Chapter-1

INTRODUCTION

This thesis analyses certain problems in Inventories and Queues. There are many situations in real-life where we encounter models as described in this thesis. It analyses in depth various models which can be applied to production, storage, telephone traffic, road traffic, economics, business administration, serving of customers, operations of particle counters and others. Certain models described here is not a complete representation of the true situation in all its complexity, but a simplified version amenable to analysis. While discussing the models, we show how a dependence structure can be suitably introduced in some problems of Inventories and Queues. Continuous review, single commodity inventory systems with Markov dependence structure introduced in the demand quantities, replenishment quantities and reordering levels are considered separately. Lead time is assumed to be zero in these models. An inventory model involving random lead time is also considered (Chapter-4). Further finite capacity single server queueing systems with single/bulk arrival, single/bulk services are also discussed. In some models the server is assumed to go on vacation (Chapters 7 and 8). In chapters 5 and 6 a sort of dependence is introduced in the service pattern in some queueing models.

This chapter reviews briefly some of the important developments in Inventories and Queues. It also explains the technical terms and notations used in this thesis. Further a brief outline of the work on which this thesis is based is also given towards the end of this chapter.

1.1. Historical background - Inventory Theory

The study of the quantitative analysis in inventory systems is considered to be originated with the work of Harris (1915) and he obtains a formula for the optimal production lot size given by the square root function of the fixed cost, holding cost and the demand. This formula referred to as the economic order quantity (EOQ) is popularised by Wilson. After World War II several authors have discussed the stochastic behaviour of the inventory in the case of scheduling the use of stored water to minimise the cost of supplying electric energy. Pierre Masse (1946), a French engineer is considered to be the first to achieve a satisfactory result regarding this problem. Arrow, Harris and Marschak(1951) have showed that the total expected cost incurred from use of an (s,S) policy satisfies a renewal equation. Further Dvoretzky, Kiefer and Wolfowitz (1952) have given some sufficient conditions for establishing that the optimal policy is an (s,S) policy for the single-stage inventory problem. A detailed account of the developments that have taken place till 1952 is given by Whitin (1953). Bellman, Glicksberg and Gross(1955) determine the optimal policy for the case in which the ordering and penalty cost are both linear. Gani (1957) studies some problems arising in the stochastic theory of storage systems.

A systematic account of (s,S) inventory type is first provided by Arrow, Karlin and Scarf (1958). Their approach is based on renewal theory. It is natural to enquire how these models could be applied in practical situations. Hadley and Whittin (1963) provides an excellent account of the applications. A lucid survey of this field through 1962 is given by Scarf (1963). A complete computational approach for finding optimal (s,S) inventory policies is developed by Veinott and Wagner (1965). There is an excellent review by Veinott (1966) which summarizes the status of mathematical inventory theory. He focusses his attention on the determination of optimal policies of multi-item and/or multi-echelon inventory systems with certain and uncertain demands. Hurter and Kamisky (1967) find the limiting distribution of the number of units in the storage for a basic single commodity storage system by applying the theory of regenerative stochastic processes. The cost analysis of different inventory systems is given in Naddor (1966). Kaplan (1970) and Gross and Harris (1971) also make distinct contributions in these directions. Inventory systems with random lead time is discussed by

Ryshikov (1973) in his monograph.

Sivazlian (1974) considers the case of a continuous review inventory system with unit demand, zero lead time and arbitrary interarrival times of demands. He obtains the transient and steady state distribution for the position inventory and shows that the limiting distribution of the position inventory is uniform and is independent of the interarrival time distribution. Richards (1975) proves the same result for the case with random demand size. Srinivasan(1979) extends the result of Sahin (1974) to the case in which lead times are i.i.d random variables following a general distribution. This is further extended by Manoharan, Krishnamoorthy and Madhusoodanan (1987) to accommodate the case of non-identically distributed interarrival times.

An (s,S) inventory system with demand for items dependent on an external environment is studied by Feldman (1978). Constant lead time (S,s) inventory policy with demand quantities forming non-negative real valued random variables is anslysed by Sahin (1979). Ramaswami (1981) obtains algorithms for an (s,S) model where the demand is according to a versatile Markovian point process. Further, Sahin (1983) obtains the binomial moments of the time dependent and limiting distributions of the deficit in the case of a continuous review (s,S) policy with random lead time and demand process following a compound renewal process. Single product inventory systems relating to production process is seen in the works of De kok, Tijms and Van der Duynschouten (1984).

Thangaraj and Ramanarayanan (1983) discuss an inventory system with random lead time and having two reordering levels. Again Ramanarayanan and Jacob (1986) consider the same problem with varying reordering levels; but in their model passage to the limit is rather difficult. Also inventory system with varying reordering levels and random lead time is discussed by Krishnamoorthy and Manoharan (1991). They obtain the time dependent probability distribution of the inventory level and the correlation between the number of demands during a lead time and the length of the next inventory dry period.

A review of the work done in perishable inventory until 1982 can be had from Nahmias (1982). Kalpakam and Arivarignan (1985) consider the case of an inventory system with arbitrary interarrival time between demands in which one item is put into operation as an exhibiting item (they have assumed that an exhibiting item has exponentially distributed life time) and obtain the transient and steady state distributions for position inventory. Again the same system having one exhibiting item subject to random failures with failure times following exponential distribution and unit demand is dealt by the same authors (1985) and the expression for the limiting distribution of the position inventory is derived by applying the techniques of semi-regenerative process. Manoharan and Krishnamoorthy(1989) consider an inventory problem with all items subject to decay and derive the limiting probability distribution. In this the quantities demanded by arrivals are i.i.d.r.vs and interarrival times have a general distribution.

Ramanarayanan and Jacob (1987) analyses an inventory system with random lead time and bulk demands. They use the matrix of transition time densities and its convolutions to arrive at the expression for the probability distribution of the inventory level. Inventory system with random lead times and server going on vacations when the inventory becomes dry is introduced by Daniel and Ramanarayanan (1987, 1988). Jacob (1988) deals with bulk demand inventory models and server vacation. Further Krishnamoorthy and Manoharan (1990) investigate an inventory problem in which the quantities demanded by successive arrivals are assumed to follow distributions depending on the availability of the items. They obtain the limiting distribution of the inventory level. A stochastic inventory system with Poisson demand and exponentially distributed delivery time is discussed by Beckmann and Srinivasan (1987).

1.2. A brief account of the Inventory theory

An Inventory is a measured stock of some goods which is held or stored for the purpose of future sale or production. So it varies in quantity over time in response to a'demand' process which operates to diminish the stock and a 'replenishment' process which operates to increase it. The obvious applications to stocks of physical goods are light bulbs, raw materials to be used in some production process etc. whereas the number of engineers employed by a company, the number of students enrolled in a college or the amount of equity capital available for corporate growth are all regarded as inventory. When production is involved, the inventory problem might require, for example, determining how much wheat to plant per year or how much gasoline of certain variety to have blended. The amount of water to be released from a dam for electricity and irrigation purposes is also an inventory problem. Again inventory problems may involve scheduling, production, determining efficient distribution of commodities in certain markets, finding proper replacement policies for old equipment, determining proper prices for goods produced, or combinations of these elements.

Demand

Inventories are held for the ultimate purpose of satisfying demands. Usually the demand is not subject to control, but the timing and magnitude of the replenishments may be regulated.

Various models of natural attrition comprise what we call the demand process and the hiring or recruitment constitute the replenishment process. Inventory theory is concerned with the analysis of several types of decisions relating primarily to the problem of when to buy and how much to buy of a given item. The analysis involves consideration of when the item should be manufactured and problems of transportation and distribution of stock etc.

Motivation for Inventory

(a) Inventories are frequently held because of economies of scale in production or procurement. If the average cost of purchasing stock decreases when larger quantities are purchased, then it is economical to purchase in relatively large quantities. The result is the accumulation of stock prior to actual need.

(b) The requirements for items may vary substantially over time and this itself may serve as an incentive for holding stock. It is advantageous to procure the item before it is needed at a lower marginal cost, thus contributing to the formation of inventories. This motive for holding inventories will be reinforced if the cost function displays decreasing average cost.

(c) Another motive for holding stock is that the costs may themselves be a function of time.

(d) Uncertainty of future requirements is also a strong motive for holding inventories.

Inventory policies and objective function

In an inventory problem that lasts for some length of time, cost will generally be incurred at various moments of time. The main costs involved are: (i) the ordering cost which is composed of a cost proportional to the amount ordered plus a set up cost which is constant when the amount z ordered is positive and zero for $z=0$, (ii) storage cost or holding cost which is incurred by the actual maintenance of stocks or the rent of storage space or a measure of obsolescence or spoilage. The cost of repairing a defective item is also considered as a storage cost, (iii) penalty cost or shortage cost which arises when supply including both current output and accumulated stocks from the past, exceeds demand. If a demand occurs beyond the available inventory, it is met by a priority shipment or it is backlogged and satisfied when the commodity becomes available. These costs involved in the inventory are to be summarised to a single number so that alternative policies can be compared. An inventory policy is a set of rules that defines when and how much quantity to be ordered.

When any inventory model is investigated first we analyse the model to get the inventory equation which represents the inventory level at any instant of time. The purpose of obtaining the inventory equation is to determine the optimal policy. A policy is called optimal if it maximises the objective function when the objective function is a profit function or minimises the total expected cost per unit time if the objective function is a cost function. Several policies can be used to control an inventory system; but if it is known before hand that the policy has a particular form, then the time to compute optimal policies can be cut substantially. The most widely used policy is the (s,S) policy where the variables s and S are the two decision variables. The variable s is referred to as the reorder level while the variables s and S together stand for how much quantity to be ordered. Whenever the inventory position is equal to or less than s for the first time after a replenishment, a procurement or replenishment is made to bring the inventory to its maximum capacity S.

An inventory system can be either a continuous review or a periodic review system. In a continuous review policy the inventory position is monitored continuously over time whereas this is done at specified points of time in a periodic review system. We concentrate only on continuous review single commodity inventory systems.

An important element in the mechanism of inventory process is the lag in delivery of the commodity after an order is placed or decision is made to produce. This time lag is called the lead time. If replenishments take place instantaneously we say lead time is zero so that the possibility of penalty cost may not occur. In some cases lead time is fixed whereas in others it is a random variable with known distribution. The time interval for which the inventory is empty is termed as a dry period.

1.3. Queueing Theory

Queueing theory had its origin in the pioneering work done by Erlang (1909) on the application of probability theory to telephone traffic problems. It soon drew attention of many probabilists. We can have a queue of broken-down machines waiting for repair at a repair shop, a queue of customers at a store cash counter or a queue formed by planes circling above an air port waiting to land. These provide obvious examples of queues. Often we have cases where a physical queue is absent, such as the waiting list of passengers for a railway or air line ticket or of persons who register their names for the purchase of a car which is not readily available and is to be supplied from future production. So queueing is a mechanism that is used to handle congestion.

A system consisting of a service facility, a process of arrival of customers who wish to be served by the facility, and the process of service is called a queueing system. A queue or waiting line develops whenever the service facility cannot cope with the number of units requiring service. The units arriving for service are called customers in a generic sense. Thus a queueing system is regarded as an arrangement where the customers requiring service form the input, the serviced customers the output and the service rendered the transformation process.

Historical review

Since the work of Erlang (1909) with telephone engineering, applications have expanded into several areas. Interesting and fruitful interactions between theoretical structures and practical applicationshave led to the rapid development of the subject in areas like production planning, inventory control and maintenance problem. For about two decades various researchers and practitioners have looked at models either to solve particular problems at hand or to develop understanding of the stochastic processes that arise from them.

In any analysis of a queueing system one or more aspects like the queue length, the waiting time and the busy period are studied through their probability distributions, from which moments like mean, variance etc. can be obtained. For an ordinary M/M/1 queueing system, the system size probabilities are obtained by solving difference-differential equations. But for most of practical applications of the queueing model, a steady state or a state of statistical equilibrium solution is necessary. The time dependent or transient solution is first given by Bailey (1954 b) making use of generating function whereas Ledermann and Reuter (1956) obtain the solution with spectral theory. While Champernowne (1956) uses combinatorial method,Conolly (1958) uses difference equation techniques for the time dependent solution to an M/M/1 system. Pegden and Rosenshine (1982) also deal with the transient solution of M/M/1 queues. Parthasarathy (1987) provides a very simple and elegant approach to obtain the time-dependent solution to the M/M/1 queue. Again Parthasarathy and Sharafali (1989) extends this to the M/M/s queue. Syski (1988) shows that the result of Parathasarathy (1987) is equivalent to that obtained by Cohen (1982).

For an M/M/1 queueing system we do not have to take into account the time since the last arrival or the elapsed service time of the unit in service because the negative exponential distribution possesses the Markovian or the forgetfulness property and so the queue length process is Markov.

Several methods are available for the analysis of non-Markovian processes. These include:

(i) the use of regeneration points ie. of an embedded Markov chain. The behaviour is considered at a discrete set of time instants chosen in such a way that the resulting process is Markovian.

(ii) Erlang's method, in which life (service time) is divided into fictitious stages such that the time spent in each stage follows an exponential distribution

(iii) supplementary variable technique, whereby the inclusion of sufficient supplementary variables such as expended life-time, in the specification of the state of the system to make the whole process Markovian in continuous time.

The system size process, at arbitrary time points, in M/G/1 and GI/M/1 queueing systems are in general a non-Markovian processes. For an M/G/1 queue the successive departure instants constitute regeneration points whereas for a GI/M/1 queue the successive arrival epochs are the regeneration points. Thus a Markov chain is embedded at these regeneration points. Kendall (1951, 1953) makes use of this method. For an M/M/1 system, all time points are regeneration points so that the whole process in continuous time is Markovian. Cox (1955) uses the supplementary variable technique to analyse non-Markovian stochastic processes. The method of supplementary variable investigated by Cox (1955) is found in the thesis of L. Kosten in 1942. Lavenberg (1975) derives an expression for the Laplace-Stieltjes transform of steady-state distribution of the M/G/1 queueing systems.

Queueing systems with server vacation arise in many computer, communication, production and other stochastic systems. Welsch (1964) characterises the transient and asymptotic distributions of the queue size, waiting time and waiting-plus service time of an M/G/1 queue in which he assumes that the first customer arriving when the server is idle has a distribution different from that when the server is busy. Miller (1964), Avi-Itzhak, Maxwell and Miller (1965), Cooper (1970, 1981), Levy and Yechiali (1975), Heymann (1977), Shantikumar (1980, 1982), Scholl and Kleinrock (1983), Ali and Neuts (1984), Doshi (1985) all deal with vacation models. An extensive survey of the queueing system with vacation to the server is given by Doshi (1986). Daniel (1985) studies some queueing models with vacation to the server where the server takes rest either after serving a certain fixed number of customers or whenever the system becomes empty, whichever occurs first. A finite capacity M/G/1 queue with server vacation is considered by Lee (1984) where the vacation is initiated if either the queue is empty or M customers have been served during a busy period. Manoharan and Krishnamoorthy (1989) also consider a model similar to Lee (1984) and obtain the time dependent queue size distribution and virtual waiting time distribution. Ramachandran Nair (1988) analyses extensively queues with vacation to the server after serving a random number of units. Jacob (1988) and Madhusoodanan (1989) deal extensively with several queueing models with server vacation and derive their time dependent behaviour.

Over the past two decades steady progress has been made towards solving increasingly difficult and realistic queueing models. Lack of results suited for ready practical implementation is observed in several areas in queueing theory. One such class of models is distinguished by the presence of a specified feature, namely, that customers arrive in groups of random size and are served in groups that are themselves of random size. Queueing models belonging to the above category are termed "bulk queues" in literature.

Bailey (1954a) is the first to carry out the mathematical investigation of queues involving batch service. He studies the stationary behaviour of the system in terms of probability generating function. This is followed by a series of papers with group arrival and/or batch service. Gaver (1959) seems to be the first to handle queues involving group arrivals. He is followed by Jaiswal (1960, 1962). Saaty (1961) provides an excellent account of some of these works. Miller (1959) is the first to examine a queueing system in which customers arrive in groups and are served in groups. He obtains the stationary distribution of the number of units in the system making use of embedded Markov chain method. Bhat (1964) studies the equilibrium behaviour of the $M^X/G^Y/1$ and the $GI^X/M^Y/1$ systems using fluctuation theory. Again bulk service queue with infinite waiting room is investigated by Bhat (1967) to obtain the busy period and the busy cycle distribution of the queue length process. Further Teghum, Loris-Teghum and Lambotte(1969) also deal with bulk arrival, bulk service queueing model. Chaudhary and Templeton (1981) obtain the limiting behaviour of an $M/G^B/1$ queueing system. The books by Chaudhary and Templeton (1984) and Medhi (1984) give a detailed account of the work done in bulk queues. Jacob (1988) and Madhusoodanan (1989) also deal extensively with several bulk service queueing models. Morse (1955), Takacs (1961, 1962), Cohen (1969), Prabhu (1965, 1980), Gnedenko and Kovalenko(1968), Cooper (1972), Gross and Harris (1974, 1984), Bagchi and Templeton (1972) and Asmussen (1987) analyse in depth several queueing problems.

Liu, Kashyap and Templeton (1987) deal with an infinite server queueing system providing both individual service and batch service and obtain the transient results for the first two moments of the system size distribution. Waiting time distribution and steady state results are also computed by them.

Another important feature of a bulk queue is that the system follows a general bulk service rule with range (a,b) and with or without vacation. In 1942 Kosten discussed a deterministic service time system with capacity range (a,∞). Further Neuts (1967), Borthakur (1971 a,b), Medhi (1975, 1984), Holman et. al. (1981), Kambo and Chaudhary (1982), Easton and Chaudhary (1982), Chaudhary and Templeton (1981) all consider bulk service queueing system with range (a,b). Fabens (1961, 1963) studies the transient state of the system by identifying the underlying semi-Markov process. Most of these works require the application of Rouche's theorem. Neuts (1979) develops an algorithmic method for the solution of $M/G^{a,b}/1$ system. His approach involves only real arithmetic and avoids the calculation of the complex roots based on Rouche's theorem. Cohen(1982) seems to be the only author to have developed waiting time results for bulk arrival and bulk service queue where the server becomes idle when the system is empty. His results are given in terms of integrals. Most of the above mentioned works concentrate on the steady state behaviour of the system. Jacob, Krishnamoorthy and Madhusoodanan (1988) obtain the time dependent solution to $M/G^{a,b}/1$ queue with finite capacity and the same model with server vacation is analysed by Jacob and Madhusoodanan (1987). Manoharan (1990) extends their result to $E_k/G^{a,b}/1$ queue with server vacation. Transient solution and virtual waiting time distribution are discussed by him. He also considers a queueing situation where the service is carried out either singly or in batches depending upon the number of customers waiting for service in the waiting room. Steady state behaviour of the system is examined by him.

Another notable feature in the queueing system is the state dependence of the service characteristics. Hiller et.al. (1964), Gupta (1967) and Rosenshine (1967) examine queueing systems in which the service rates are an instantaneous functions of the system state. Harris (1967) considers the standard M/G/1 system in which the service time parameter is a random variable dependent upon the state of the system at the moment the customer's service is begun. Murari (1969) and Harris (1970) discuss bulk arrival queue with state dependent service rate. Ponser (1973) investigates a queueing model in which the service time of a customer depends upon his waiting time in the queue and at the same time independent of all other parameters associated with the system size. Shantikumar (1979) discusses a class of queueing models in which the service time of a customer at a single server facility is dependent on the queue size at the onset of its service. He extends Harris's two state, state dependent service to M/G/1 queue.

1.4. Relation between Queues and Inventories

Applications of and fruitful connections between queueing theory and inventory theory occur numerously. Steady progress has been made to solve problems which are difficult but realistic in inventory and queues. Similarities between the mathematical formalisms of both models have been observed from early times.

The amount of goods or material held in stock for future purpose can be identified as a group of customers waiting for some sort of service at a service facility. The arrival of an order or a demand for an item is likened to a service completion since such an arrival or demand results in the departure of a customer in the queue which corresponds to the depletion of the inventory level. The demand for an item to an inventory arrives singly or in batch of fixed or variable size. The bulk demand corresponds to the bulk arrival in queueing theory and single demand that of single arrival. The interarrival times of demands are regarded as the service time.

A better correspondence between an inventory system and a queueing system is seen by regarding the demands occuring to the inventory system as the arrival of customers to the queue because both of these are more or less uncontrollable. The inventory replenishment time or leadtime can be compared to the service time of the queueing system and both of these are, in general controllable by the management of the system.

1.5. Renewal process

Let $\{X_n, n=1,2,...\}$ be a sequence of non-negative independent and identically distributed random variables with $X_1, X_2, X_3, ...$ representing the times between successive occurrences of a fixed phenomenon. Then $S_o=0$; $S_{n+1}=S_n+X_{n+1}$, $n=0,1,2,...$ define the times of occurrence of 1st, 2nd,... events, assuming that the time origin is taken to be an instant of such an occurrence. Then S_n's are called renewal times.

Let F(.) denote the distribution of the interrenewal times. Assume that $Pr\{X_o=0\} < 1$. Since X_n's are non-negative $E(X_n)$ exists.

Define N(t) as $Sup\{n|S_n \leq t\}$. Then the process $\{N(t), t \geq 0\}$ is called a renewal process or a counting process. Obviously the state space of the renewal process consists of a single element. The random variable $N(t)$ gives the number of renewals in the interval $(o,t]$. The distribution of S_n is given by $Pr\{S_n \leq x\} = F_n(x)$, where $F_n(x) = F^{*n}(x)$, (since X_i's are i.i.d random variables) and $F^{*n}(.)$ denotes the n-fold convolution of $F(.)$ with itself. $(F^{*o}(.) \equiv 1)$.

It is easily verified that

$$N(t) \geq n \iff S_n \leq t$$

so that the distribution of $N(t)$ is

$$Pr\{N(t) = n\} = F^{*n}(t) - F^{*(n+1)}(t).$$

Using this distribution, the expected number of renewals in $(o,t]$ denoted by $M(t)$ is given by

$$M(t) = E[N(t)] = \sum_{n=1}^{\infty} F^{*n}(t)$$

$M(t)$ is called the renewal function.

Consider a stochastic process $Z = \{Z(t), t \geq 0\}$ with state space E. Assume that every time a certain event occurs, the future of the process Z after that time is a probabilistic replica of the future after time 0. Such times are called regeneration times of Z and the process Z is said to be a regenerative process. If $T_1, T_2, T_3, \ldots$ constitute a sequence of regeneration points, then $\{T_n, n=1,2,....\}$ forms a renewal process and the time between successive renewal points is called a cycle of the process. Cox and Smith (1961), Cox (1962), Feller (1965), Ross (1975), Cinlar (1975 b), Bhat (1984) give a detailed account of renewal theory.

1.6. Semi-Markov and Markov renewal process

Consider a stochastic process which moves from one state to another of a countable number of states in such a way that the successive states visited forms a Markov chain. Assume that the process remains in a given state for a random length of time whose distribution depends upon the state being visited and the one to be visited next. Such a process is defined as a semi-Markov process since it is a Markov chain with the time scale being randomly selected. Thus a semi-Markov process identifies or gives the state of the process at each time point. For the same stochastic process, let $N_i(t)$ denotes the number of transitions or renewals into the state i (E be the state space of the Markov chain) which occur in (o,t]. Set

$$N(t) = ((N_1(t), N_2(t), \ldots)$$

Then the stochastic process $\{N(t), t \geq 0\}$ is a Markov renewal process. Thus a Markov renewal process is a counting process which records at each time point t the number of times each of the possible states have been visited. Such a process becomes a Markov process if the sojourn times are all exponentially distributed independent of the next state to be visited; it reduces to a Markov chain if sojourn times are all equal to one, and becomes a renewal process if there is only one state. This means that a stochastic process $\{(X,T)\}=\{(X_n,T_n),\ n\in N\}$ defined over a finite set E is a Markov renewal process if

$$\Pr\{(X_{n+1}=j;\ T_{n+1}-T_n\leqslant t \mid X_o,X_1,\ldots,X_n;T_o,T_1,\ldots,T_n\}$$

$$= \Pr\{X_{n+1}=j;\ T_{n+1}-T_n\leqslant t \mid X_n\} \text{ for all } n\in N \text{ and } i,j\in E \text{ and } t\geqslant 0 \qquad (1)$$

Denote the R.H.S. of (1) by $Q(i,j,t)$, if $X_n=i$.

Clearly

$$Q(i,j,t) \geqslant 0;\quad i,j\in E;\ t\geqslant 0$$

$$\sum_{j\in E} Q(i,j,\infty) = 1$$

The family of probabilities

$\mathbb{Q} = \{Q(i,j,t),\ i,j\in E;\ t\geqslant 0\}$ is called a semi-Markov kernel.

For this Markov renewal process, the expected number of returns to state j in an amount of time t given that the system has started from state i is the Markov renewal function $R(i,j,t)$

which is given by

$$R(i,j,t) = \sum_{n=o}^{\infty} Q^{*n}(i,j,t) \quad \text{where}$$

$$Q^{*(n+1)}(i,j,t) = \sum_{k \in E} \int_o^t Q(i,k,du)\, Q^{*n}(k,j,t-u) \quad \text{for } n \geq 0$$

and

$$Q^o(i,j,t) = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}$$

Define the process $Y = \{Y(t),\ t \geq 0\}$ with state space E by $Y(t) = X_n$ for $T_n \leq t < T_{n+1}$. Then the process $\{Y(t),\ t \geq 0\}$ is called the semi-Markov process defined over the state space E with the semi-Markov transition kernel $\mathbb{Q} = \{Q(i,j,t)\}$. Thus the semi-Markov process Y provides a picture which is convenient in describing the Markov renewal process underlying it.

Markov renewal equations

Let (X,T) be a Markov renewal process defined over a finite state space E with the semi-Markov kernel Q(i,j,t) and Markov renewal function R(i,j,t), $i,j \in E$, $t \geq 0$. Let R_+ and R denote the set of non-negative real numbers and real numbers respectively. Assume that f is a function defined by $f: E \times R_+ \longrightarrow R$ such that for every $i \in E$, the mapping $t \longrightarrow f(i,t)$ is Borel measurable and bounded over finite intervals. Let $\mathcal{F}$ be the class of functions f. Then a function $f \in \mathcal{F}$ is said to satisfy the Markov renewal equation if f(i,t) can be written as

$$f(i,t) = g(i,t) + \sum_{j \in E} \int_0^t Q(i,j,du)\, f(j,t-u), \quad i \in E,\ t \in R_+ \qquad (2)$$

for some function $g \in \mathcal{F}$. Here Q(i,j,t) and g(i,t) are known and so the problem is to solve for f(i,t). Further the Markov renewal function (2) has one and only one solution given by

$$f(i,t) = \sum_{j \in E} \int_0^t R(i,j,du)\, g(j,t-u), \quad i \in E,\ t \in R_+$$

Levy (1954) and Smith (1955) independently introduced semi-Markov processes. A detailed description of the Markov renewal process is given in Pyke (1961 a,b). Cinlar (1969, '75 a,b) provide a detailed account of Markov renewal and semi-Markov processes. Inventory and queueing models based on the theory of semi-Markov process is studied by Fabens (1961, '63). Further Schal (1971) analyses M/G/1 and G/M/1 queues and obtains their assymptotic behaviour and rates of convergence. His approach is also based on the theory of semi-Markov process.

1.7 A brief account of the results in this thesis

The aim of the thesis is to study the time-dependent and steady state behaviour of certain problems in Inventories and Queues. This is achieved by identifying the underlying semi-Markov processe and the embedded Markov renewal process of the basic process. It is assumed that the inventory (assumed to be single commodity) is continuously monitored over an infinite horizon period. In the case of some of the problems discussed we have analysed certain control problems associated with them. All queueing problems investigated deal with finite capacity.

Chapter 2 deals with an (s,S) inventory policy where each arrival demands a random number of items, the maximum size being a with $a \leq s$. We assume that the successive quantities demanded form a Markov chain. Replenishment is instantaneous and the quantity replenished is such that the inventory is brought back to its maximum capacity S. The probability distribution of the stock level at arbitrary time points and also the steady state inventory level distribution are obtained. The optimal value of the pair (s,S) is computed.

In chapter 3, the dependence structure is introduced in the (s,S) inventory problems in two different ways. Model I discusses a bulk demand inventory policy with the successive quantities replenished forming a Markov chain. Model II studies a unit demand (s,S) policy with the successive reorder levels varying according to a Markov chain. In Model II, the replenishment quantity is always equal to $M=S-s$. In both Models lead time is assumed to be zero. The inventory level at arbitrary time point and its limiting distribution are computed for both Models. Some control problems associated with the Models are investigated.

Some numerical illustrations are provided at the end of chapters 2 and 3.

Chapter 4 considers a bulk demand inventory problem with zero lead time and the server taking vacation each time the inventory becomes dry after the previous replenishment. The system size probabilities and the reliability of the system at arbitrary time epochs are obtained.

Chapter 5 introduces a class of finite capacity single server queueing models in which the server offers a random number of stages of service to each unit depending upon the system size at the onset of its service. A three dimensional Markov chain with the first coordinate representing the system size, the second one representing the number of stages of service given to the unit undergoing service and the third one denoting the number of stages of service completed by the unit underoing service is identified. The system size probabilities and the limiting distributions are computed. Numerical illustration is also provided.

Chapter 6 generalises the $M/G^{a,b}/1$ queueing system with finite capacity. The services are in batches of sizes between a and b and is such that the size of a batch to be served is determined based on the time taken to serve the previous batch. System size probabilities and steady state analysis are carried out. Distribution of the busy period and the busy cycle are studied. Virtual waiting time distribution is also derived. A control problem associated with the model is discussed.

In chapter 7, we consider two cases of single server queueing systems of finite capacity. Model I discusses a $G/E_k/1$ queueing system whereas Model II investigates a queueing system of general bulk service rule with batch size varying from a to b. Expressions for the time dependent system size probabilities at arbitrary time point for Model I and II, Limiting distribution for Model I and virtual waiting time distribution for Model II are obtained.

Chapter 8 discusses a bulk arrival,bulk service queue of finite capacity b. We assume that a service commences only when the system is full and then only a random number of units are taken for service. On completion of the service of a batch if the system is not full, the server goes for vacation of random duration. System size probabilities are computed. In this model the time duration for which the system remains non-empty continuously is defined as the busy period of the system. Expressions for the distribution of the above defined busy period gives an upper bound for the virtual waiting time. By restricting b=2, the virtual waiting time at time t is computed.