CHAPTER 17

PRODUCTION PLANNING & CONTROL

Production planning and control (PPC) are the two important components which entail the acquisition and allocation of limited resources to production activities so as to satisfy customer demand over a specified time period. *Production* is a process of converting raw material into semi-finished products, and thereby adding the value of utility to the products, encompassing the activities of procurement, allocation, and utilization of resources. *Planning* is the systematic preparation for future activities based on assumptions and projections about how the object (being planned) and its environment will develop in the future. The *control* function involves supervising the operations with the aid of control mechanism and feedback information about the progress of work.

17.1 FUNCTIONS OF PPC

Production planning and control (PPC) are inter-related in a complex manner such that they look like single function of management of an enterprise. Their ultimate objective is to contribute the profits of the enterprise.

Production planning is a partial planning approach for a particular function of an organization. It encompasses coordination of parallel activities related to the production processes in order to find suitable measures to eliminate non-allowable deviations from planned production. Planning undertakes many assumptions and uncertainties in outlining the future production processes. However, an uncertain information becomes certain during production, and then it is taken as an input for control decisions to adjust the process performance. Thus, the distinction between planning and control can be justified by the uncertainty about resource availability, process performance and process results.

It will be more appropriate to describe the functions of production planning and control under respective headings [Fig. 17.1].

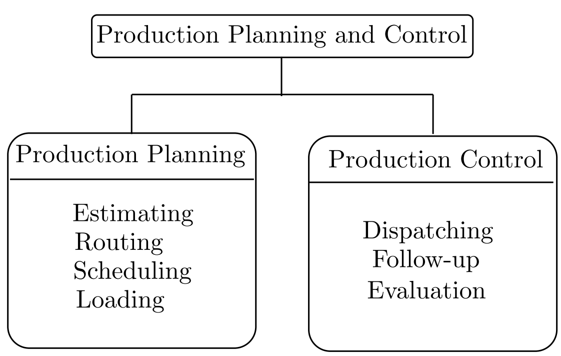

Figure 17.1 Functions of PPC.

17.1.1 Production Planning

Production planning precedes control function. It is the formulation of a new plan in-line with the objective of production. It begins with the analysis of given data on the basis of which a scheme of utilization of firm's resources can be outlined to result in desirable outputs in an efficient manner. It determines the optimal schedule and sequence of operations, economic batch size for production, machine assignments, and dispatching priorities for sequencing.

Production planning is in fact synonymous to process analysis that leads to elimination of undesirable process elements as well as improvement of certain processes. It involves the following functions:

1. *Estimating* *Estimating* function involves deciding the quantity of production and associated cost on the basis of sale forecast. This function uses forecasting techniques in assessing the production quantity.

2. *Routing* *Routing* is to determine and ensure the best and cheapest route or sequence of operations to be followed by the raw material in acquiring the shape of the finished product. Thus, routing involves two elements: operations, and their sequence.

   Routing is performed through *route sheets* for different manufacturing orders. Route sheet of a product defines each step of the production operation and lays down the precise route of operations. It includes important details, such as product identification number, symbol for identification of parts, number of pieces in each lot, operation data, production rate. An assembly of a number of component parts, like a printer or a laptop, requires separate route sheets for each of its parts, sub-assemblies and final assembly.

   In continuous manufacturing systems, such as cement or food industries, routing function is built-in with the original design of the plant and sequencing of machines. However, it is a major planning activity in the case of intermittent production, particularly in customer-oriented products.

   Routing can be generalized or detailed. Generalized routes are established by work-stations or workshops, whereas detailed routes are incorporated within a work-station for different work centers or machines.

3. *Loading* Once the route has been established, specific jobs are then assigned to work centers in view of relative priorities and optimum utilization of capacity of each work center. This is called *loading*. Total time required to perform operations at each work center can be computed by using standard process sheets. These details are then added to the work already planned for each work station. This forms a chart, known as *machine loading charts* or *Gantt charts* that show utilization of men and machines as per priority established in routing and scheduling.

   Loading can be done at various levels of an enterprise. Loading at the level of work station or machine is called *detailed loading*. Loading can also be *generalized* by assigning specific jobs to a group of machines or department as a single unit.

4. *Scheduling* *Scheduling* is deciding the priorities for each job and planning the time-table of production through considerate allocation of start and finishing time for each operation and entire series as routed. Objective of scheduling is to prevent unbalanced use of time among work centers and to utilize resources within established cycle time. Scheduling is closely related to routing because routing cannot be done without scheduling. Therefore, routing and scheduling are generally done by the same team or person of an enterprise. Routing and scheduling need immediate change in the event of contingencies, such as machine breakdowns, delays in supply of raw materials.

   Master schedule is the key plan of production, prepared on the basis of production program. It contains details of the products to be produced, their quantities, and time of delivery. Two major types of master schedules are *rolling master schedules* and *jumping master schedules*. In rolling master schedule, only a short period of production is added, whereas in jumping master schedule a long period can be covered. In the mass production system, involving repetitive or continuous type of manufacturing processes, the master schedule is prepared on the basis of anticipated demand depending on market survey, competitor position, and consumers' preferences regarding quality, price, and design. In intermittent production system, master schedule is prepared based on the orders from customers. Manufacturing schedule are based on the master schedule.

17.1.2 Production Control

Production control is a device to attain the highest efficiency in production by producing the required quantity of production of the required quality at the required time by the best and the cheapest method. This involves control on production quantity, materials and tools, spares and maintenance, labor efficiency, delivery schedule, etc. A production activity of an enterprise is said to be in control when the actual performance is within the objective of planned performance. The following are the functions of production control:

1. *Dispatching* *Dispatching* is putting the plan into effect by authorized release of resources to plant locations along with the necessary instructions to commence the production in accordance with the requirements of route sheets and schedule charts. This task is performed by a person called the *dispatcher*. Dispatching can be centralized (by central office) or decentralized (i.e. dispatching as per plans, only timing of start is then advised.).

   Quantity of production is controlled during dispatching the manufacturing order. The enterprise is required to exercise effective control over its inventory (both material and tools) to prevent over-stock and out-of-stock situations. Immediate replacement of obsolete and breakdown parts is essential to continue production. However, machine efficiency is significantly affected by the system of periodic maintenance of plant and machinery.

2. *Follow-up* Every production program involves determination of the progress of work, removing bottlenecks in the flow of work, and ensuring that the production operations are taking place in accordance with the plan. This process is called *follow-up*. Follow-up is comparing the actual performance to the planned performance in order to identify the discrepancies in the production for appropriate corrective actions.

3. *Evaluation* Deviation of actual performance from planned schedule is an usual phenomenon. This can result from breakdown of machines, unavailability of raw materials, poor performance of workmen, etc. A break at any point in the supply chain hampers the complete series of operations. Therefore, periodic *evaluation* is essential to formulate corrective action in order to bring the operations back on the schedule.

A detailed discussion on the techniques and procedures of production planning and control is beyond the scope of present context. However, important procedures are described in the following sections of this chapter and in the next two chapters.

17.2 FORECASTING

Estimating stage of production planning seeks for appropriate estimates of production quantity. This is considerably affected by growing competition, frequent changes in customer's demand, and the trend towards automation. Thus, decisions in production should not be based purely on guesses, rather on a careful analysis of data concerning the future course of events. When estimates of future conditions are made on a systematic basis of historical and current data, the process is called *forecasting* and the statement thus obtained is defined as *forecast*.

17.2.1 Forecasting Methods

The ultimate objective of forecasting is to reduce the uncertainty in management decision-making with respect to costs, profit, sales, production, pricing, capital investment, and so forth. Forecasting methods are broadly categorized into two sets: qualitative forecasting methods and quantitative forecasting methods [Fig. 17.2].

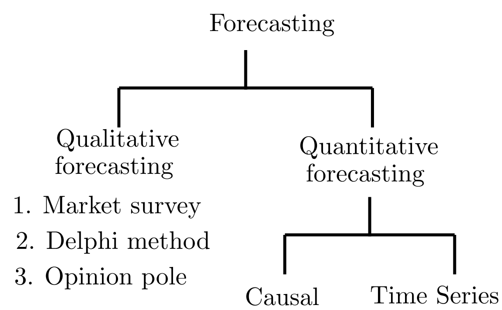

Figure 17.2 Forecasting methods.

Forecasting methods are also classified into intrinsic forecasting and extrinsic forecasting. Intrinsic forecasting uses historical data often availability within the organization itself. Extrinsic forecasting involves external factors.

17.2.1.1 Qualitative Forecasting *Qualitative forecasting* is based on intuitive information that does not have a well-defined analytic structure. This option is essentially used in absence of historical data, such as for a new product. Qualitative forecasting tends to be subjective and is often biased, depending upon the potentially optimistic or pessimistic position of the forecasting team.

Some common methods of qualitative forecasting are described as follows:

1. *Market Survey* Market survey involves structured questionnaire submitted to potential customers in the market. It is fairly expensive and time consuming task.

2. *Opinion Polls* Opinion polls are conducted to assess the opinion of the knowledge personnel and experts in the field whose views carry lot of weight.

3. *Delphi Technique* The Delphi technique[1] is a group process used to survey and collect the opinions of experts on a particular subject. Delphi technique in forecasting is based on a panel of experts in a way that eliminates the potential dominance of the most prestigious, the most verbal and best salespeople. The expert opinion is consensus instead of a compromise; the experts review each others' ideas.

4. *Life Cycle Analogy Forecasting* This technique is applicable to new product or service. It is based on the assumption that most products have a fairly well-defined life cycle consisting of the stages of gradual growth in early stage, maturity, and eventual decline in the demand.

17.2.1.2 Quantitative Forecasting-Causal *Quantitative forecasting methods* are based on the apparent causal relationship between two variables such that change in a measurable variable causes a predictable change in each other. The measurable variable effecting the change is called a *leading indicator*. Sufficient number of leading indicators enable bringing excellent forecasting results. For some objects, market survey adds the forecasting. Quantitative forecasting methods based on causality are often time-consuming and expensive primarily because of developing the relationship and obtaining the causal data.

The best example of causal forecasting is the *regression analysis*. It is a statistical technique in which past demand data is used to establish a functional relationship between two variables. One variable is known or assumed to be known; and used to forecast the value of the other unknown variable. Trend line analysis is another statistical method of causal forecasting. In this analysis, the trend line (line of best fit) is drawn on a scatter diagram, which represents the trend in the data. The trend line tells whether a particular data set has increased or decreased over a period of time.

17.2.1.3 Quantitative Forecasting-Time Series Most common quantitative forecasting methods are based on the assumption that past demand follows some pattern, which can be used in developing projections for future demand. The first step in making a forecast is gathering information from the past in the form of statistical data recorded at successive intervals of time. Such a data is usually referred to as *time series*. Demand data is plotted on a time scale to study and look for consistent shapes and patterns.

[1]The Delphi technique was named after the Ancient Greek oracle at Delphi, who could predict the future. An oracle refers to a statement from someone of unquestioned wisdom and knowledge or of infallible authority. The technique was developed by Olaf Helmer and his associates at the Rand Corporation in the early 1950s.

Such methods are commonly known as time-series forecasting because time is the only real independent variable in statistical data. Since these methods are based on internal data (sale), they are some times called *intrinsic forecasts*. Understanding the reason behind the trend of demand in time series helps in forecasting. The pattern pattern can be random (unpredictable), trending (apparently definite) or seasonal (periodic).

Some common time-series forecasting models are described as follows:

1. *Simple Average Method* A *simple average method* forecasts the demand of the next time period as the average of demands occurring in all previous time periods.

2. *Simple Moving Average Method* In *simple moving average method*, the average of the demands from several of the most recent periods is taken as the demand forecast for the next time period. The number of past periods to be used in calculations is selected in the beginning and is kept constant.

3. *Weighted Moving Average Method* In *weighted moving average method*, unequal weights are assigned to the past demand data while calculating simple moving average as the demand forecast for the next time period. Usually, the most recent data is assigned the highest weight factor.

4. *Exponential Smoothing* In *exponential smoothing method*[2], weights are assigned in exponential order. The forecast for time period t is related to the demand for the previous period D_{t-1} and forecast of previous period F_{t-1} as[3]

$$F_t = F_{t-1} + \mu\left(D_{t-1} - F_{t-1}\right)$$

where μ is called *smoothing constant*. This method takes care of error $\left(D_{t-1} - F_{t-1}\right)$ in old forecasting. The weights decrease exponentially from most recent demand data to older demand data.

By direct substitution of the defining equation for simple exponential smoothing back into itself,

$$\begin{aligned} F_t &= \mu F_{t-1} + \mu\left(1-\mu\right) F_{t-2} \\ &\quad + \mu\left(1-\mu\right)^2 F_{t-3} + \ldots. \\ &\quad + \mu\left(1-\mu\right)^{t-1} F_0 \end{aligned}$$

Thus, F_t is found to be weighted average with general proportion of μ, $\mu\left(1-\mu\right)$, $\mu\left(1-\mu\right)^2$.., which is geometric progression and discrete form

[2]Exponential smoothing was first suggested by Charles C. Holt in 1957.

[3]The formulation is attributed to Brown and is known as Brown's simple exponential smoothing.

of exponential function as

$$e^{\mu} = \sum_{n=0}^{\infty} \frac{\mu^n}{n!}$$

Therefore, this method of forecasting is called exponential smoothing.

17.2.2 Forecasting Errors

The numeric difference in the forecasted demand and actual demand is known as *forecasting error* (e). The cost of a forecasting error can be substantial. Forecasting can be improved by examining some objective evaluations of alternative forecasting techniques. Various statistical measures can be used to measure forecasting errors of various models.

Regardless of the object of forecasting, the following are some important fundamental characteristics of forecasting error:

1. When forecasts are almost wrong, the potential error in the forecast can be accommodated through use of buffer capacity.
2. Forecasting is easier for a product line because forecasting errors for an individual product tend to cancel each other.
3. Forecasts for short time periods are more accurate due to possible disruptions in the product demand.
4. Every forecast is incomplete without mentioning estimate of forecast error.
5. Forecasting cannot be substituted by calculation of demand based on actual data for a given time period.

The commonly used measures for summarizing historical errors include following:

1. Mean Absolute Deviation *Mean absolute deviation* (MAD) is determined as

$$\text{MAD} = \frac{\sum |e|}{n}$$

2. Bias *Bias* is defined as

$$\text{Bias} = \frac{\sum e}{n}$$

3. Mean Square Error *Mean square error* (MSE) is determined as

$$\text{MSE} = \frac{\sum e^2}{n-1}$$

where e is the forecasting error for a period in the data, and n is the total number of periods of forecasting.

Forecasting can be monitored by a parameter known as *tracking signal*. It is based on the ratio of cummulative forecast error to the corresponding value of MAD:

$$\text{Tracking signal} = \frac{\sum e}{\text{MAD}}$$

17.3 AGGREGATE PLANNING

Aggregate planning is aimed at pre-estimating the procurement quantity and scheduling the output over an intermediate range by determining optimum levels of the production rate, employment, inventory and other controllable variables. It involves aggregate decisions rather than stock-keeping unit (SKU) level decisions. The resources cannot be changed as fast a rate as the demand. Therefore, aggregate planning offers strategies to the production system so that it can absorb the fluctuations in demand by trade-offs among controllable factors, such as capacity, time, inventory. This can also be done by accepting back orders and subcontracting.

Several methods have been suggested by researchers to absorb fluctuating demand. Notably among those are *linear decision rule*, *graphical charting*, and mathematical programing. Linear decision rule (LDR) was developed by Holt, Modagiliani, Muth and Simon, therefore, it is known as *HMMS rule*. The model initiated with two sets of decision variables, namely, work force and production rate. Graphical techniques are popular because they are easy to understand and use. Linear programming (e.g. transportation model) is a mathematical approach that produces optimal plan for minimizing the costs.

17.4 DISAGGREGATION

The aggregate plan works on the level of workforce and production for medium and long-term periods. For practical implementations, the aggregate plan for each product family is subjected to disaggregation into operational production plans over a short scheduling period. In simple words, the data of aggregate plan are further broken down to detailed levels.

Disaggregation is not a complex process because it involves several steps and a variety of trade-offs that must be made before a final item-by-item production schedule is obtained. The most successful schemes use a hierarchical planning model that explicitly ties together the decisions at each level of planning in order to be internally consistent. The disaggregated plan is assigned the limited resources to meet the cumulative production data for each product family over a given periodicity of production.

17.5 MATERIAL REQUIREMENT PLANNING

For dependent demand situations, normal reactive inventory control systems, such as economic order quantity models, are not suitable because they result in high inventory costs and unreliable delivery schedules. Unavailability of even one component can cause discontinuity in the production. *Material requirement planning*[4] (MRP) is a special technique to plan the requirement of materials for production. It deals with the materials which directly depend upon the requirement of production. This technique employs production plan or schedule to arrange for the raw materials, rather than depending upon EOQ models. MRP is a simple system of calculating arithmetically the requirement of input materials at different points of time based on the actual production plan. It can be simply defined as a planning and scheduling system to meet time-phased material requirements for production operations, without any probability.

Bill of materials (BOM) is a detailed list of materials required to produce a product. It is constructed in a way that reflects the manufacturing process so that it can be used in material requirement planning.

Fig. 17.3 shows a product structure which consists of three assemblies: A, B and C. There are five materials (1 to 5) required to produce this product.

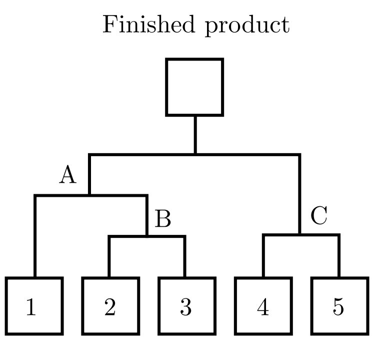

Figure 17.3 Product structure.

Using product structure, the bill of materials recognizes the dependence of certain components on subassemblies, which in turn depend on the final product.

[4]The concept of material requirement planning was introduced by Joseph Orlicky in the early 1960s.

17.6 BREAK-EVEN POINT ANALYSIS

The *break-even point analysis* is a valuable planning and control technique that uses the relationship between production rate and costs of profits. The analysis is performed on a chart that shows a *break-even point* where profit is zero [Fig. 17.4].

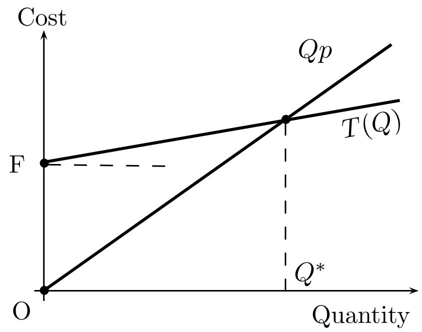

Figure 17.4 Break-even analysis.

Production costs consist of two elements:

1. *Fixed Cost* Costs that remain relatively constant regardless of the level of activity are known as *fixed costs* or *indirect costs*. It is considered as preparation expenses to produce a product or service.

2. *Variable Cost* Costs that are generally proportional to output are called *variable costs*, or *direct costs*. Such costs are relatively easy to determine.

Revenues results from sales of output. *Profit* represents the difference between revenue and total cost.

Suppose a factory is planning to produce a quantity Q units and sale it at the price rate Rs. p/unit. Fixed cost for the production is Rs. F and variable cost is Rs. V/unit. The revenue on sales of Q units will be Qp. Break-even quantity is determined as

$$Q^* p - F + VQ^* = 0$$

$$Q^* = \frac{F}{p - V}$$

Total cost $T(Q^*)$ and profit $P(Q^*)$ at break-even point are found as

$$T(Q^*) = F + VQ^*$$

$$P(Q^*) = (p + V)\,Q^* - F$$

The business should be immediately closed when variable cost line becomes vertical.

17.7 LOT SIZING RULES

Production planning essentially require the lot size, which is most economical for the manufacture of goods. For this purpose, the cost of production of each lot size is to be worked out and the lot size giving the least cost for unit is selected for manufacturing purpose. *lot-sizing rules* tend to offer lumpy demand conditions and result in inventory for the lower-level components (raw materials).

Some of the commonly used *lot-sizing rules* are summarized as follows:

1. Fixed Order Quantity In this policy, the order quantity is fixed on the basis of intuitive method. The method is suitable for selected items having ordering costs sufficiently high to rule out in net requirement quantity per period.

2. Economic Order Quantity The economic order quantity (EOQ) method is preferable when relatively constant independent demand exists. It is a statistical technique using averages, whereas the MRP procedure assumes known demand reflected in the master production schedule.

3. Lot for Lot Lot-for-lot technique is based on the principle that an MRP system should produce only as needed, with no safety stock and no anticipation of further orders. The technique is efficient when frequent orders are economical and just-in-time inventory technique is implemented.

4. Fixed Period Requirement In this policy, the lot quantity per order is specified to cover certain number of periods.

5. Periodic Order Quantity In *periodic order quantity* (POQ) lot sizing, ordering intervals related to the quantity is computed, and hence, carrying cost tends to be lower. Therefore, it is more effective than EOQ.

6. Part-Period Balancing Pure EOQ approach does not take into account the variations in requirements over time. A simple heuristic method that does this is the *part period balancing* (PPB) method developed by DeMatteis. A part period is defined as a unit of measure that is equivalent to carrying one unit of an item (a part) in inventory for one period. The PPB heuristic is based on the observation that the optimal order quantity in the basic EOQ model occurs when the total ordering or setup cost equals the total holding cost.

17.8 ASSEMBLY LINE BALANCING

The work in an *assembly line* passes through a series of workstations in a uniform time interval, known as workstation cycle time (t). Assembly work performed at each workstation takes its own time that depends upon the work elements. Work time at different workstations can be different. Objective of *assembly line balancing* is allocating operations to each work station of the assembly line without violating the precedence and without exceeding the cycle time. Balance in an assembly line is examined through the following features:

1. Cycle Time Theoretically, workstation cycle time for an assembly is defined as

$$t_c = \frac{\text{production time per day}}{\text{required output per day}}$$

2. Number of Workstations Theoretical minimum number of work stations (n_t) required to satisfy the workstation cycle time constraint is determined as

$$n_t = \frac{\sum t}{t_c}$$

3. Cycle Efficiency *Cycle efficiency* or line efficiency (η) of the assembly line is determined as

$$\eta = \frac{n_t}{n_a} = \frac{\sum t}{n_a t_c}$$

where n_a is the actual number of workstations.

The objective of the assembly line balancing can be viewed as to minimize the number of workstations (n_t), which is equivalent to maximizing the *cycle efficiency* (η).

4. Balance Delay *Balance delay* is the amount of idle time on production assembly lines caused by the uneven division of work among operators or stations. It is related to line efficiency as

$$\text{Balance delay} = 1 - \eta$$

CHAPTER 18

INVENTORY CONTROL

Inventory refers to any kind of resource that has economic value and is maintained to fulfill the present and future needs of an organization. In a production system, *inventory* accounts for a large percentage of working material. It can be of raw material, goods in process, finished goods, and goods assisting in production. Creation and maintenance of inventory is very essential for continuity of production by providing decoupling function between stages of production, ensuring full utilization of sources, and successfully meeting the variations in customer demands. It also takes advantage of lowering the total material cost by availing the quantity discounts. Inventories involve capital tied up, storage and handling cost, deterioration, pilferage, and obsolescence. Inventory control is a procedure of production control by means of which inventory of appropriate quantity is availed for manufacturing operations without entailing the cost involved in overstocking or under-stocking at any instant.

18.1 BASIC CONCEPTS

Informal procedure of inventory management by intuitive determinations can work well with a small company as the number of items are few. However, for an enterprise requiring a wide variety of inventory items having different usage rates, informal systems tend to create problems that result in higher costs and interruptions in production. In such cases, a formal system of inventory management can produce substantial savings. This section describes basic principles that are important for designing a formal system of inventory management.

18.1.1 Types of Inventory

Inventory can be following types:

1. *Lot Size Inventory* Lot size inventory is the periodic inventory to meet the average replenishment in production. Decision on lot size and timings are very important for economical use of personnel and equipment. Continuous production is suitable for high-volume items. However, low-volume items should be preferably produced periodically and in economic lots.

2. *Transit Inventory* Movement of inventory cannot be instantaneous. To prevent delay in supply to work centers, some inventory is essentially in transit between work centers. This is called *transit or process inventory.*

3. *Buffer Inventory* Additional inventory is required to accommodate the uncertainties of the demand and the *lead time*. This stock is called *buffer* or *safety inventory*. However, there is a minor difference between these two terms; buffer inventory

refers inventory reserved to meet customer demand when customer ordering patterns vary, whereas *safety inventory* refers the inventory reserved to meet the customer demand when internal constraints or inefficiencies disrupt in the process flow.

4. Decoupling Inventory Stock points are created between adjacent stages of production in order to achieve decoupling of the stages. This is essential to prevent disruptive effect of breakdown on the entire system of production. Stock points are supplied with *decoupling inventory*.

By convention, manufacturing inventory generally refers to items that contribute or become part of a product. Manufacturing inventory is typically classified into raw materials, finished products, component parts, supplies, and work-in-process inventory.

18.1.2 Costs of Inventory

Apart from purchase cost, two types of costs are associated with inventory:

1. Carrying Cost The cost associated with carrying the inventory in the stores is called *carrying cost* or *holding cost* ($H(Q)$). It includes rent, interest of the money locked up, insurance premium, salaries of store keepers, deterioration, etc.

   If the average inventory held is $\bar{Q}$ units and average holding cost is h (Rs. per unit per annum), then carrying cost is written as

$$H(Q) = \bar{Q}h$$

2. Ordering Cost The cost associated with the procedure for the placement of purchase orders of the inventory is called *ordering cost* or *setup cost* ($O(Q)$). It includes the cost of stationary, postage, telephone, traveling expenses, material handling, etc. The ordering cost is independent of the batch size of the order.

   If A is the ordering cost per order and annual demand is D for ordered quantity Q, then total D/Q orders will be placed per annum. Therefore, ordering cost is

$$O(Q) = \frac{D}{Q}A$$

The total *annual inventory cost* $T(Q)$ is the sum of total annual ordering cost $O(Q)$ and total annual holding or carrying cost $H(Q)$, mathematically

$$T(Q) = O(Q) + H(Q)$$

The costs associated with non-delivery of a demanded item is called *stock-out cost* or *shortage cost*. The backlogging of orders often involves extra cost for administration, price discounts for late deliveries, material handling, and transportation. There can be situation of loss in sale due to selection of alternative suppliers. *Service level* is a measure of the degree of stock-out protection provided by a given amount of safety inventory.

18.1.3 Inventory Demand

Size of the demand is the number of units required in each period. The demand pattern of an item can be of two types:

1. Deterministic Demand The inventory model using the assumption of constant and known *demand* for the item and *lead time* are called *deterministic models*. Here, the stock is replenished as soon as the stock reaches the point of exhaustion. In such a situation, there is no need to maintain any extra stock.

2. Probabilistic Demand When the demand over a period is uncertain, but can be predicted by a probability distribution, the demand is called *probabilistic demand*.

18.1.4 Inventory Replenishment

The size of replenishment orders affects inventory level to be maintained at various stock points. Inventory policies are aimed at determining when to replenish the inventories and how much to order at one time. This is compounded by price discounts and by the need to prevent disruption in operations due to delays in supply time and temporary increase in requirements.

The period of time between two consecutive placements of orders is called *ordering cycle*. *Lead time* is the time between placing an order and its delivery. *Reorder point* (ROP) is defined as the amount of stock equal to the demand during the *lead time*, which will be consumed by the time the fresh delivery is due to arrive.

Overstocking, the hard core of inventory mismanagement, results in the wastage of scarce resources of a business enterprise. Inventories are financed through borrowings, thus involve interest charges payable to the creditors, irrespective of the profits. Understocking is as disastrous or even more fatal to the manufacturing enterprise. It involves the risk of short supply and as a result the capacity cannot be utilized optimally.

18.1.5 Inventory Control Systems

An inventory system provides the organizational structure and the operating policies for maintaining and controlling the inventory. Inventory systems can be broadly classified into two categories:

1. *Static Inventory System* These are single purchase decision systems for a single period without replenishment. The inventory cost is optimized between the cost of having and the cost of not having an item in stock.
2. *Dynamic Inventory System* These are multi-purchase decision systems, concerned with consumable spares, that make replenishment decisions on an ongoing basis over time. Majority of inventory problems belong to this type of situation.

18.2 EOQ MODELS

Objective of the EOQ models is to determine the optimal order quantity that minimizes the total incremental cost of holding an inventory and processing order when demand occurs at a constant rate. The optimum level of quantity is called *economic order quantity* (EOQ). There are various EOQ models for different situations, but the following two models are relevant in the present context of study.

18.2.1 Simple EOQ Model

Simple EOQ model[1] is based on the assumptions evident from the inventory replenishment cycle depicted in Fig. 18.1. The demand is deterministic and constant. The depletion rate of the inventory is also constant. The inventory is replenished immediately when stock level reaches zero level.

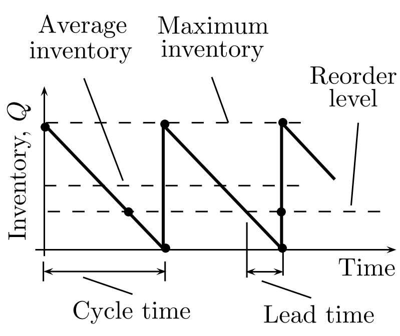

Figure 18.1 Simple EOQ model.

Let Q units be the order quantity to cover annual demand D (units per annum). The components of inventory costs are determined as follows:

[1]Simple EOQ model was developed in 1913 by F. Wilson Harris but R. H. Wilson did its in-depth analysis.

1. *Ordering Cost* There will be D/Q number of orders per year. Therefore, annual ordering cost of the inventory is determined as
$$O(Q) = \frac{D}{Q}A$$
where A is the cost of one order.
2. *Holding Cost* The inventory is maximum (Q) at the replenishment and gradually decreases to zero. Annual average holding cost of the inventory is determined as
$$H(Q) = \frac{Q}{2}h$$
where h is the annual holding cost per unit of the inventory.

Total inventory cost is determined as

$$T(Q) = \frac{D}{Q}A + \frac{Q}{2}h \tag{18.1}$$

Figure 18.2 shows the variation of $O(Q)$, $H(Q)$, $T(Q)$ with respect to order quantity Q.

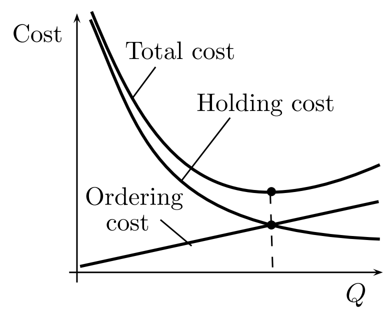

Figure 18.2 Total cost in simple EOQ model.

The order quantity associated with the minimum value of $T(Q)$ can be determined as

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

$$-\frac{DA}{Q^2} + \frac{h}{2} = 0$$

$$Q = \sqrt{\frac{2AD}{h}}$$

This value of order quantity is called the *economic order quantity* and it is denoted by Q^*. Therefore,

$$Q^* = \sqrt{\frac{2AD}{h}} \tag{18.2}$$

This equation is called *Wilson Harris formula.*

Total annual inventory cost at economic order quantity can be determined by using Eq. (18.1) as

$$T(Q^*) = \frac{D}{Q^*}A + \frac{Q^*}{2}h$$
$$= \sqrt{2ADh}$$

The optimal number of production runs (n^*) is

$$n^* = \frac{D}{Q^*}$$

The optimum purchase cycle time (t^* in days) is

$$t^* = \frac{Q^*}{\text{Demand per day}}$$
$$= \frac{Q^*}{D/365}$$
$$= \frac{365}{n^*}$$

Using Eq. (18.2), the economic order quantity is found to be inversely proportional to the square root of the holding cost:

$$Q^* \propto \frac{1}{\sqrt{h}}$$

Let s be the stock-out cost per incidence. Thus, the system has to attain maximum inventory level Q^* at holding cost of $h+s$. The carrying cost associated with safety stock will be s. Therefore, the safety stock can be determined as

$$Q_s = Q^* \times \sqrt{\frac{s}{s+h}}$$

18.2.2 Build-Up EOQ Model

Build-up EOQ model is applied when inventory is not replenished in one shot but rather continuously over a time period. Let the total order quantity Q be the build-up at a constant *production rate* p over a period t_p ($= Q/p$) [Fig. 18.3].

The average inventory level would be determined not only by the lot size Q, but will also be affected by the production rate (p) and the *depletion rate* (d):

$$\bar{Q} = \frac{p-d}{2}t_p$$
$$= \frac{Q}{2}\left(1-\frac{d}{p}\right)$$

Total inventory cost is

$$T(Q) = \frac{D}{Q}A + \frac{Q}{2}\left(1-\frac{d}{p}\right)h$$

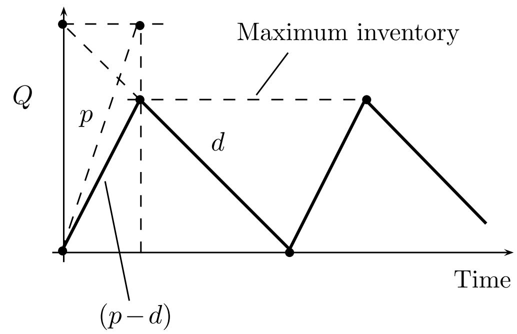

Figure 18.3 | Build-up EOQ model.

Minimization of $T(Q)$ with respect to Q gives the expression for build-up EOQ as

$$Q^* = \sqrt{\frac{2AD}{h\,(1-d/p)}}$$

The effective holding cost in build-up model against simple EOQ model is given by

$$h' = h\left(1-\frac{d}{p}\right)$$

18.3 PROBABILISTIC INVENTORY MODELS

Simple inventory models are based on constant demand and supply lead time. However, in real applications, the demand can be uncertain and lead time often varies significantly. In such situations, the risks of stock-out can be reduced by carrying safety or buffer stock, which requires additional funds. Therefore, probabilistic inventory models are developed to balance the risks and minimize the incremental costs. A detailed discussion on these models is out of context of the book.

Single period model is used in uncertain demand situations. The key trade-off in this model is to balance the cost of overstocking the inventory if there are leftovers at the end of the cycle with the cost of understocking in terms of the profit foregone if the demand turns out to be higher than the stock at hand.

Order quantity - reorder point (Q-ROP) model is developed for situations when item is continuously demanded at constant rate for a long time horizon but with wide variations in lead time. This model takes into consideration both the expected demand during lead time and safety stock.

Least unit cost (LUC) technique chooses a lot size that equals the demand of some k (> 0) periods in future. The average holding and ordering cost per unit is computed for each $k = 1, 2, 3$, etc. The computation starts from $k = 1$ and increasing k by 1 until the average cost per unit starts increasing. The best k is the last one up to which the average cost per unit decreases.

18.4 SELECTIVE APPROACHES

In a large production system, all items of the inventory cannot be controlled uniformly. Therefore, selective approaches are used for inventory control on priority basis. These techniques are based on nature and usage rate of inventories. Some of them are described as follows:

1. ABC Analysis *ABC analysis* is based on the concept "thick on the best, thin on the rest." All inventory items are classified in three distinct categories: (A) High consumption - strict control, (B) Moderate consumption - fair control, and (C) Low consumption - open storage. The analysis is also used as an acronym of Always Better Control.

   The analysis is accomplished by plotting the usage value of the items to obtain the ABC distribution curve which is also called the *Pareto curve*[2] [Fig. 18.4].

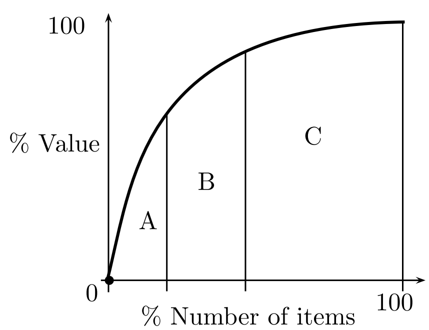

Figure 18.4 ABC Classification.

2. VED Analysis *VED analysis* is applied for spare parts. The items are classified as vital, essential and desirable based on their importance to the production process. Vital items are the items without which the production process would come to a standstill. Essential items would adversely affect the efficiency of the production system although the system would not altogether stop for want of these items. Desirable items do not cause an immediate loss of production.

3. HML Analysis In *HML analysis*, the items are classified based on their unit value (not annual usages) as high value, medium value, and low value items. This analysis helps in administrative decisions in an organization, such as procurement authority which is generally based on the hierarchy and price of material.

4. SDE Analysis *SDE analysis* is done based on purchasing problems, such as long lead time, scarcity and low availability, geographically scattered sources, uncertainty in supply. The items are classified as Scarce, Difficult, and Easy items.

5. FSN Analysis In *FSN analysis*, the items are classified into Fast, Slow-moving, and Non-moving items, based on their issue rates or consumption pattern. This analysis enables in controlling the obsolescence. High consumption items desired attention for their uninterrupted procurement. Non-moving items (almost nil consumption) indicate obsolete inventories due to changes in their specifications.

6. GOLF Analysis *GOLF analysis* is carried out mainly based on the source of material. GOLF stands for Government, Ordinary, Local and Foreign. This classification helps in describing the special procedure to be followed for procurement of materials from specialized sources.

7. XYZ Analysis *XYZ analysis* is based on the value of the inventory undertaken during the closing of annual accounts. X items are those having high value, whereas Z items are of low value. This analysis is used to identify items causing locking up money in the stock.

18.5 JUST-IN-TIME PRODUCTION

In a *just-in-time* (JIT) production system[3], materials are produced only at the time when they are needed in required quantity. For this, the companies make agreements with their vendors. This is also called *kanban system*. Literal meaning of *kanban* is visual record, hence it is also known as *card system*.

Jidoka, along with just-in-time, is one of the two main pillars of the Toyota production system. Japanese word "ji-do-ka" means automation with a human mind, and implies the ability of production lines to be stopped in the event of such problems as equipment malfunctions, quality problems or work being late. This is desired to prevent passing on defects.

[2]The Pareto curve is named after the Italian economist Vilfredo Pareto (1848-1923). In 1906, he made the famous observation that twenty percent of the population owned eighty percent of the property in Italy, later generalised by Joseph M. Juran into the Pareto principle (also termed the 80-20 rule).

[3]This system was found by Taiichi Ohono, vice president of Toyota Motor Company of Japan.

CHAPTER 19

OPERATIONS RESEARCH

Operations research (OR) is the application of scientific methods primarily in addressing the problems of decision making by providing systematic and rational approaches. The term 'operations research' (coined by McClosky and Trefthen in 1940 in the UK) came from military operations during World War-II when scientists dealt with strategic and tactical problems. These techniques were later applied in business, industry and research. Operations research provides a quantitative basis for decisions regarding the operations kept under control. Operations research is based on the scientific approach through theoretical models to understand and explain the phenomenon of operating systems.

19.1 SIMPLEX METHOD

Dantzig's *simplex method* is a general procedure of iterative nature for obtaining systematically the optimal solution to a linear programming problem (LPP). The method is based on the property that if objective function does not take the maximum value in a vertex, then there is an edge starting at that vertex along which the value of the function grows.

The procedure is comparable with the *graphical method* wherein *feasible solution* is located at corner points of the feasible region determined by the constraints of the system. The simplex method is just the same but it is done by table. The algorithm is very simple and efficient, as it considers only those feasible solutions which are provided by the corner points, and that too not all of them. Thus, the technique considers minimum number of feasible solutions to obtain an optimum one.

The procedure begins by assigning values to an appropriately selected set of variables introduced into the problem, and primary (decision) variables of the problem are all set equal to zero. This assumption is analogous to starting the evaluation in graphical approach at the point of origin. The algorithm then replaces one of the initial variables by one of decision variable which contributes most to the desired optimal value. This is repeated until the algorithm terminates indicating the optimal solution of the problem.

19.1.1 Problem Definition

In vector form, the problem is to maximize or minimize the objective function:

$$Z = c_j x_j \tag{19.1}$$

This is subjected to constraints:

$$a_{ij} x_j \leq b_i \tag{19.2}$$

where all x_j and b_i are non-negative (≥ 0).

The following notations are employed:

Z = Objective function
x_i = Variables of the problem
c_j = Coefficients of x_j in the objective function
a_{ij} = Coefficient of x_j in the ith constraint
b_i = Right hand side value of the ith constraint

The coefficients c_j represent the contribution of variables per unit in the objective function.

Implied summation is used in writing the above expressions. However, the problem can be written without implied summation as follows:

1. Objective function [Eq. (19.1)]

$$Z = c_1x_1 + c_2x_2 + c_3x_3 + \cdots$$

2. Constraints [Eq. (19.2)]

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots &\leq b_2 \\ \ldots\ldots &\leq .. \\ a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots &\leq b_i \end{aligned}$$

19.1.2 Conditions for Applicability

Simplex method is applicable to an LPP which satisfies the following two conditions:

1. Right hand side of each constraint (b_i's) is non-negative (≥ 0).

   To ensure this, the constraints having negative b_i's should be multiplied by '-1' in order to reverse the direction of inequality. For example, the equivalent of constraint $x_1 - 3x_2 \geq -6$ is $-8x_1 + 3x_2 \leq 6$.

2. Each of the decision variables (x_i's) of the problem is non-negative (≥ 0).

   To ensure this, the decision variables without any restriction (if any) on their sign are dealt by treating such variables at the difference of two non-negative variables. For example, an unrestricted variable x_1 can be assumed to be $x_2 - x_3$ where x_2 and x_3 are non-negative variables (≥ 0). This shall also need modifications in all the constraints and objective function (Z).

19.1.3 Simplex Algorithm

The simplex method proceeds by preparing a series of tables called *simplex tableaus*. Steps of the method are described under the following headings.

19.1.3.1 Standardization of the Problem Unless otherwise specified, the following are the modifications required to have constraints in the desired format irrespective of minimization or maximization types of problems:

1. *Modification of Constraints* Irrespective of the problem type, the constraints of LPP require modifications using the following variables:

   (a) *Slack Variables* - Inequalities ($\leq$) of constraints ($a_{ij}x_j \leq b_i$) need conversion into equations by introducing positive non-negative *slack variable*, denoted by s_i (≥ 0), on the left hand side to get the following form of constraints:
   $$a_{ij}x_j + s_i = b_i \quad (19.3)$$
   These are called *slack variables* because they take up any slack between the left and right hand side of the inequalities upon conversion into equation. Slack variables can take values $0 \leq s_i \leq b_i$.

   (b) *Artificial Variables* - *Artificial variables* are used to deal with the inequality of the constraints $a_{ij}x_j \geq b_i$ type for obtaining the initial solution. Such constraints take the following form:
   $$a_{ij}x_j + A_1 = b_i \quad (19.4)$$
   Use of artificial variables in turn helps in preventing use of negative slack variables.

2. *Modification of Objective Function* The objective function of the problem should contain every variable in the system, including slack variables (s_i's) and artificial variables (A_i's), if any. These are incorporated as follows:

   (a) *Slack Variables* - The coefficients of variables in objective function represent the cost involved in using the raw materials or labor hours. However, unused resources represented by slack variables have no cost and effect on the profits. Thus, the coefficients of slack variables in the objective function are assigned zeros, as follows:
   $$Z = c_ix_i + 0 \times s_i \quad (19.5)$$

   (b) *Artificial Variables* - Artificial variables do not represent any quantity relating to the decision problem, they must be driven out of the system and must not reflect in the final solution. This can be ensured by assigning high cost to them. Therefore, a value M, higher than finite number, is assigned to each artificial variable in the objective function:
   $$Z = c_ix_i + 0 \times s_i \pm MA_i \quad (19.6)$$

To effect the proper direction of iterations, negative sign of M is used in maximization type problems and positive sign is used in minimization type problems. For this reason, the technique of using artificial variables is known as *big-M method*.

Conclusively, the *objective function* (Z) is modified by incorporating all the slack variables (s_i) with their zero coefficients, and all the artificial variables (A_i) with the same coefficient M having $-$ve sign for maximization and +ve sign for minimization type problems.

19.1.3.2 Obtaining the Simplex Tableau

Initial simplex tableau is written in following format:

	c_j	c_1	c_2	c_3	$\cdots$	
e_i	CSV	x_1	x_2	s_1	$\cdots$	b_i
e_1	s_1	a_{11}	a_{12}	a_{13}	$\cdots$	b_1
e_2	s_2	a_{21}	a_{22}	a_{23}	$\cdots$	b_2

This tableau reflects the following additional elements:

1. *Identity Matrix* After writing the initial simplex tableau in the above format, the next step is to locate the identity (matrix) and variables involved in the simplex tableau. The identity contains all zero except diagonal column of positive 1's. The identity must have this square form with all zeros and a diagonal of (+ ve) ones.

   An *identity matrix square* is obtained when slack variable are introduced with coefficient of unity. It appears automatically where coefficients of other variables in the identity are equal to zero.

2. *Basic Variables* The variables in the identity matrix are called *basic variables* and the remaining are called *non-basic variables*.

   In general, if the number of variables (x_i 's and s_i) is n, and number of constraints (excluding $x_1, x_2 \geq 0$) is m, then, the number of basic variables is m, and the number of non-basic variables is $n-m$. In turn, size of square of identity is determined by the number of constraints.

3. *Current Solution Variables* The basic variables form the basis of solution, and are known as the *current solution variables* (CSV) in the simplex tableau of the latest feasible solution. In every iteration, one of the CSV's is replaced by a new basic variable and corresponding value (equal to b_i). Additional column is added on the left side to repeat the respective value of coefficients of the basic variables in the current solution (e_i) to ease the calculations.

19.1.3.3 Obtaining Feasible Solution

A feasible solution is obtained by assigning zero to all variables except basic variables, and then assign the values of the constraints (b_i's) of the variable in the identity.

	c_j	c_1	c_2	c_3	$\cdots$	
e_i	CSV	x_1	x_2	s_1	$\cdots$	b_i
e_1	x_1	a_{11}	a_{12}	a_{13}	$\cdots$	b_1
e_2	x_2	a_{21}	a_{22}	a_{23}	$\cdots$	b_2

For the above table, the initial solution will be $x_1 = b_1$, $x_2 = b_2$, $s_1 = 0$, and so on.

Solution from the initial simplex tableau is not always the optimal solution but indicates the value of optimal function at the origin of x_1 and x_2 coordinates.

19.1.3.4 Testing the Optimality

The feasible solution reflected in CSV of the simplex tableau is tested for optimality under the following steps:

1. *Obtaining z_j* Coefficients z_j are defined with implied summation as

$$z_j = a_{ij} e_i \tag{19.7}$$

   Their value is obtained under each (jth) variable column head. For this, each element of the jth column (a_{ij}) is multiplied by the corresponding coefficient of the solution variables appearing in the basis (e_i), and then the products are added up to get z_j. Here, z_j represents the reduction in profit if one unit of any of the variables is added to the matrix.

2. *Obtaining Δ_j* A new set of coefficients Δ_j is defined as

$$\Delta_j = c_j - z_j \tag{19.8}$$

   The row constituted by Δ_j is called *net after opportunity cost row*. The calculations are done in the simplex tableau as shown below:

	c_j	c_1	c_2	c_3	$\cdots$	
e_i	CSV	x_1	x_2	s_1	$\cdots$	b_i
e_1	x_1	a_{11}	a_{12}	a_{13}	$\cdots$	b_1
e_2	x_2	a_{21}	a_{22}	a_{23}	$\cdots$	b_2
	z_j	z_1	z_2	z_3	$\cdots$	
	Δ_j	Δ_1	Δ_2	Δ_3	$\cdots$	

3. *Optimality Test* Except when an artificial variable is included in the basis, a simplex tableau depicts an optimal solution if all entries in the net after opportunity cost row (Δ_j) are non-positive for maximization type problems or non-negative for minimization type problems:

$$\Delta_j = \begin{cases} \leq 0 & \text{maximization type} \\ \geq 0 & \text{minimization type} \end{cases} \tag{19.9}$$

19.1.3.5 Improving the Solution If the solution is not found optimal using Eq. (19.9), the next step is to improve the solution by the following steps:

1. *Finding Key Element* Presence of positive Δ_j (for maximization type problems) or negative Δ_j (for minimization type problems) indicates that the solution can be improved. This is done by replacing the least potent basic varible in CSV by the most potent non-basic variable, described as follows:

   (a) The variable with largest positive Δ_j value (for maximization type) or negative Δ_j (for minimization type) is selected as the incoming variable to CSV. The corresponding column is called *key column* or *pivot column*, of the row $j = k$.

   (b) The values of b_i's are divided by the corresponding values in the key column to get b_i/a_{ik}, which is called *replacement ratios*. The row with the least non-negative quotient is then selected as the *key row* and corresponding variable to this represents the *outgoing variable* from CSV.

   The key column and key row meet at *key element*, can be marked by asterisk $(*)$ in the superscript. The outgoing variable is replaced by incoming variable along with corresponding e_i.

2. *Optimizing the Solution* To find the next optimized solution of the problem, another simplex tableau is derived by obtaining identity through the following steps:

   (a) Divide each element of the key row (including b_i) by the key element to get the corresponding values in the tableau. The row so derived is called the *replacement row*, and the values of this row are called *replacement ratios*, in which the value at the key element is 1.

   (b) Get the other elements of key column zero by subtracting each row (other than key row) by replacement row multiplied by the corresponding row element in the key column.

   This process is also known as *pivoting operation* that determines a revised solution.

3. *Iteration* This new simplex tableau represents the new solution which needs to be again subjected to the optimality test in previous manner. If the solution is not optimal, it must be improved in the similar manner again and again until it satisfies the condition for optimality [Eq. (19.9)].

The simplex algorithm is also explained in the section of solved examples of this chapter. Besides the simplex method, several other efficient methods for solving large linear-programming problems have been developed. Few of them are the revised simplex method, the duplex method, the revised duplex method, Dantzig–Wolfe decomposition method. However, discussion on these methods (suitable for large linear-programming problems) is not useful in the present context of the book.

19.1.4 Exceptional Cases

An LPP can have the following exceptional situations:

1. *Degeneracy and Cycling* For n variables and m constraint problems, there would be m basic and $n-m$ non-basic variables. The basic variable would assume positive values. However, during any stage of iteration, one or more basic-variables can become zero. When this happens, it is said that degeneracy had set in. At this point, there is no assurance that the value of object function would improve in the next iteration, and the new solution can remain degenerate or it can become a loop. This is called *cycling*.

2. *Infeasibility* Infeasibility is said to exist when a given problem has no feasible solution. It is evident graphically when no common point is found in the feasible region (two-dimensional) of all the constraints of a problem. In simplex algorithm, the presence of artificial variable at a positive value in the final solution indicates state of infeasibility. Such a problem arises when the constraints are conflicting each other. The infeasibility solely depends upon the constraints and has nothing to do with the objective function [Fig. 19.1].

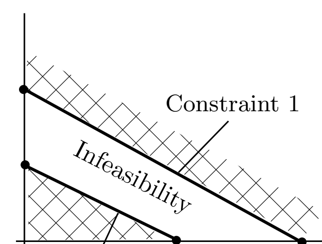

Figure 19.1 Infeasibility in LPP.

3. *Unboundness* *Unboundness* of the objective function occurs when one or more of the decision variables is permitted to increase infinitely without violating any of the constraints. Thus, value of objective function can be increased indefinitely [Fig. 19.2].

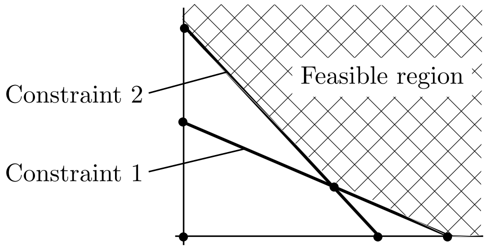

Figure 19.2 | Unboundness in LPP.

In simplex method, the row with the smallest non-negative replacement ratio is selected for the outgoing variable. Absence of non-negative replacement ratio, (all ratios are negative) indicates unbounded solution.

4. *No Feasible Solution* If there is no feasible region, the LPP is said to have no feasible solution [Fig. 19.3].

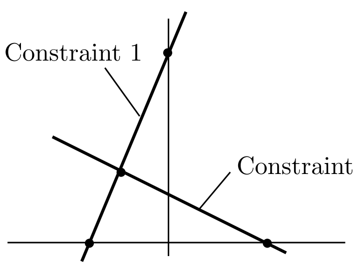

Figure 19.3 | No feasible solution in LPP.

19.1.5 Duality

The optimal solution of maximization type problem can yield complete information about optimal solution of minimization type problem, and vice versa. Then, one is called *primal*, and other is called *dual*. Such an existence of two interdependent problems is called *duality*. When the primal problem is of maximization type, the dual would be of the minimization type.

Every problem, in general form with implied summation, to maximize $c_j x_j$, subjected to $a_{ij}x_j \le b_i$, and $x_j \ge 0$, can be converted so as to minimize $b_i x_i$, subjected to $y_i a_{ij} \le c_j$, and $y_i \ge 0$, using the following conversions for duality:

$$b_i \to c_j, \quad c_j \to b_i, \quad a_{ij} \to a_{ji}, \quad \ge \to \le$$

Let a maximization type problem for an objective function

$$Z = 20x_1 + 30x_2$$

be subjected to constraints:

$$2x_1 + 3x_2 \le 66$$
$$4x_1 + 3x_2 \le 84$$
$$x_1, x_2 \ge 0$$

Its dual is minimization of the objective function

$$Z = 66y_1 + 84y_2$$

and subjected to constraints:

$$2y_1 + 4y_2 \ge 20$$
$$3y_1 + 3y_2 \ge 30$$
$$y_1, y_2 \ge 0$$

19.1.6 Limitations of Simplex Method

Following are the limitations of simplex method.

1. *Linearity* The object function and every constrain must be linear, but in practice, the variations are non-linear.
2. *Additivity* The activities must be additive w.r.t. the measurement of effectiveness and usage of each resources
3. *Divisibility* The solutions are integral.
4. *Deterministic* The coefficients can change with time, and at the time of implementation the solution can become wrong.

19.2 TRANSPORTATION PROBLEM

Transportation problems involve a mathematical approach that produces optimal plan for minimizing the transportation costs of goods and services from several supply centers to several demand centers. Transportation problems can be solved using simplex algorithm, but involve a large number of variables and constraints. Therefore, separate algorithms, such as *stepping-stone method*, *modified distribution method*, have been developed to solve transportation problems.

19.2.1 Problem Definition

Consider a system of n number of sources S_i ($i = 1, 2, 3 \cdots, m$) and m number of destinations D_j ($j = 1, 2, 3 \cdots, n$). Let a_i be the capacity of source S_i and let b_j be the requirement at destination D_j. Let c_{ij} represent the cost of transportation of a unit from source S_i to destination D_j. Let x_{ij} (≥ 0) be the number of units transported from source S_i to destination D_j. The problem is to determine x_{ij} that minimizes the total transportation cost. With these details, a transportation problem can be presented in the following format:

1. Objective function (aggregate transportation cost):

$$Z = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \tag{19.10}$$

2. Constraints (rim requirements):

$$\sum_{j=1}^{n} x_{ij} = a_i \quad \text{for } i = 1, 2, 3 \cdots m$$

$$\sum_{i=1}^{m} x_{ij} = b_j \quad \text{for } j = 1, 2, 3 \cdots n$$

$$x_{ij} \geq 0 \quad \text{for all } i, j$$

This can also be portrayed in a matrix of $m \times n$ size by means of a transportation tableau:

Sources	Destinations (D_j)				Supply
(S_i)	1	2	$\cdots$	n	(a_i)
1	c_{11}	c_{22}	$\cdots$	c_{1n}	a_1
2	c_{21}	c_{22}	$\cdots$	c_{2n}	a_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
m	c_{m1}	c_{m2}	$\cdots$	c_{mn}	a_m
Demand (b_j)	b_1	b_2	$\cdots$	b_n	

The allotment of the transport quantity x_{ij} is written in a similar tableau

Origin	Destinations				Supply
	1	2		n	
1	x_{11}	x_{22}	$\cdots$	x_{1n}	a_1
2	x_{21}	x_{22}	$\cdots$	x_{2n}	a_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
m	x_{m1}	x_{m2}	$\cdots$	x_{mn}	a_m
Demand b_j	b_1	b_2	$\cdots$	b_n	

Before proceeding for the solution, a transportation problem should be looked for the following aspects:

1. *Balanced Transportation Problem* Transportation problem is said to be balanced when the total demand is equal to the total supply:

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j \tag{19.11}$$

2. *Basic Feasible Solution* The number of constraints in the above transportation problems (x_{mn}) is $m+n$ (sum of the numbers of a_i and b_j). The number of variables required for forming a basis is one less, that is, $m+n-1$. This means, with known values of $m+n-1$ variables, the other will be automatically determined by the compatibility derived using Eq. (19.11). Therefore, the number of basic variables is $m+n-1$, and others are non-basic. Thus, a feasible solution of transportation problem of size $m \times n$ is called *basic feasible solution* if it contains no more than $m+n-1$ non-negative allocations.

19.2.2 Solution Procedure

Solution of a transportation problem consists of the following steps.

19.2.2.1 First Infeasible Solution Feasible solution of a transportation problem is defined as a set of non-negative allocations that satisfies the supply and requirement constraints (called *rim requirements*). This can be obtained by north-west corner rule, least cost method, or *Vogel's approximation method* (VAM).

A feasible solution achieved by the Vogel's approximation method[1] yields a very good initial solution which some items can be an optimal solution. This method consists of the following four steps:

1. Enter the cost difference between the two least cost cells, (i.e. smallest and second smallest element in each column) below the corresponding column and do the same in each row. To right of the row, put these differences in bracket.
2. Select the row or column with the greatest difference and allocate as much as possible to the lowest cost (so that the maximum units are transported with minimum cost, at the first sight) while satisfying the supply and requirement constraints. To do this, first allocate the minimum supply and requirement for the lowest cost cell.
3. Reduce supply and demand units by the amount assigned to the cell and cross the completely satisfied column and row. Repeat the above 1 to 3 steps until all allocations have been made.
4. Write down the reduced transportation table after omitting rows or columns.

The Vogel's approximation method is also called the *penalty method* because the cost differences that it uses are the penalties of not using the least cost routes. Since the objective function is the minimization of the transportation cost, in each iteration that route is selected which involves the maximum penalty of not being used.

[1]was developed by William R. Vogel.

19.2.2.2 Testing the Optimality Optimality test can be performed only on a feasible solution, that satisfies following two conditions:

1. Number of allocations is $m+n-1$.
2. All allocations are in independent[2] positions.

If above conditions are not satisfied, it becomes the case of degeneracy. The optimality test can be performed on a feasible solution by either stepping-stone method or modified distribution method. The stepping stone method is applied by calculating the opportunity cost of each empty cell.

The modified distribution (MODI) method, also called *u*-*v* method, is based on the concept of the dual variables that are used to evaluate the empty cells. The test is performed in the following steps:

1. Find u_i and v_j for each row and column such that

$$u_i + v_j = c_{ij}$$

   This can be started by assigning 0 to the first row ($u_1 = 0$).

2. Fill the vacant cell by its $u_i + v_j$ value.
3. Subtract this table/matrix from the original model to find *cost evaluation matrix* (CEM).

The CEM represents the maximum error in allocations in comparison to the optimal solution; the solution could be optimum if this cell should have been chosen for allocation amount equal to the negative value of the cell. A negative cell element of the CEM indicates that the basic solution is not optimal and it can be improved. A feasible solution is optimum if all cells of the CEM are positive.

19.2.2.3 Improving the Solution If the solution is not optimal, rearrangement is made by transferring units from an occupied cell to an empty cell that has the largest opportunity cost but satisfying the rim requirements. The iteration toward an optimal solution is done in the following steps:

1. Identify the CEM cell having most negative entry.
2. Check mark the empty cell on the most negative entry. This cell is called *identified cell*.
3. Assign +ve sign to the identified cell and draw alternating horizontal and vertical lines such that corner is at allocated cells.
4. Assign alternate +ve and −ve signs to corners.
5. Add the absolute of largest negative value of the CEM to all to find next feasible solution.

[2]Independent allocations means it is not possible to increase or decrease any allocation without either changing the position of the allocation or violating the row and column constraint.

Again check for optimality test and iterate towards optimal solution (i.e. until all elements of the CEM become non-negative) in the way mentioned above.

The procedure is also explained in the section of solved examples of this chapter.

19.2.3 Exceptional Cases

A transportation problem can have the following exceptional situations:

1. *Maximization Problem* A transportation tableau can represent unit profits, instead of unit costs, and the objective is the maximization of profit. Such cases are solved by converting the profit matrix into cost matrix, thus, objective function needs to be minimized. This is obtained by subtracting the profit matrix from highest profit value in the matrix. The problem is then solved in usual way, and the objective function is determined with reference to the original profit matrix.

2. *Unbalanced Problem* Transportation problems are solved assuming a balanced problem in which aggregate supply is equal to the aggregate demand. If they are unequal, the problem is known as an *unbalanced transportation problems*.

   Unbalanced problems are solved by converting them into a balance problem through the following changes:

   (a) *Supply* $>$ *Demand* - Such problems are solved by creating a dummy destination having zero cost of transportation.

   (b) *Supply* $<$ *Demand* - Such problems are solved by creating a dummy source having zero unit cost of transportation.

3. *Prohibited Routes* Some routes in a transportation system can be strictly prohibited due to a variety of unfavorable reasons. In such cases, the problem of optimization is solved by assigning a very large cost, represented by M, to each of the prohibited routes, thus prevent the allocations in those routes. The problem is then solved in the usual way.

4. *Degeneracy* In transportation problem, if the number of allocations is less than $(m+n-1)$ then this is the case of degeneracy. In such cases, the solution cannot be improved upon because the algorithm cannot be applied.

19.3 ASSIGNMENT PROBLEM

Resources possess varying abilities for performing different jobs, therefore, the costs of performing those jobs are different. Optimal assignment of resources for production is an essential aspect of production planning and control. An *assignment problem* can be viewed as a reduced or degenerate form of transportation problem obtained by incorporating the following changes:

1. Sources are the assignees and destinations are tasks.
2. Taking demand (b_i) and supply (a_i) as 1, (there will be only one assignment); the units available at each origin and units demanded at each destination are all equal to one.
3. The numbers of origins and destinations should be made exactly equal. This makes the problem a square matrix.

The objective is to minimize the cost of production by optimal assignments of the job to most suitable worker or machine.

19.3.1 Problem Definition

An assignment problem is a $n \times n$ square problem of cost matrix c_{ij} to assign jobs J_i to workers W_j such that working time or cost is minimum. Formulation of this problem requires that the decision variables take only one of the two values 1 or 0, accordingly as an assignment (of a worker to a job) is made or not. The decision variable x_{ij} answers whether ith worker is assigned jth job:

$$x_{ij} = \begin{cases} 1 & \text{true} \\ 0 & \text{false} \end{cases} \tag{19.12}$$

Assignment problem is to minimize the objective function:

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{19.13}$$

This is subjected to the following constraints:

$$\sum_{j=1}^{n} x_{ij} = 1 \quad \text{for } i = 1, 2, 3 \cdots n$$

$$\sum_{i=1}^{n} x_{ij} = 1 \quad \text{for } j = 1, 2, 3 \cdots n$$

$$x_{ij} = 0 \text{ or } 1 \quad \text{for all } i, j$$

The problem can also be portrayed in a $n \times n$ matrix of a transportation tableau:

Workers (W_j)	Jobs (J_i) 1	2		n	x_{ij}
1	c_{11}	c_{22}	$\cdots$	c_{1n}	1
2	c_{21}	c_{22}	$\cdots$	c_{2n}	1
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n	c_{n1}	c_{n2}	$\cdots$	c_{nn}	1
x_{ij}	1	1	$\cdots$	1	

19.3.2 Solution of Problem

For the $n \times n$ assignment problem, the maximum number of possible allocations is $n!$, but only few of them can represent optimal allocations. Because of degeneracy, an assignment problem cannot be easily solved by either simplex algorithm or transportation algorithm. Solutions with simplex methods require $n \times n$ decision variables and $2n$ inequalities as constraints. Solution with transportation problem require $n-1$ dummy allocations. Thus, different algorithms have been developed to solve assignment problems efficiently. Out of these, *Hungarian assignment method*[3] (HAM) is most simple and frequently used. The method is performed through the following steps:

1. *Opportunity Cost Table* Obtain a reduced cost table, known as *opportunity cost table*, through the following steps:
   (a) Find minimum of c_{ij} in each row. Subtract this value in the respective row elements. Thus, each row has at least one zero.
   (b) Find minimum of c_{ij} in each column. Subtract this value in the respective column elements. Thus, each column has at least one zero.
2. *Feasible Solution and Optimality Test* Draw minimum number of horizontal and vertical lines to cover all the zeros. For this, first cover the row or column having maximum number of zeros, and then in descending order. If the number of these lines is equal to n then optimal solution is obtained, and the allocations are made after scanning the rows and columns for single zero. Work is assigned first to the worker having minimum number of zeros in his row, and similarly for all. If solution is not optimal, follow the next step.
3. *Improving the Solution* Find the smallest uncovered elements and subtract it from all uncovered elements and add it to all elements at the intersection of lines. Other elements will not change. Repeat step (2) until an optimal solution is obtained.

[3]Hungarian assignment method was developed by Harold Kuhn in 1955. The algorithm is largely based on the earlier works of two Hungarian mathematicians, so the name comes.

The procedure is also explained in the section of solved examples of this chapter.

19.3.3 Exceptional Cases

An assignment problem can have the following exceptional situations:

1. *Maximization Assignments* Assignments model can be used to optimize the profit or revenue. Such problems can be solved by converting them into minimization type assignment problems. For this, each cell element of the cost matrix (c_{ij}) is subtracted from the largest value of c_{ij} in the matrix. The problem is then solved in usual manner.

2. *Unbalanced Assignments* A assignment problem can be unbalanced (non-square matrix); number of jobs is not equal to the number of workers. Such problems are solved by introducing a dummy worker or job with zero cost (in both cases). The problem is then solved in usual manner.

3. *Multiple Optimal Solutions* Assignment algorithm can face tie situation when it is possible to have two or more ways to cover zeros and assign the job to workers. This indicates existence of multiple optimal solutions but with the same optimal value of the objective function. Selection of optimal solution then depends upon the discretion of the decision maker.

4. *Restricted Assignments* Some cells (workers and jobs) in an assignment problem can be strictly restricted due to a variety of unfavorable reasons. In such cases, the problem of optimization is solved by assigning a very large cost, represented by M, to each of the restricted cell, thus preventing allocations in those cells. The problem is then solved in usual way.

19.4 SEQUENCING

Routing is an important function of production planning. Most of the effectiveness measures, such as time, cost, distance, depend on the order of performing a series of jobs. The jobs can be scheduled by using priority sequencing rules whenever the workstation becomes available for further processing. Some of the priority sequencing rules are enlisted as follows:

1. *First-Come, First-Served Rule* The first-come, first-served (FCFS) rule gives the job arriving at the workstation first the highest priority.

2. *Earliest Due Date Rule* Earliest Due Date (EDD) rule orders the jobs in the order of earliest due date.

3. *Shortest Processing Time Rule* The shortest processing time (SPT) rule orders the jobs in the order of increasing processing times.

However, the priority sequencing rules cannot be applied in large production systems. Theoretically, n number of jobs can be performed on single machine in $n!$ number of ways. Similarly, n number of jobs can be performed through m number of machines or workers in $(n!)^m$ possible sequences. Out of the theoretically possible sequences, only a few or one can be the optimal sequence. The selection of optimal sequence of jobs through a system is called *sequencing*. *Sequencing problem* arises when the system has more than one option of the order in performing a series of jobs.

19.4.1 Problem Definition

Suppose n number of jobs J_i $(i = 1, 2, \cdots, n)$ are to be performed on m number of machines M_j $(j = 1, 2, \cdots, m)$ in order from M_1 to M_m. A machine can process one job at time. Let c_{ij} be the cost (or time) of completion of ith job on jth machine, which is independent of the order of performing jobs. This description can be represented in a tabular format:

	Machines			
Jobs	M_1	M_2	$\cdots$	M_m
J_1	c_{11}	c_{22}	$\cdots$	c_{1m}
J_2	c_{21}	c_{22}	$\cdots$	c_{2m}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
J_n	c_{n1}	c_{n2}	$\cdots$	c_{nm}

The objective of the model is to determine the sequence of jobs (J_i) which will allow execution of all jobs so that the cost or time, from the beginning of the first job till the completion of the last job is minimum.

19.4.2 Solution of Problem

There is no general method of solving $n \times m$ sequencing problem. Therefore, these problems are solved by converting them into $n \times 2$ sequencing problems. Therefore, sequencing of n jobs through 2 machines is discussed first.

19.4.2.1 Processing n Jobs through Two Machines Sequencing of n jobs through two machines can be solved by *Johnson's algorithm*. Tabular form of the $n \times 2$ problem is shown as follows:

	Machines	
Jobs	M_1	M_2
J_1	c_{11}	c_{12}
J_2	c_{21}	c_{22}
$\cdots$	$\cdots$	$\cdots$
J_n	c_{n1}	c_{n2}

Johnson's algorithm consists of the following steps:

1. Make a horizontal box with n number of cells for sequencing of n jobs in a manner that reduces the cost of performing all the jobs.

Sequence	1	2	3	$\cdots$	$\cdots$	n
Jobs						

   Since a job has to be performed first by machine M_1 then by machine M_2, left side of the box will be filled by for minimum c_{i1}'s and right side will be filled for minimum c_{i2}'s.

2. Find the minimum of c_{i1}'s and c_{i2}'s. If c_{i1}'s is minimum out of these, put corresponding J_i in the first empty column from left in the box. If c_{i2}'s is minimum out of these, put corresponding J_i in the first empty column from right in the box.

   If more c_{i1}'s or c_{i2}'s are minimum and equal, then several alternatives of optimal solution are possible by putting corresponding J_i in the box. Value of objective function remains the same.

   A job can be allocated in the box only once. Therefore, an allocated job should be crossed from the table.

3. Repeat the above process until all jobs are fitted in the box. The sequence of J_i's from left to right in the box is optimal sequence.

4. Determine the cost of performing all the jobs in optimal sequence by adding the time taken in the two machines for each job.

The procedure is also explained in the section of solved examples of this chapter.

19.4.2.2 Processing n Jobs through Three Machines

Problems of sequencing n jobs through three machines can be written in the following format:

	Machines		
Jobs	M_1	M_2	M_3
J_1	c_{11}	c_{12}	c_{13}
J_2	c_{21}	c_{22}	c_{23}
$\cdots$	$\cdots$	$\cdots$	$\cdots$
J_n	c_{n1}	c_{n2}	c_{n3}

These types of problems can be solved by converting them into processing of n jobs through two machines. This can be done if any one of the following two conditions is valid:

1. Minimum processing time on machine M_1 is at least greater than the maximum processing time on M_2:

$$\min(c_{i1}) \geq \max(c_{i2}) \quad \text{for } i = 1, 2, \cdots, n$$

2. Minimum processing time on machine M_3 is at least great as the maximum processing time on machine M_2:

$$\min(c_{i3}) \geq \max(c_{i2}) \quad \text{for } i = 1, 2, \cdots, n$$

If any of the above conditions is valid, the three machine problem is converted into two machines problem by taking combined time in M_1 and M_2 as an apparent single machine, and similarly that on M_2 and M_3 to time on another apparent machine.

	Apparent machines	
Jobs	M_{a1}	M_{a2}
J_1	$c_{11}+c_{12}$	$c_{12}+c_{13}$
J_2	$c_{21}+c_{22}$	$c_{22}+c_{23}$
$\cdots$	$\cdots+\cdots$	$\cdots+\cdots$
J_n	$c_{n1}+c_{n2}$	$c_{n2}+c_{n3}$

The problem is then solved like processing of n jobs through two machines.

19.5 QUEUING THEORY

A queue is a waiting line. It is formed when arrival rate of customers is greater than the serving rate during a period of time. Any service system involving queuing situation has to achieve an economic balance between the percentage utilization of server and cost of waiting line. *Queuing theory*[4] utilizes mathematical models and performance measures to assess and improve the flow of customers through a queuing system.

19.5.1 Elements of Queuing Models

A queuing model consists of the following elements [Fig. 19.4]:

[4] Queuing theory was born in the early 1900s with the work of A. K. Erlang, a Danish mathematician, statistician and engineer of the Copenhagen Telephone Company.

1. *Input Process* The queuing model where in customers' arrival times are known with certainty are called *deterministic*. The essence of queuing theory is that it takes into account the randomness of the arrival process and the randomness of the service process. Hence, the distribution of inter-arrival time is seen as per a prescribed probability law, mainly *Poisson distribution*.

2. *Queue Discipline* This is the law according to which the customers form a queue and the manner in which they are served. Following are few examples of queue disciplines:

   (a) First-come first-served (FCFS)
   (b) Last-come first-served (LCFS)
   (c) Service in random order (SIRO)
   (d) Priority service.

   These queue disciplines can be seen in reservation counter (FCFS), government office files (LCFS), treatment of VIP (priority service), etc.

3. *Service Mechanism* Service mechanism is the facility available in the system to serve the customers. *Service time* is the amount of time needed to serve a customer, and *service rate* is number of customer served per unit time. Service times are assumed to be exponentially distributed about some average service time.

4. *Capacity* Capacity is the number of customers who can be in the queue. Some of the queuing systems has limited capacity beyond which customers are not permitted to enter into the system until the space is availed by serving the customers already in the queue.

5. *Notation* Using *Kendall's notation*, the queuing model is denoted in terms of arrival rate distribution (A), service rate distribution (B), number of servers (C), capacity of the system (D) and service discipline (E) as

   $$(\mathrm{A/B/C}) : (\mathrm{D/E})$$

   In this, M denotes Poisson distribution of arrival rate and exponential distributed service rate. For example, a queuing model "(M/M/1) : (∞/FCFS)" means arrival in Poisson distribution, service in Poisson distribution, infinite (∞) customers, and first come first served (FCFS) basis of queue discipline. Also, this is an example of single server, infinite population queuing model.

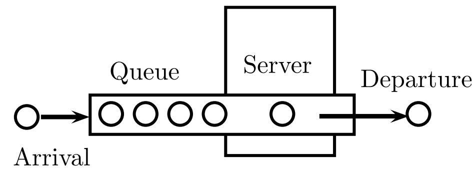

Figure 19.4 Queuing system.

19.5.2 Model (M/M/1):(∞/FCFS)

Consider a (M/M/1) : (∞/FCFS) system in which customer arrive at average rate λ (customers per unit time) and they are served at average rate of μ (customers per unit time). Average utilization, termed as *traffic intensity* (ρ), is expressed as

$$\rho = \frac{\lambda}{\mu}$$

Using the Poisson distribution in arrival of the customers, the probability that n customers will arrive in system during a period of interval t is given by

$$p(n,t) = e^{-m}\frac{m^n}{n!}$$

where $m = \lambda t$ is number of customers arrived in time interval t.

Using exponential distribution in serving the customers, the probability that no more than t time period is needed to serve a customer is given by

$$p(\bar{t}) = 1 - e^{-\mu t}$$

where μt is average number of customer served in time period t.

The following expressions can be derived using the concept of probability:

1. The probability of having exactly n customers in the system:

   $$p_n = \rho^n (1-\rho)$$

   Using this, following expressions can be derived:

   (a) The probability of having exactly one customer in the system:

   $$p_1 = \rho(1-\rho)$$

   (b) The probability that there are no customers in the queue:

   $$p_0 = 1-\rho$$

(c) Expected number of customers in the system:

$$n_s = \sum_{n=0}^{\infty} p_n = \frac{\rho}{1-\rho} = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda}$$

(d) Expected number of customers in the queue:

$$n_q = n_s - \rho = \frac{\rho^2}{1-\rho}$$

2. Mean waiting time in queue:

$$t_q = \frac{\rho}{\mu-\lambda}$$

3. Mean time in the system:

$$\bar{t}_s = \frac{n_s}{\lambda} = \frac{1}{\mu-\lambda}$$

Little's law tells us that the average number of customers in the store n_q is the effective arrival rate λ times the average waiting time in the queue t_q.

$$n_q = \lambda t_q$$

This can be proved as

$$n_q = \frac{\rho^2}{1-\rho} = \frac{\lambda}{\mu} \times \frac{\rho}{1-\lambda/\mu} = \lambda \times \frac{\rho}{\mu-\lambda} = \lambda t_q$$

19.6 PERT AND CPM

Project evaluation and review technique (PERT) and critical path method (CPM) are the *network based techniques* of project scheduling. A *project* is a well-defined task having definable beginning and end points. It requires resources for the completion of the inter-related constituent activities. Project scheduling aims to develop an optimal sequence of activities of the project so that the project completion time and cost are properly balanced and kept at the optimum level. Project scheduling techniques are described as follows:

1. *Critical Path Method* This technique is based on only single estimate of completion time for each activity of the project of repetitive nature, therefore, CPM is an activity-oriented technique.

2. *Project Evaluation and Review Technique* This technique is applicable for the projects having non-repetitive and stochastic in nature. The technique emphasizes on the completion of task rather than the individual activities of project, therefore, it is an event-oriented technique.

   Each activity has three values of time estimates from β-distribution [Fig. 19.5]:

   (a) *Optimistic Time* - The shortest possible time required for the completion of activity is called *optimistic time* (t_o).
   (b) *Most-Likely Time* - The time required for the completion of activity under normal circumstances is called *most-likely time* (t_m).
   (c) *Pessimistic Time* - The longest possible time required for the completion of activity is called *pessimistic time* (t_p).

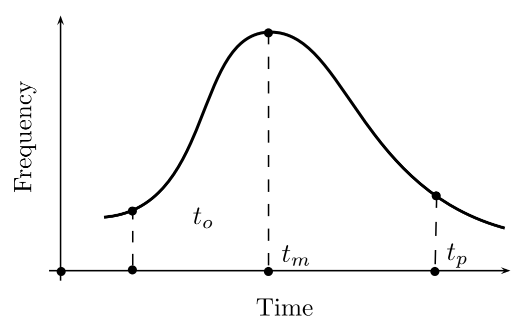

Figure 19.5 β-distribution of time.

PERT utilizes weighted average of the probability distribution of expected completion time of an activity, given by

$$t_e = \frac{t_o + 4t_m + t_p}{6}$$

Variance and standard deviation of the time estimates are determined as

$$v = \left(\frac{t_p - t_o}{6}\right)^2$$

$$\sigma = \sqrt{v} = \frac{t_p - t_o}{6}$$

19.6.1 Project Network Components

Project network is the graphical representation of the project activities and events arranged in a logical sequence and depicting the inter-relationships among them: Thus, a network consists of two major components [Fig. 19.6]:

1. Activity An *activity* is a physically identifiable part of a project which consumes both time and resources. It is represented by an arrow in a network diagram. Tail of an arrow represents the start of activity while head of the arrow represents its completion of the activity. This representation also includes description and estimated completion time over the arrow.

   An activity has two terminal events which can be starting or completion points of other activities. Thus, an activity can have two types of associated activities:

   (a) *Predecessor Activity* - All those activities which must be completed before the start of activity under consideration, are called its *predecessor activities*.

   (b) *Successor Activity* - All those activities which have to follow the activity under consideration are called its *successor activities*.

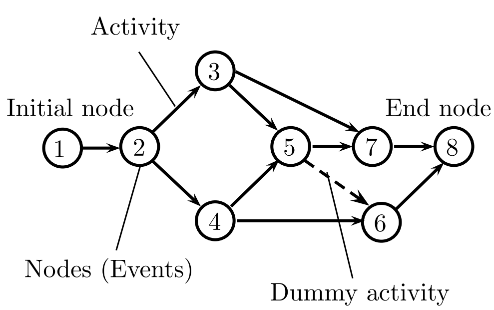

Figure 19.6 Elements of project network.

   A *dummy activity* is used to maintain the pre-defined precedence relationship only during the construction of the project network. It does not consume any time and resource, therefore, it is represented by a dotted arrow. For example, the dummy activity shown in Fig. 19.6 indicates that the activity 4-5 is the predecessor of the activity 6-8.

   Dummy activities can be used to maintain precedence relationships only when actually required. Their use should be minimized in the network diagram.

2. Event Beginning and ending of an activity are represented as *events*. Each event is shown as a node represented by a circle. An unbroken chain of activities between any two events is called a *path*.

Numbers should be so assigned to the events that they reflect the logical sequence of events in the network.

19.6.2 Critical Path

A project network can have numerous paths between the initial event and the last event of the project. Duration of a particular path would be the sum of durations of all the activities lying on that path. The path that has the longest duration is called the *critical path* and the activities that lie on the critical path are called *critical activities*. A delay in any of the critical activities can affect all other succeeding activities. A network can have more than one critical path of the same duration.

Critical path is identified by a systematic procedure comprising two series of computations:

1. Forward Pass Computation This is the method of computation of *earliest start time* (ES) of the events. The computation begins from the initial event ($ES_1 = 0$) and moves towards the final event to arrive at the *earliest start time* of all the events. When two or more arrows terminate at an event, then ES is taken as the maximum value of ES's of all such activities. Earliest finishing time (EF) can be calculated by using

$$EF_{ij} = ES_{ij} + t_{ij}$$

2. Backward Pass Computation This is the method of computation of *latest start time* (LS) of the events. The computation begins from the final event (by assigning LS = EF) and moves towards the initial event to arrive at the *latest start time* of all the events. When two or more arrows terminate at an event, then LS is equal to the minimum of LS's of all such activities. Latest finishing time (LF) is calculated by using

$$LF_{ij} = LS_{ij} + t_{ij}$$

Slack time of an event is the difference between LS and ES. For a critical activity, the earliest start time (ES) and the latest start time (LS) are the same. Therefore, the path connecting events with zero slack is the critical path. Since ES = LS, a succeeding activity in a critical path shall commence immediately after its proceeding activity is completed. If there is any delay in either starting or completion of a critical activity, the project implementation period will get extended.

The procedure is explained in the section of *solved examples* of this chapter.

19.6.3 Activity Float Analysis

The free time available for an activity is called *float*. All paths in a network other than the critical path are called non-critical paths. If the activities in non-critical path are so delayed that they exceed the duration of the critical path, the overall project completion time will get delayed. Hence, a detailed study of non-critical activities with regard to the float is worth doing since it helps in better control of the project implementation and allocation of resources.

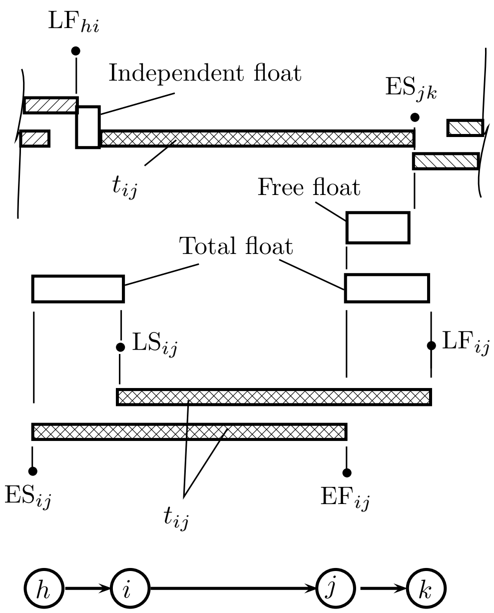

Figure 19.7 | Activity float analysis.

An event has only one estimate of slack but an activity has three types of float estimates [Fig. 19.7]:

1. Total Float *Total float* of an activity is the time by which an activity can be delayed without affecting the project completion time:

$$\begin{aligned}\text{Total float}_{ij} &= \text{LS}_{ij} - \text{ES}_{ij}\\ &= \text{LF}_{ij} - \text{EF}_{ij}\end{aligned}$$

2. Free Float *Free float* of an activity is the delay that can be permitted in an activity so that succeeding activities in the path are not affected. For this, the earliest start time of the head event of the activity shall not exceed. Therefore, free float of an activity ij is

$$\text{Free float} = \text{ES}_{jk} - \text{EF}_{ij}$$

3. Independent Float *Independent float* of an activity is the spare time available for the activity if its preceding activities are completed at their latest, and its succeeding activities start at their earliest. Therefore, independent float of an activity ij is

$$\text{Independent float} = \text{ES}_{jk} - \text{LF}_{hi} - t_{ij}$$

19.6.4 Time-Cost Trade-Off Analysis

Project cost is linked with the requirements of time and resources in completing a project. Project costs are generally a function of time. The shorter the period, lesser the overhead charges, but it would increase the direct cost.

The *critical path method* in network analysis can evaluate the alternate ways to expedite some of the activities and then analyze their effect in the cost of the project. This analysis is referred as the *time-cost trade-off analysis*.

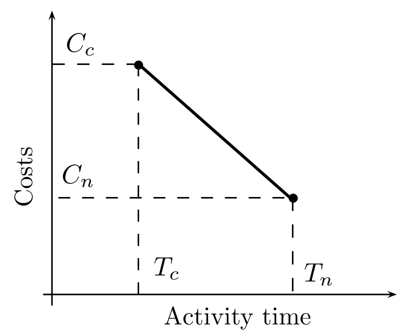

Figure 19.8 | Incremental cost.

Crashing is a process for reducing the duration of an activity by allocating more resources. This can affect the critical path. The minimum possible duration of an activity is called *crash duration* (T_c), and corresponding cost of the activity is called *crash cost* (C_c).The crash cost of an activity is always greater than the normal cost C_n. Assuming linear approximation [Fig. 19.8], the *incremental cost* is the slope of crashing an activity:

$$\text{Incremental cost} = \frac{\text{Crash cost} - \text{Normal cost}}{\text{Normal time} - \text{Crash time}}$$

An activity with the least incremental cost should be crashed on priority. Crashing is preferred only for critical activities. Crashing of non-critical activities would not reduce the project duration.