_k)P(y_r) + P(x_k/y_r)\}\log e$$

$$= \sum_{r=1}^{m}[P(y_r) - P(y_r)]\log e = 0.$$

Hence, $H(X/Y) - H(X) \le 0$ or $H(X) \ge H(X/Y)$.

Similarly, it can be shown that

$$H(Y) \ge H(Y/X).$$

Example 1. A transmitter has a character consisting of five letters $(x_1, x_2, x_3, x_4, x_5)$ and the receiver has a character consisting of four letters (y_1, y_2, y_3, y_4). The joint probability for the communication is given below:

$P(x_i/y_j)$	y_1	y_2	y_3	y_4	$P(x_i)$
x_1	0.25	0	0	0	0.25
x_2	0.10	0.30	0	0	0.40
x_3	0	0.05	0.10	0	0.15
x_4	0	0	0.05	0.10	0.15
x_5	0.	0	0.05	0	0.15
$P(y_j)$	0.35	0.35	0.20	0.10	0.05

(a) Determine the different entropies for the channel, assume $0 \log 0 = 0$.

(b) Determine $H(X)$, $H(Y)$, $H(X, Y)$ and $H(Y/X)$.

Solution. To determine different entropies for the channel, joint probabilities are to be calculated for *i* and *j*. Then different marginal and conditional probabilities may be calculated with the help of joint probabilities as given below:

$$P(x_1) = 0.25 + 0.0 = 0.25$$

$$P(x_2) = 0.10 + 0.30 = 0.40$$

$$P(x_3) = 0.05 + 0.0 = 0.05$$

$$P(x_4) = 0.05 + 0.10 = 0.15$$

$$P(x_5) = 0.05 + 0.0 = 0.05$$

$$P(y_1) = 0.35, P(y_2) = 0.35, P(y_3) = 0.20, P(y_4) = 0.10.$$

The conditional probabilities $P(x_i/y_j)$ are shown below.

$P(x_i/y_j)$	y_1	y_2	y_3	y_4
x_1	1	0	0	0
x_2	0.25	0.75	0	0
x_3	0	0.33	0.66	0
x_4	0	0	0.33	0.66
x_5	0	0	1	0

Marginal entropies

$$H(X) = -\sum_{i=1}^{5} P(x_i)\log P(x_i)$$

$$= -(0.25)\,log\,(0.25) - (0.40)\,log\,(0.40) - (0.15)\,log\,(0.15) - (0.15)\,log\,(0.15) - (0.05)\,log\,(0.05)$$

$$= \frac{1}{4}\log 4 + \frac{2}{5}\log\left(\frac{5}{2}\right) + \frac{3}{20}\log\left(\frac{20}{3}\right) + \frac{1}{20}\log 20$$

= 1.326 bits.

$$H(Y) = -\sum_{j=1}^{4} P(y_j)\log P(y_j)$$

$$= -(0.35)\,log\,(0.35) - (0.35)\,log\,(0.35) - (0.20)\,log\,(0.20) - (0.10)\,log\,(0.10)$$

$$= \frac{7}{10}\log\left(\frac{20}{7}\right) + \frac{1}{5}\log 5 + \frac{1}{10}\log 10$$

= 1.855 bits.

Conditional entropies

$$H(X/Y) = -\sum_{i=1}^{5}\sum_{j=1}^{4} P(x_i, y_j)\log P(x_i/y_j)$$

$$= -\left\{(0.25)\log\left(\frac{0.25}{0.35}\right) - 0.10\log\left[\left(\frac{0.10}{0.30}\right)\right] + 0.30\log\left(\frac{0.30}{0.35}\right)\right.$$

$$\left. + 0.05\log\left(\frac{0.05}{0.35}\right) + 0.10\log\left(\frac{0.10}{0.20}\right) + 0.05\log\left(\frac{0.05}{0.20}\right) + 0.05\log\left(\frac{0.05}{0.20}\right) + 0.10\log\left(\frac{0.10}{0.10}\right)\right\}$$

$$= -\left\{\frac{1}{4}\log\frac{5}{6} + \frac{1}{10}\log\frac{1}{3} + \frac{3}{10}\log\frac{6}{7} + \frac{1}{20}\log\frac{1}{7}\right.$$

$$\left. + \frac{1}{10}\log\frac{1}{2} + \frac{1}{20}\log\frac{1}{4} + \frac{1}{20}\log\frac{1}{4} + \frac{1}{10}\log 1\right\}$$

= 0.0704 bits.

Hence $H(Y/X) = H(Y) + H(X/Y) - H(X)$

= 1.855 + 0.0704 – 1.326 = 0.599 bits.

Joint entropy

$$H(X,Y) = H(X) + H(Y/X) = 1.326 + 0.559 = 1.925.$$

120.8. Expected Mutual Information

The mutual information of the meassage transmitted x_i and the message received y_j is defined as follows:

$$h(x_i, y_j) = \frac{P(y_j/x_i)}{P(y_j)} = \log\left\{\frac{P(x_i, y_j)}{P(x_i)P(y_j)}\right\};$$

$i = 1, 2, ..., m; j = 1, 2, ..., n.$

The following two cases may arise:

(a) If no information is provided by one event about the other, then

$$h(x_i, y_j) = 0 \text{ for all } i \text{ and } j.$$

In other words, the events X and Y are independent, in which case one event cannot give information about the other.

(b) If $h(x_i, y_j) > 0$ or <0, for a fixed x_i, then more or less information is received from the transmission of the message x_i given that message y_j has already been received. This is different from the independence pattern of a set of messages tramsitted $X = \{x_1, x_2, ..., x_m\}$ and set of messages received $Y = \{y_1, y_2, ..., y_n\}$.

Now, the expected mutual information or the averages amount of information about X that is provided by the occurrence of Y event may be defined as :

$$I(X \leftarrow Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j) h(x_i, y_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j) \log \frac{P(x_i / y_j)}{P(x_i)}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j) \log\left[\frac{P(y_j/x_i)P(x_i)}{P(y_j)P(x_i)}\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} P(x_i, y_j) \log\left[\frac{P(y_j/x_i)}{P(y_j)}\right]$$

$$= I(Y \leftarrow X).$$

This shows that the average amount of information that an event Y provides about the occurrence of an event X is equal to the average amount of information that an event X provides about the occurrence of an event Y.

It may be noted that the expected mutual information is always non-negative. Thus, on the average mutual information is not misleading, the knowledge of the occurrence of an event in one set will on the average provide information about the occurrence of an event in the other set.

Expected mutual information measures will be zero if and only if X and Y are independent.

Theorem 1.

(a) $I(X,Y)=H(X)-H(X/Y)=H(Y)-H(Y/X)$

(b) $I(X,Y)=H(X)+H(Y)-H(X,Y)$.

Proof. We know that

$$H(X)-H(X/Y)=-\sum_{i=1}^{m}\sum_{j=1}^{n}P(x_i,y_j)\log P(x_i)+\sum_{i=1}^{m}\sum_{j=1}^{n}P(x_i,y_j)\log P(x_i/y_j)$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{n}P(x_i,y_j)\log\frac{P(x_i/y_j)}{P(x_i)}$$

$$=I(X,Y).$$

Similarly, it can also be proved that

$$I(X,Y)=H(Y)-H(Y/X).$$

(b) $I(X,Y)=H(X)-H(X/Y)$

$$=H(X)-\{H(X,Y)-H(Y)\}$$

$$=H(X)+H(Y)-H(X,Y).$$

120.9. Axiom for Entropy Function

Axiom 1. The entropy function takes its maximum value when all the events have equal probabilities.

That is,

$$\max H(p_1,p_2,\ldots,p_n)=H\left(\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n}\right)$$

where $p_1=p_2=\ldots=p_n=\frac{1}{n}$.

Axiom 2. The information provided by the joint occurrence of the pair (X, Y) is equal to the sum of the information provided by the occurrence of X and that provided by the occurrence of Y given that X has already occurred, i.e.,

$$H(X,Y)=H(X)+H(Y/X).$$

Axiom 3. The entropy of the scheme remains unchanged with the addition of an impossible event in that scheme, i.e.,

$$H(p_1, p_2, ..., p_n, 0) = H(p_1, p_2, ..., p_n).$$

Axiom 4. The entropy is continuous with respect to all its arguments.

120.10. Channel Capacity, Efficiency and Redundancy

Channel Capacity

The amount of average mutual information processed by the channel in a communication system is defined by

$$I(X,Y) = H(X/Y).$$

As the information processed by a channel depends upon the input probability distribution $P(x_i)$ of x_i's therefore, it can be varied until the maximum of $I(X,Y)$ is reached. Hence channel capacity, say C, can be defined as

$$C = \max I(X,Y) = \max\{H(X) - H(X/Y)\}$$

for all $P(x_i)$.

But, for a noise free channel, we have

$$I(X,Y) = H(X) = H(Y)$$

and $I(X,Y) = H(X,Y)$.

Therefore, channel capacity C in this case may be redefined as

$$C = \max I(X,Y) = \max H(X)$$

$$= \max\left[-\sum_{i=1}^{m} P(x_i)\log P(x_i)\right]$$

$$= -\log(1/n)$$

$$= \log n \text{ bits/symbol.}$$

This is due to the fact that maximum of $H(X)$ occurs when $P(x_1) = P(x_2) = ... = P(x_n)$.

Special types of channels

(a) Binary symmetric channel. The channel capacity for the channel matrix

$$P = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

is given by $C = 1 - H(p, 1-p)$.

(b) **Channels with non-singular channel matrix.**

For a square and non-singular channel P, the channel capacity is defined as:

$$C = \log_2 \sum_{j=1}^{n} \exp\left[-\sum_{i=1}^{n} a_{jk} H(Y/X = x_k)\right]$$

where a_{jk} is the (j, k)th element of p^{-1}.

Efficiency

The efficiency of the noise-free system is given by, $\eta = \dfrac{H(X)}{L}$, where L is the length of the code.

Redundancy

The difference between the expected or average mutual information $I(X, Y)$ and its maximum value is defined as the redundancy of the communication system. The ratio of given redundancy to channel capacity is know as relative redundancy, i.e.,

Redundancy for noise-free channel

$$\beta = C - I(X, Y) = \log n - H(X).$$

Relative redundancy for noise-free channel β

$$= \frac{\log n - H(x)}{\log n} = 1 - \frac{H(x)}{\log n}$$

= 1 – efficiency of the system.

Example 1. Find the capacity of the memory less channel specified by the channel matrix

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

Solution. The capacity of the memory less channel is given by

$$C = \max I(X,Y) = \max\{H(X)+H(Y)-H(X,Y)\}$$

$$= -\sum_{i=1}^{4} P(x_i,y_j)\log P(x_i,y_j);\ j = 1, 2, 3, 4$$

where

$$P(x_i,y_j) = (1/2,1/4,1/4,0)$$

$$P(x_i,y_2) = (1/4,1/4,1/4,1/4)$$

$$P(x_i,y_3) = (0,0,1,0)$$

$$P(x_i,y_4) = (1/2,0,0,1/2).$$

Thus $C = \frac{1}{2}\log\frac{1}{2} + 2\left(\frac{1}{4}\log\frac{1}{4}\right) + 4\left(\frac{1}{4}\log\frac{1}{4}\right) + 1\log 1 + 2\left(\frac{1}{2}\log\frac{1}{2}\right)$

$$= \frac{3}{2}\log 2 + 3\log 2 = \frac{9}{2} \text{ bits/ symbol.}$$

120.11. Encoding

Encoding may be defined as a transformation procedure of a message from sources to receiver through a noiseless channel in some code language. In other words, if $X = \{x_1, x_2, x...., x_m\}$ be the set of messages to be transmitted then codes may be defined as a relationship between all possible sequences of symbols of the set X with another set $Y = \{y_1, y_2, ..., y_n\}$ of code character of alphabet.

Objectives of encoding

(a) It is used to increase the efficiency of transmission.

(b) It is used to minimize the expected code word length. If the code word associated with x_i is of length l_i, i=1, 2, ...,m; then the expected length of messages is given by

$$L = \sum_{i=1}^{m} l_i P(x_i).$$

If L is minimum then the transmission is called efficient transmission.

(c) It is used to minimize the cost of transmission. If $C_i\ (i = 1, 2, ..., m)$ is the cost of transmission of some words

W_i with probability of transmission $P(W_i)$, then in a message of m words, the expected cost per message is given by

$$M = \sum_{i=1}^{m} C_i P(W_i).$$

The transmission for which M is minimum is considered to be efficient.

The coding are of different kinds, some of them are discussed below:

Block Code

A code which make a relationship with each of the symbols of the set X to a fixed sequence of symbols of the set Y is called a block code. That is, each symbol x_i $(i = 1, 2, \ldots, m)$ is to be assigned a fixed sequence of symbols of Y called the code words, associated with x_i.

For example, x_1 may correspond to y_1, y_2 and x_2 may correspond to y_7, y_8, y_9.

Bindary Code

In particular, if the set $X = \{0,1\}$, then a block code is said to be binary code. An example of binary code is

$$x_1 \to 01,\ x_2 \to 010,\ x_3 \to 10,\ x_4 \to 1101.$$

Non-singular Code

A block code is said to be non-singular code if all words of the code are distinct. An example of non-singular code is

$$x_1 \to 00,\ x_2 \to 01,\ x_3 \to 10,\ x_4 \to 11.$$

Uniquely decodable (separable) code

A code is said to be uniquely decodable (separable) code if every finite sequence of symbols of the set Y is associated to at most one symbol of the set X. Examples of this code are given below:

(a) $x_1 \to 0,\ x_2 \to 10,\ x_2 \to 110,\ x_3 \to 111$

(b) $x_1 \to 0,\ x_2 \to 01,\ x_3 \to 011,\ x_3 \to 0111.$

120.12. Shannon-Fano Encoding Procedure

In this method, a sequence of binary numbers {0, 1} is used for encoding messages through a memoryless communication channel. Let $X = \{x_1, x_2, \ldots, x_m\}$ be the list of the messages to be transmitted from some source and $P = \{p_1, p_2, \ldots, p_m\}$ be their corresponding probabilities. Our aim is to devise an encoding procedure so that a sequence of binary numbers $(0,1)$ of unspecified length can be associated to each meassage x_i. The sequence obtained must satisfy the following conditions:

(a) No sequence of binary numbers can be obtained from any other sequence by adding additional binary terms to sequences of shorter lengths.

(b) Binary numbers associated with each meassage x_i to form a sequence occur independently with equal probability.

The algorithm of this procedure is given below:

Step 1. Arrange the messages (words) $x_1, x_2, ..., x_m$ in descending order in terms of their probabilities. Without loss of generality, let

$$p_1 > p_2 > .... > p_m$$

so that, we have

Message :	x_1	x_2	$x_3 .... x_i .... x_m$
Probability :	p_1	p_2	$p_3 .... p_i .... p_m$.

Step 2. Divide the set of messages $X = \{x_1, x_2, ..., x_m\}$ into two subsets, say X_1 and X_2 of equal probabilities i.e.,

Set	Message	Probabilities
X_1	x_1, x_2	$P(X_1) = p_1 + p_2$
X_2	$x_3, x_4, ..., x_m$	$P(X_2) = p_3 + ... + p_m$

such that $P(X_1) = P(X_2)$.

Step 3. Again, divide both subsets X_1 and X_2 into two subsets, say X_{11}, X_{12} and X_{21}, X_{22} with equal probabilities respectively.

Step 4. Assign binary number 0 to the first position of the coded word in each message in subset X_1 and binary number 1 to the first position of the coded word in each message in subset X_2. The similar procedure of assigning binary number 0 and 1 must be repeated for subsets of X_1 and X_2.

Step 5. The division and assigning binary digits 0 and 1 continue till each subset contains only one message (word).

Example 1. A source memory has six characters with the following probabilities of transmission:

A	B	C	D	E	F
$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{2}$

Devise the Shannon-Fano encoding procedure to obtain uniquely decodable code to the above message ensemble. What is the average length, efficiency and redundancy of the code that you obtain.

Solution.

Step 1. The ensembled message are already in descending order of probabilities.

Step 2. Divide the elements of the set X into two subsets X_1 and X_2 with approximately equal probabilities as shown below

$$X_1 = \{A, B\} \text{ and } X_2 = \{C, D, E, F\},$$

with probabilities

$$P(X_1) = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$

$$P(X_2) = \frac{1}{8} + \frac{1}{8} + \frac{1}{12} + \frac{1}{12} = \frac{5}{12}.$$

Step 3. Further divide the set X_2 into two subsets of equal probabilities as given below:

Subsets	**Probabilities**
$X_{21} = \{C, D\} = \{1/8, 1/8\}$	$P(X_{21}) = 1/4$
$X_{22} = \{E, F\} = \{1/12, 1/12\}$	$P(X_{22}) = 1/16.$

Step 4. Assign binary number 0 and 1 to first position of all code words in X_1 and X_2 respectively as shown below:

Character	**Probabilities (p)**	**Partitioning**	**Code word**	**Code word lenght (l)**
A } X_1	$\frac{1}{3}$	X_{11}	00	2
B } X_1	$\frac{1}{4}$	X_{12}	10	2
C } X_2	$\frac{1}{8}$ } X_{21}	X_{211}	100	3
D } X_2	$\frac{1}{8}$ } X_{21}	X_{212}	101	3
E } X_2	$\frac{1}{12}$ } X_{22}	X_{221}	110	3
F } X_2	$\frac{1}{12}$ } X_{22}	X_{222}	111	3

Steps 5. Subsets X_{21} and X_{22} contain two elements each, therefore these can be further subdivided into two subsets as shown below:

Subsets	**Probabilities**
$X_{211} = \{C\} = \left\{\frac{1}{8}\right\}$	$P(X_{211}) = \frac{1}{8}$
$X_{212} = \{D\} = \left\{\frac{1}{8}\right\}$	$P(X_{212}) = \frac{1}{8}$
$X_{221} = \{E\} = \left\{\frac{1}{12}\right\}$	$P(X_{221}) = \frac{1}{12}$
$X_{222} = \{F\} = \left\{\frac{1}{12}\right\}$	$P(X_{222}) = \frac{1}{12}$.

(a) The entropy of the source is given by

$$H(X) = -\sum_{i=1}^{6} P(x_i)\log P(x_i)$$

$$= -\left[\frac{1}{3}\log\frac{1}{3} + \frac{1}{4}\log\frac{1}{4} + \frac{2}{8}\log\frac{1}{8} + \frac{2}{12}\log\frac{1}{12}\right]$$

$$= \frac{1}{3}\log 3 + \frac{1}{4}\log 4 + \frac{2}{8}\log 8 + \frac{2}{12}\log 12$$

$$= \frac{1}{3}\log 3 + \frac{1}{4}\log 2^2 + \frac{2}{8}\log 2^3 + \frac{2}{12}\log\left(2^2 \times 3\right)$$

$$= \left(\frac{1}{2} + \frac{4}{12} + \frac{6}{8}\right)\log 2 + \left(\frac{1}{3} + \frac{2}{12}\right)\log 3$$

$$= \frac{19}{12} + \frac{1}{2}\log 3 = 2.3752 \text{ bits.}$$

(b) Average code length of the message is given by

$$L = \sum_{i=1}^{6} l_i P(x_i)$$

$$= \frac{2}{3} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8} + \frac{3}{12} + \frac{3}{12} = \frac{29}{12} \text{ bits/ symbol.}$$

(c) Efficiency of the code

$$\eta = \frac{H(X)}{L} = \frac{2.3752}{29/12} = \frac{12 \times 2.3752}{29} = 0.9828.$$

(d) Redundancy of the code

$$\beta = 1 - \eta = 0.0172.$$

120.13. Necessary and Sufficient Condition for Noiseless Encoding

Theorem 1. (Noiseless coding theorem). The necessary and sufficient condition for the existence of an irreducible noiseless encoding procedure with specified word length $(n_1, n_2, \ldots, n_N)$ and that a set of positive integers $n_1, n_2, \ldots, n_N$ can be found such that

$$\sum_{i=1}^{N} D^{-n_i} \le 1,$$

where D is the number of symbols is encoding alphabet.

Proof. Condition is necessary :

Let x_i be the number of coded messags of length n_i. Since such messages having only letter cannot be greater than D, therefore $x_1 \le D$.

Also, due to the coding restriction, the number of messages encoded of length 2 cannot exceed $(D - x_1)D$, therefore we have

$$x_2 \le (D - x_1)D = D^2 - x_1 D.$$

Similarly,

$$x_3 \le \{(D - x_1)D - x_2\}D = D^3 - x_1 D^2 - x_2 D.$$

$$\vdots$$

$$x_m \le D^m - x_1 D^{m-1} - x_2 D^{m-2} - \ldots - x_{m-1} D$$

or, $x_m D^{-m} \le 1 - x_1 D^{-1} - x_2 D^{-2} - \ldots - x_{m-1} D$

or, $\sum_{i=1}^{m} x_i D^{-1} \le 1,$ .... (*i*)

where m is the maximum length of any message.

The left hand side of inequality (*i*) can also be written as:

$$\sum_{i=1}^{m} x_i D^{-i} = x_1 D^{-1} + x_2 D^{-2} + \ldots + x_m D^{-m}$$

$$= \left[\frac{1}{D} + \frac{1}{D} + ... + x_1 \text{ times}\right] + \left[\frac{1}{D^2} + \frac{1}{D^2} + ... + x_2 \text{ times}\right]$$

$$+ .... + \left[\frac{1}{D^m} + \frac{1}{D^m} + ... + x_m \text{ times}\right]. \qquad (ii)$$

Each term in the bracket of above equation corresponds to a specified message length, such as in the first bracket, x_1 message is of length 1, in second bracket x_2 message is of length 2 and so on. Hence the total number of messages are

$$x_1 + x_2 + ... + x_m = N.$$

If $i = n_i$, then terms in (*ii*) can be rewritten as:

$$\sum_{i=1}^{n} x_i D^{-1} = \sum_{i=1}^{n} D^{-n_i} \leq 1.$$

This proves the necessary condition.

Condition is sufficient:

We have to show that the condition

$$\sum_{i=1}^{n} x_i D^{-i} \leq 1$$

is sufficient for the existence of desired codes.

Since the terms $x_1 D^{-1}, x_2 D^{-2} ..., x_m D^{-m}$ are all positive, each term must be less than 1. Thus, it can be concluded that

$$x_1 D^{-1} \leq 1 \text{ or } x_1 \leq D \text{ and } x_1 D^{-1} + x_2 D^{-2} \text{ or } x_2 \leq D(D - x_1)$$

and so on. Since these are the conditions we have to satisfy in order to guarantee that no encoded message can be obtained from any other source by the addition of a sequence of letters of the encoding alphabet.

As an application of this theorem, let D be a binary set, i.e., $A = [a_1, a_2]$, then the encoding theorem requires that

$$\sum_{i=1}^{N} 2^{-n_i} \leq 1.$$

As an application of the foregoing, consider the existence of a separable code book having N words of equal length n. The noiseless coding theorem suggests that such codes exist if

$$\sum_{i=1}^{N} D^{-n_i} \leq 1, \text{ where } n_1 = n_2 = \ldots = n_N = n$$

or, $D^{-n} + D^{-n} + \ldots + N$ times ≤ 1 or $ND^{-n} \leq 1$

or, $\log N + (-n)\log D \leq 0$ or $\log N \leq n \log D$.

This relation between N, n and D guarantees the existence of desired codes. Hence the theorem.

Example 1. There are 12 coins, all of equal weight except one which may be lighter or heavier. Using concepts of information theory show that it is possible to determine which coin is heavier.

Solution. There are 12 coins and one of which is heavier. In order to isolate the heavier coin an equal arm balance is used for weighing. Here weighing implies putting of a subset of the coins on each of the balance pans and then observing the result. The problem is to find the heavy coin in the smallest number of weighings.

Since we have 12 coins, therefore minimum three weighings are required to isolate the heavy coin. The procedure of isolating the heavier coin can be summarized as follows:

Place four coins on each pan of the balance.

Two possibilities may arise :

(a) if left (or right) pan is heavier, then the heavy coin is on the left (or right pans);

(b) if pans are balanced, then the heavy coin among the four not weighed.

In the case of (a), the heavy coin is found in two weighings whereas in case of (b) one more weighing suffices.

The expected amount of information necessary to isolate the heavy coin in case there are n coins is given by

$$H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = -\log_2\left(\frac{1}{n}\right) = \log_2 n \text{ bits.}$$

The all possible cases would be

Coin 1 is heavy with probability $\frac{1}{n}$

Coin 2 is heavy with probability $\frac{1}{n}$

....

....

Coin n is heavy with probability $\frac{1}{n}$

Thus in other words $log_2 n$ bits of information has to be accumulated to isolate the heavy coin. Hence, for i=12, the expected amount of information received is $log_2 12$ bits.

Suppose in the first weighing, there are x coins on each pan and 12–2x coins not weighed. Since coins are equally likely to be the odd one, therefore at each weighing, the following three probabilities may arise

$$P\,(\text{left pan down}) = \frac{x}{12}$$

$$P\,(\text{right pan down}) = \frac{x}{12}$$

$$P\,(\text{pans balanced}) = \frac{12-2x}{12}.$$

Then the expected amount of information obtained by the outcome of the weighing is given by

$$H = \left\{\frac{x}{12}, \frac{x}{12}, \frac{12-2x}{12}\right\}.$$

If the total number of coins are divisible by 3, then $x = 12/3 = 4$ and expected amount of information necessary to isolate the heavy coin is

$$H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = -\log_2\left(\frac{1}{3}\right) = \log_2 3.$$

In general, if one is able to get maximum $log_2 3$ bits of information in each weighing, then k number of weighings provide $k\,log_2 3$ bits of information. To isolate the heavy coin after these k number of weighings, it is required that

$$k\log_2 3 \geq \log_2 n \text{ or } 3^k \geq n.$$

120.14. Module Summary

In this module, we have introduced the concept to measure the information. The function, H, called entropy, which used to measure uncertainty, is defined and studied several properties. The marginal, joint entropies and conditional entropies are defined here. The expected mutual information is introduced. The concept of channel capacity, efficiency and redundancy are discussed. The encoding and Shannon-Fano encoding procedure are studied here. The module is ended with an exercise and the references.

120.15. Self Assessment Questions

1. Define entropy function and established its formal requirements.
2. Show that the entropy function is maximum when mutually exclusive events are equi-probable. Show that the partitioning of events into sub-events cannot decrease the entropy of the system.
3. Evaluate the entropy associated with the following probability distribution:

Event:	A	B	C	D
Probability:	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

4. Prove that $H(p_1, p_2, ..., p_n) \leq \log_2 n$, and equality holds if and only if, $p_k = \frac{1}{n}; k = 1, 2, ..., n$.
5. The following two finite probability schemes are given by $(p_1, p_2, ..., p_n)$ and $(q_1, q_2, ..., q_n)$, with

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i.$$

Then show that

$$-\sum_{i=1}^{n} p_i \log p_i \leq -\sum_{i=1}^{n} p_i \log q_i$$

with equality if and only if $p_i = q_i$ for all i.

6. Let X be a discrete random variable taking values $x_1, x_2, ..., x_n$ with probability $P(X = x_k) = p_k$; $k = 1, 2, ..., n$ where $p_k \geq 0$ and $\sum_{k=1}^{n} p_k = 1$. Define the entropy $H(p_1, p_2, ..., p_n)$ of the probability distribution to X and prove that

$$H(p_1, p_2, ..., p_n) = H(p_1, p_2, ..., p_n, p_n) + (p_{n-1} + p_n) H\left(\frac{p_{n-1}}{p_n + p_{n-1}}, \frac{p_n}{p_n + p_{n-1}}\right).$$

7. If H denotes the entropy function, then prove that

$$H(p_1,p_2,\ldots,p_n,q_1,q_2,\ldots,q_m)=H(p_1,p_2,\ldots,p_n)+p_nH\left(\frac{q_1}{p_n},\frac{q_2}{p_n},\ldots,\frac{q_m}{p_n}\right)$$

where $P_n=\sum_{i=1}^{m}q_i$. Verify the formula, defining additivity of entropies for events *A*, *B*, and *C* with probabilities $\frac{1}{5},\frac{4}{15}$ and $\frac{8}{15}$ respectively.

8. A word consists of three letters with respective probabilities $\frac{5}{12},\frac{1}{2}$ and $\frac{1}{12}$. Find the average amount of information associated with the tramsmission of letters.

9. A transmitter and receiver has an alphabet consisting of three letters each. The joint probabilities for communication are given below:

$P(x_i,y_j)$	y_1	y_2	y_3
x_1	0.45	0.45	0.01
x_2	0.02	0.02	0.01
x_3	0.01	0.02	0.01

Determine the different entropies for this channel.

10. Apply Shannon's encoding procedure to the following message ensemble

X:	A	B	C	D
P:	0.4	0.3	0.2	0.1

11. Find the capacity of the memory less channel specified by the channel matrix.

$$p=\begin{bmatrix}\frac{1}{3} & \frac{2}{3} & 0\\ \frac{2}{3} & \frac{1}{3} & 0\\ 0 & 0 & 0\end{bmatrix}.$$

120.16. Suggested Further Readings

1. S.D. Sharma, Operations Research, Kedar Nath Ram Nath & Co., Meerut.
2. J.K. Sharma, Operations Research, Macmillan.
3. M.P. Gupta and J.K. Sharma, Operations Research for Management (2nd ed.), National Publishing House, Delhi.
4. K. Swarup, P.K. Gupta and M. Mohan, Operations Research, Sultan Chand, Delhi.

---- 0 ----

Published by : Directorate of Distance Education,
Vidyasagar University, Midnapore - 721 102, W.B.
Printed by : Trade Made, Howrah-711103 & ISO 9001 : 2008 Company