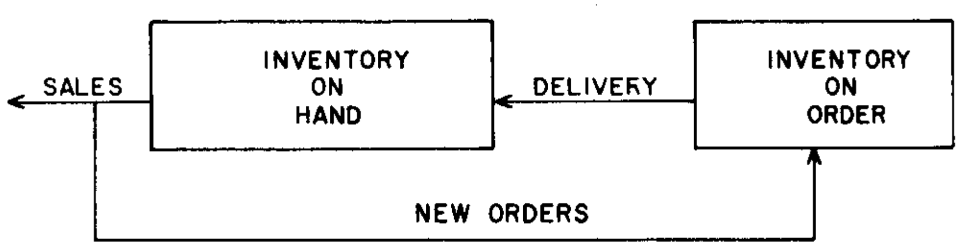
16

STATIONARY OPERATING CHARACTERISTICS OF AN INVENTORY MODEL WITH TIME LAG

HERBERT SCARF

In the last several years a number of papers have appeared which discuss the general problem of the control of inventories. In some of these papers the problem of selecting appropriate inventory policies has been treated as a decision problem. This point of view emphasizes the determination of optimal inventory policies, that is, policies which maximize some desired objective associated with the continual build-up and utilization of inventories. A certain amount of success has been obtained in characterizing the form of optimal policies when various assumptions are placed on the relevant cost functions. However, the actual determination of optimal policies in any specific situation is a task which generally cannot be solved analytically, and frequently exceeds even the capabilities of modern high speed computers. This makes it necessary for some compromises to be introduced.

The point of view which we shall adopt in this chapter is to restrict our attention to a rather simple class of inventory policies which are



actually found in practice and to describe the effects of these policies. If we introduce a specific inventory policy, the resultant fluctuating inventory level is a stochastic phenomenon, governed primarily by the statistical aspects of demand. The underlying stochastic process may be described schematically as follows.

Items are ordered according to the inventory policy being used ; they are delivered after a certain length of time has elapsed, which may be random or constant, and they are then sold according to the demand distribution. Most of the inventory policies in actual use have a self-regulatory aspect, in the sense that a depletion of the inventory on hand induces substantial ordering, whereas an excess of the inventory on hand results in diminished ordering. This is essentially a feedback phenomenon, which modulates the size of the inventory.

In our model we shall assume that the demand is an arbitrary process, in the sense that the distribution of the time between successive demands is an arbitrary distribution function $\Theta(\tau)$. We shall denote the probability distribution of the time lag by $\Omega(\lambda)$.

At any time we shall have a certain number, say x units, on hand, and a corresponding number, say w units, on the books, with some delivery dates associated with each shipment on the books. The policy that we shall examine is the very simple one which tells us to order an amount $D = S - s$ whenever $x + w$ falls to s, and otherwise to order nothing.

A distinction must be made between two possible interpretations of this policy. We have not yet specified whether or not to allow negative inventories on hand. For example, if we assume that a demand which occurs when the inventory on hand is zero, may be satisfied by an item which is delivered in the future, then we are essentially permitting the inventory on hand to be negative. We may make the alternative assumption that a demand which occurs when the inventory on hand is zero, is either neglected entirely or satisfied by priority shipment. In either of the latter cases the inventory on hand is always nonnegative. We shall refer to the first of these models as the infinite model, and the second as the finite.

If a choice of S and s is made, and the infinite or finite model is selected, then a description of the system at, say, $t = 0$, gives rise to a well-defined stochastic process for the inventory model. In order to compare these stochastic processes to those previously studied, let us assume for the moment that the infinite model has been selected. In this case the number of shipments on the books is precisely the same as the number of busy servers in a queue with an infinite number of servers ; customers being assumed to arrive according to the interarrival distribution $\Theta_D(\tau)$ (the D-fold convolution of Θ), and with service performed according to the distribution $\Omega(\lambda)$. This model is also identical to the telephone trunking problem [2]. Both of these forms of the process have been discussed quite thoroughly [4].

In the finite model with $s = S - 1$ the number of shipments on the books is the same as the number of busy servers in a queue with a finite number of servers, with the assumption that if customers arrive when all servers are busy, they leave immediately, and a waiting line is not formed. This model may also be described as an extension of the type I and type II counter problems, in which incoming pulses are disregarded if S pulses are still affecting the counter. This model, with very special assumptions on the distributions involved, has been studied by Erlang [1].

Let us say a word about the method to be used in analyzing these models. Suppose for the moment that we are considering the infinite model, and let us say that the system is in state (m, j) if there are m shipments on the books and $x = S - mD - j$ units in current inventory $(0 \leqq j \leqq D - 1)$. If $m < [S/D]$, and $j = 0, 1, \cdots, D-1$ or $j = 0, 1, \cdots, S - D[S/D]$ if $m = [S/D]$, the number of units held in current inventory is nonnegative and it is natural to assume that a holding cost h_x per unit time will be charged. On the other hand x may be negative, i.e., demand has exceeded supply, and we assume that a penalty cost p_x per unit time will be charged. If the system starts out in any specific state, then as time goes on the system will wander from one state to another, accumulating both holding and penalty costs. Let us denote the expected costs accumulated in t units of time by $C(t)$. Then it may be shown that as t becomes infinite $C(t)/t$ approaches a limit which may be computed in the following way. Let $\pi_{m,j}$ be the average length of time spent in state (m, j). Then

$$\lim_{t\to\infty} \frac{C(t)}{t} = \sum_{x\geq 0} h_x \pi_{m,j} + \sum_{x<0} p_x \pi_{m,j},$$

where x is $S - mD - j$. The same procedure may be used to determine the average value per unit time of any cost or revenue which is a linear function of the time spent in a given state. The costs may themselves be nonlinear in the state variables.

There are, of course, other types of costs or revenues associated with the inventory process. For example, ordering costs do not depend on the length of time spent in any given state, but are incurred instantaneously whenever a new shipment is placed on the books. Before we discuss these costs let us discuss the specific expressions obtained in this chapter for the numbers $\pi_{m,j}$.

1. If $D = 1$ $(s = S - 1)$ and demands arrive according to a Poisson distribution, then for the infinite model,

$$\pi_m = e^{-\alpha} \frac{\alpha^m}{m!}, \quad \text{where} \quad \alpha = \frac{E(\lambda)}{E(\tau)}.$$

It is to be noticed that this result is correct for an arbitrary distribution of the time lag. This result, however, is not original with this book; it has been previously obtained by a number of authors, and the proof will not be repeated here.

If $D = 1$ $(s = S - 1)$ and demands arrive according to a Poisson distribution, then for the finite model,

$$\pi_m = \frac{e^{-\alpha}\dfrac{\alpha^m}{m!}}{e^{-\alpha}\sum_{j=0}^{S}\dfrac{\alpha^j}{j!}} \quad \text{for } m = 0, 1, \cdots, S.$$

The value of α is the same as in the infinite case, and again the result is correct independently of the distribution of the time lag. An heuristic proof of this result is given in the body of this chapter. A rigorous proof is given in Chapter 17 of this book.

2. In both this result and the subsequent one the time lag distribution is restricted to be a negative exponential distribution, i.e.,

$$\Omega(\lambda) = 1 - e^{-(\lambda/L)},$$

where L represents the average time lag. In order to describe this result we shall introduce some notation. Let $\Theta_D(\tau)$ represent the D-fold convolution of $\Theta(\tau)$, that is, the length of time for D successive demands to occur. We define

$$M(\tau) = \sum_1^\infty \Theta_{nD}(\tau).$$

The quantity $M(\tau)$ is a very common one in the study of renewal processes and represents the average number of groups of D successive demands that occur in the time period $(0, \tau)$ or the average number of times that orders are placed in this time period. We shall use the following notation for Laplace transforms:

$$\tilde{G}(s) = \int_0^\infty e^{-sy}\, dG(y).$$

As a final definition, let us write

$$a_n = \tilde{M}\left(\frac{1}{L}\right)\tilde{M}\left(\frac{2}{L}\right) \cdots \tilde{M}\left(\frac{n}{L}\right). \tag{1}$$

Our main result as far as the infinite model is concerned is to give an explicit expression for the generating function associated with the quantities $\pi_m = \pi_{m,0} + \cdots + \pi_{m,D-1}$. This result is that if

$$\pi(y) = \sum_0^\infty \pi_m y^m,$$

then

$$\pi'(y) = \frac{L}{DE(\tau)} \sum_0^\infty a_n (y-1)^n, \quad \text{and } \pi(1) = 1. \tag{2}$$

Let us, first of all, remark that corresponding expressions may be obtained for

$$\pi_j(y) = \sum_0^\infty \pi_{m,j}\, y^m,$$

using precisely the same techniques as were used in obtaining the above expression. We shall not perform the explicit computations here.

The method that we shall use in this chapter for obtaining the above result is that of the imbedded Markov chain, a method which has been used by numerous authors on queueing theory [3]. Of course, as was previously mentioned, the infinite model is actually a queue with an infinite number of servers. An alternative and somewhat simpler procedure for obtaining the above result is given in Chapter 17, and is based on the work of L. Takács [4].

In using his ideas, it is not necessary to assume an exponential time lag. Takács' procedure, however, is inapplicable to the finite model, and we are forced to return to some modification of the imbedded Markov chain technique.

Let us illustrate the above result when the individual demands are assumed to arrive according to a Poisson process, with an average number of demands per unit time equal to μ. Then $\Theta(\tau) = 1 - e^{-\mu\tau}$, and a simple computation gives

$$a_n = \frac{1}{\prod_{j=1}^{n}\left[\left(1 + \frac{j}{\mu L}\right)^D - 1\right]} .$$

For $D = 1$, we obtain $a_n = (\mu L)^n/n!$, so that

$$\pi'(y) = (\mu L)\sum_0^\infty \frac{(\mu L)^n (y-1)^n}{n!} = \mu L \exp \mu L(y-1) ,$$

and $\pi(y) = \exp \mu L(y-1)$. This latter function is, as we would expect, the generating function of a Poisson distribution with mean μL.

For D = 2, it may be shown that

$$\pi(y) = \Gamma(2\mu L)\frac{J_{2\mu L-1}(2\mu L\sqrt{1-y})}{(\mu L\sqrt{1-y})^{2\mu L-1}} ,$$

where $J_{2\mu L-1}$ is the Bessel function of order $2\mu L - 1$. The quantities π_m may be found by successive differentiation to be

$$\pi_m = \frac{\Gamma(2\mu L)}{m!}\,\frac{J_{2\mu L-1+m}(2\mu L)}{(\mu L)^{2\mu L-1-m}} .$$

For $D > 2$, the functions $\pi(y)$ are the solutions of a generalized hypergeometric equation.

3. The finite model represents a substantially new stochastic process, which we solve for an arbitrary interarrival time and a negative exponential distribution for the time lag. For simplicity, we shall describe the result for $s = S - 1$ or $D = 1$, though the general result is quite similar and is given in Section 3 (specifically, equations 25 and 26).

We are interested in obtaining an expression for the generating function

$$\pi(y) = \sum_0^S \pi_m y^m ,$$

and in order to do this, we define the function

$$G_j(y) = \sum_{k=0}^{j} a_j(y-1)^j,$$

where a_j is defined as before (we remember that for the exposition of this section $D = 1$). Then our result is

$$\pi'(y) = \frac{L}{E(\tau)} \frac{\sum_{j=1}^{s} \frac{1}{a_j}\binom{s}{j} G_{j-1}(y)}{\sum_{j=0}^{s} \frac{1}{a_j}\binom{s}{j}}. \tag{3}$$

(It was brought to the author's attention, after the present chapter had been written, that Takács in [5] had obtained this solution for $D = 1$.)

4. The previous discussion has shown us how to compute the average value of those costs and revenues which depend merely on the amount of time spent in the states of the model, under the assumption of a negative exponential distribution of the time lag. There are, of course, other costs which are relevant to the choice of an inventory policy. One of the most important of these is related to the following considerations. Let us consider the finite model and assume that all demands which occur when no inventory is on hand are satisfied by priority shipment. All of the sales may be placed into two categories, routine sales and priority sales; let us suppose that we are interested in determining the average ratio of routine sales to total sales, induced by the inventory policy that has been adopted.

It may be shown, for general distributions Θ and Ω, that

$$\frac{\text{average inventory on order}}{\text{average routine sales per unit time}} = L.$$

This permits us to obtain an explicit expression for the average number of routine sales per unit time, inasmuch as the average inventory on order is given by $D\pi'(1)$. More specifically, if R represents the average number of routine sales per unit time, then assuming that the time lag is represented by a negative exponential distribution, and that $D = 1$, we obtain

$$1 - RE(\tau) = \frac{1}{\sum_{j=0}^{S} \frac{1}{a_j}\binom{s}{j}}. \tag{4}$$

A similar expression is obtained for $D \geqq 1$, in Section 4 (equation 27).

There are several other figures of merit associated with an inventory problem, which are readily calculable in terms of R and the data of the problem. For example, the average inventory on hand is given by

$$E(x) = S - E(\lambda)R - r + E(\tau)R\left[r - \frac{D-1}{2}\right], \tag{5}$$

where $S = nD + r$, with $0 \leqq r < D$. Another example is the frequency of ordering, which is given by

$$\frac{R}{D}. \tag{6}$$

It is quite possible that a knowledge of the three quantities discussed in this section would be sufficient to form a judgment about the merits of a given (s, S) policy. For example, if the holding cost is proportional to the inventory on hand, i.e., $h_x = hx$, and if the penalty cost is proportional to the number of lost sales per unit time, then the average costs per unit time may be determined from a knowledge of these quantities. Let us denote the cost of purchasing D units by C_D. Then the average costs per unit time are

$$hE(x) + p\left\{\frac{1}{E(\tau)} - R\right\} + \frac{C_D R}{D}.$$

The stockage policy may then be chosen so as to minimize this cost. Even if the costs are not known to be of the specific form mentioned above, knowledge of the quantities involved in (4), (5), and (6) could be of considerable importance in selecting an appropriate (s, S) policy.

We shall present some tables of these quantities at the end of this chapter, on the assumption that the demand follows a Poisson distribution. If $D = 1$, then the result of Section 1 demonstrates the insensitivity of these computations to the time lag distribution. On the other hand, it may be shown that if $D = S$, the results are again insensitive to the time lag distribution. This indicates that the value of R, while computed on the basis of a negative exponential time lag distribution, will be approximately the same regardless of the time lag distribution, at least for extreme values of D.

1. The Case D = 1 (Poisson Demand)

In this section we shall introduce the very special inventory policy which consists in keeping the amount on hand plus the amount on order equal to a constant S. This corresponds to the case $D = 1$. We shall also assume that the demand is given by a Poisson distribution with mean μ per unit time. In the next section we shall consider more general types of processes.

We shall begin by assuming that the time between placing an order and receiving it, is distributed according to a negative exponential distribution with mean L, i.e., according to the density function

$$\frac{1}{L}e^{-(\lambda/L)}.$$

If the inventory policy is to keep the amount on hand plus the amount on order equal to S, then the state of the system may, at any time, be described by an integer w ranging from 0 to S, which represents the amount on order. At first glance, it would seem necessary to record the time at which each shipment was originally ordered. However, the negative exponential distribution for the time lag makes it unnecessary to do this; for if a shipment was originally ordered at time $\bar{t}$, and if it has not been delivered at time $t > \bar{t}$, then the conditional distribution of delivery time T is

$$1 - e^{-(1/L(T-t))},$$

which is the same as if the shipment were originally ordered at time t.

In the time interval $(t, t + \delta)$ any shipment on order has a probability of being delivered equal to

$$\frac{\delta}{L} + o(\delta),$$

regardless of when the shipment had been originally ordered. It is easy to see that the amount on order varies according to a continuous time Markov process with the following transition matrix ($i = 0, \cdots, S$).

$$(P_{ij}(t)) = I + t\begin{pmatrix} -\mu & \mu & \cdots & 0 \\ \frac{1}{L} & -\left(\mu + \frac{1}{L}\right), \ \mu & \cdots & 0 \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ 0 & 0 & \cdots & -\frac{S}{L} \end{pmatrix} + o(t)$$

$$= I + tQ + o(t), \quad \text{for } t \text{ small.}$$

The limiting distribution π_m is given by the solution of the equation $\pi Q = 0$, and turns out to be

$$\pi_m = \frac{\dfrac{(\mu L)^m}{m!}}{\displaystyle\sum_{j=0}^{S} \frac{(\mu L)^j}{j!}},$$

which is a truncated Poisson distribution. This result was known to Erlang [1]. The average inventory on order is given by

$$\mu L \frac{P_{S-1}(\mu L)}{P_S(\mu L)}, \quad \text{where} \quad P_n(x) = e^{-x}\sum_{0}^{n} \frac{x^j}{j!}.$$

It may be shown that the long term average number of routine sales per unit time is given by

$$\mu(1 - \pi_S) = \mu \frac{P_{S-1}(\mu L)}{P_S(\mu L)},$$

and we may therefore conclude that

$$\frac{\text{average inventory on order}}{\text{average routine sales per unit time}} = L.$$

The next problem that we wish to discuss, again under the assumption that demands arrive according to a Poisson process and that $s = S - 1$, is the finite model with an arbitrary distribution for the time lag $\Omega(\lambda)$. As was mentioned in the introduction, the proportion of time spent in any given state is the same as under the assumption of a negative exponential time lag as long as the mean of $\Omega(\lambda)$ is L. We shall not give a completely rigorous proof of this fact for all distributions, but only for a subclass of distributions, which is, however, dense in the class of all distributions. Let us assume that $\Omega(\lambda)$ is a distribution for the recurrence time for a state of a continuous time, finite state Markov process. What this means is that when a shipment is placed on the books, it passes through a number of stages not necessarily sequentially, prior to delivery, and that the length of time spent in each one of these stages is a negative exponential distribution.

Let us denote the states of the Markov process by $1, 2, \cdots, k$, and let $\Omega(\lambda)$ be the distribution of time for a particle to leave state 1 and then return to it. Let us suppose that the transition probabilities are given by $P_{ij}(t) = \delta_{ij} + q_{ij}t + o(t)$ for small t. The stationary probabilities satisfy the equations

$$\sum_i \pi_i q_{ij} = 0 \ ,$$

and the mean of the distribution $\Omega(\lambda)$, which we have called L, is given by $-1/\pi_1 q_{11}$.

It is clear that the inventory model may be described, at any time, by an integer w, which represents the total number of shipments on order, allocated to the k states of the Markov process. That is, the system is described by nonnegative integers $w_1, \cdots, w_k$, with $w_1 + \cdots + w_k = w$. Transitions of various kinds may take place. For example, if the system is in any state with $w < S$, it is possible for a new order to be placed, and the probability that this occurs in a length of time t is given by $\mu t + o(t)$. Again if the system is in a state with $w > 0$, a delivery may occur, and the probability that this happens is the probability that any one of the particles in the states $2, \cdots, k$, makes a transition to state 1, or $(w_2 q_{21} + \cdots + w_k q_{k1})t + o(t)$. The shipments may, in a small time t, move around within the Markov process with neither a delivery nor a new shipment occurring. Lastly there may be no transition at all.. Combining these alternatives, we see that the stationary probabilities $\pi_{w_1 \cdots w_n}$ for the inventory model are given by

$$\pi_{w_1 \cdots w_k}\Big[\mu - \sum_{i=1}^{k} w_i q_{ii}\Big] = \mu \pi_{w_1 - 1 \cdots w_k} + \sum_{i=2}^{k} \pi_{w_1 \cdots w_i + 1 \cdots w_k}(w_i + 1) q_{i1}$$
$$+ \sum_{j=2}^{k} \sum_{i=1}^{k} \pi_{w_1 \cdots w_i + 1 \cdots w_j - 1 \cdots w_k}(w_i + 1) q_{ij} \, .$$

It may be shown, by direct substitution in these equations, that the limiting distribution $\pi_{w_1 \cdots w_k}$ is given by the following rule : The probability of having w shipments on order is

$$\frac{\dfrac{(\mu L)^w}{w\,!}}{\displaystyle\sum_{J=0}^{S} \frac{(\mu L)^J}{J\,!}},$$

and if there are w shipments on the books, then the state in which any shipment may be found is independent of the state of any other shipment, and given by the probability distribution (π_j). In other words, the conditional distribution of $(w_1, \cdots, w_k)$ given that $w_1 + \cdots + w_k = w$ is given by

$$\frac{w\,!}{w_1\,! \cdots w_k\,!} \pi_1^{w_1} \cdots \pi_k^{w_k}.$$

This demonstrates the result discussed in paragraph 1 of the introduction.

2. The Infinite Model with an Arbitrary Interarrival Distribution and a Negative Exponential Distribution for the Time Lag

In the previous section the stationary probabilities for the number of outstanding orders were computed under certain assumptions. The nature of these assumptions was essentially to make the number of outstanding orders a continuous time Markov process, and thereby permitted us to compute the limiting probabilities by means of standard techniques.

However, when the interarrival distribution is permitted to be arbitrary, the number of outstanding shipments is no longer a continuous time Markov process and some alternative procedure must be used to evaluate the average time spent in each state. In this chapter we use the method of the imbedded Markov process. As was mentioned in the introduction, it is also possible to discuss the infinite model by means of a method due to Takács, with no assumptions on either the interarrival distribution or the time lag distribution. This latter method, however, is inapplicable to the finite model, which is discussed by means of the imbedded Markov process in Section 3.

The method of the imbedded Markov process depends on an analysis of the number of outstanding orders at a discrete set of time points, at which points the process is a discrete time Markov process. Let us refer to the notion of the state of the inventory model, defined in the introduction. The system is in state $S_{m,j}$ if the number of outstanding shipments is m (each shipment consisting of D items), and if the inventory on hand is $x = S - mD - j$ where $j = 0, 1, \cdots, D-1$. For the infinite model, m may be an arbitrary nonnegative integer.

We start the process off at time $t_0 = 0$ in state S_{10}. Let t_1 be the time it takes for D demands to arrive. At time t_1, the process will be in state S_{10} or in state S_{20}, depending on whether or not a delivery has been made. Let t_2 be the time when D additional demands arrive. At this time, the process will be at one of the states S_{10}, S_{20}, or S_{30}. If we examine the process at the points t_0, t_1, t_2, $\cdots$, with the property that between any pair of consecutive points D demands have been made, then at these times the process will be in states $S_{m,0}$ which we shall designate by S_m. The fact that the time lag distribution is assumed to be a negative exponential distribution implies that the sampled process is a Markov process. We shall first determine an explicit formula for the transition probabilities of the sampled process, and then determine the limiting distributions. We shall then relate the limiting probabilities to the average time spent by the original process in any state.

Let us assume, then, that we are at state m, where m is any positive integer. The time that it takes for D additional units to be sold is a random variable τ drawn from a distribution $\Theta_D(\tau)$ which is a D-fold convolution of $\Theta(\tau)$. If no shipments are delivered in time τ, then since a new shipment is requested whenever D demands have been made, the system will move from state m to state $m+1$ at the next sampling point. In general, if $m-k$ shipments are delivered in time τ, the system moves from m to $k+1$. Now for any definite time τ, the probability that $m-k$ shipments out of the m are actually delivered is given by

$$\binom{m}{k}(e^{-(\tau/L)})^k(1-e^{-(\tau/L)})^{m-k} .$$

The reasoning behind this formula is that for any shipment on the books, the probability of its being delivered in time τ is

$$1-e^{-(\tau/L)} ,$$

and that delivery is independent for all outstanding shipments. Since τ is distributed by $\Theta_D(\tau)$, we see that

$$p_{m,k+1}=\binom{m}{k}\int_0^\infty (e^{-(\tau/L)})^k(1-e^{-(\tau/L)})^{m-k}\,d\Theta_D(\tau) . \tag{7}$$

The sampled process ranges through the states 1, 2, $\cdots$, ∞ with transition probabilities $p_{m,k+1}$ as given above. We are interested in computing the stationary distribution $(p_1, \cdots)$ associated with this process. The p's satisfy the equations

$$p_{k+1}=\sum_{m=1}^{\infty} p_m p_{m,k+1} ,$$

and therefore if

$$G(x)=\sum_{k=0}^{\infty} p_{k+1}x^k ,$$

then we have

$$
\begin{aligned}
(8) \qquad G(x) &= \sum_{m=1}^{\infty} p_m \sum_{j=0}^{m} p_{m,\,k+1} x^k \\
&= \sum_{m=1}^{\infty} p_m \int_0^{\infty} \sum_{k=0}^{m} \binom{m}{k} (x e^{-(\tau/L)})^k (1 - e^{-(\tau/L)})^{m-k}\, d\Theta_D(\tau) \\
&= \sum_{m=1}^{\infty} p_m \int_0^{\infty} \Big[1 - (1 - x) e^{-(\tau/L)}\Big]^m d\Theta_D(\tau) \\
&= \int_0^{\infty} \Big[1 - (1 - x) e^{-(\tau/L)}\Big] G\Big[1 - (1 - x) e^{-(\tau/L)}\Big] d\Theta_D(\tau)\,.
\end{aligned}
$$

This rather complex equation can be solved quite directly if we substitute

$$G(x) = \sum_0^{\infty} a_n (x - 1)^n, \quad \text{with} \quad a_0 = 1.$$

The integral becomes

$$\sum_0^{\infty} a_n \int_0^{\infty} \Big[1 - (1 - x) e^{-(\tau/L)}\Big] (x - 1)^n e^{-(\tau n/L)}\, d\Theta_D(\tau),$$

and if we equate coefficients we obtain

$$a_n = a_{n-1} \tilde{M}\Big(\frac{n}{L}\Big), \quad \text{where} \quad \tilde{M}\Big(\frac{n}{L}\Big) = \int_0^{\infty} e^{-(\tau n/L)}\, dM(\tau),$$

so that

$$(9) \qquad G(x) = \sum_{m=0}^{\infty} \tilde{M}\Big(\frac{1}{L}\Big) \cdots \tilde{M}\Big(\frac{m}{L}\Big) (x - 1)^m \,.$$

The functions $G(x)$ may be used to compute the limiting distribution (p) for any D. However this limiting distribution is not, in itself, applicable to the computation of the average revenues and costs. (p) represents the limiting distribution of the process which is sampled at time intervals in which precisely D demands have been made, and therefore at times when a new order has just been placed on the books. In the time between two successive samplings the process will be moving from state to state, and we have to come to grips with this fluctuation in order to compute the average time spent in any state. If $\pi_{m,\,j}$ represents the average time in $S_{m,\,j}$ of the continuous time process, we propose to compute

$$\pi_m = \sum_{j=0}^{D-1} \pi_{m,\,j}\,.$$

This distribution will be directly applicable to the average costs. The determination of this distribution will make use of the function $G(x)$.

At a sampled time the process moves into one of the states $S_{m,\,0}$. Let us denote by σ_m the state $S_{m,\,0} + \cdots + S_{m,\,D-1}$. During the time in which the next demands are made, the process moves from σ_m to σ_{m-1}, and so forth. We wish to compute the expected length of time spent in σ_m, $\sigma_{m-1}, \cdots, \sigma_0$, in one cycle of demands. Let us begin by starting off the system in state $S_{m,0}$ at time 0 and determine the expected length of time spent in the states $\sigma_m, \cdots, \sigma_0$ during the time interval $(0, t)$, where t is a fixed number. We denote the expected length of time in σ_k by $T_{m,k}(t)$, for $k = 0, 1, \cdots, m$. Of course

$$\sum_{k=0}^{m} T_{m,k}(t) = t .$$

The fact that the length of time in σ_k is a negative exponential distribution with mean L/k enables us to write the following relation.

$$T_{m,k}(t+\delta) = \frac{\delta m}{L} T_{m-1,k}(t) + \left(1 - \frac{\delta m}{L}\right) T_{m,k}(t) + o(\delta),$$

for $k < m$, and this yields

$$T'_{m,k}(t) = \frac{m}{L} T_{m-1,k} - \frac{m}{L} T_{m,k},$$

for $k < m$. A simple calculation shows us that

$$T_{m,m}(t) = \frac{L}{m}(1 - e^{-(mt/L)}) , \quad \text{if } m > 0 .$$

It is easy to see that these equations are satisfied by

$$T_{m,k}(t) = \int_0^t \binom{m}{k} e^{-(\xi k/L)} (1 - e^{-(\xi/L)})^{m-k} \, d\xi,$$

and therefore the average length of time spent in state σ_k during one cycle is

$$a_{m,k} = \int_0^\infty \int_0^\tau \binom{m}{k} e^{-(\xi k/L)} (1 - e^{-(\xi/L)})^{m-k} \, d\xi \, d\Theta_D(\tau) . \tag{10}$$

The average time in the states $\sigma_0, \sigma_1, \cdots$ is given by the average length of time in each state, for one cycle, assuming that we start off the cycle with a stationary distribution, divided by the average length of each cycle. If we denote this set of frequencies by (π_m), then

$$\pi_m = \frac{1}{DE(\tau)} \sum_{v=m}^{\infty} p_v a_{v,m} .$$

Let us define

$$\pi(y) = \sum_0^\infty \pi_m y^m .$$

Then

$$\begin{aligned}\pi(y) &= \frac{1}{DE(\tau)} \sum_{m=0}^{\infty} \sum_{v=m}^{\infty} p_v a_{v,m} y^m \\ &= \frac{1}{DE(\tau)} \int_0^\infty \int_0^\tau [1 - (1-y)e^{-(\xi/L)}] G[1 - (1-y)e^{-(\xi/L)}] \, d\xi \, d\Theta_D(\tau) .\end{aligned}$$

If we substitute

$$G(y) = \sum_0^\infty a_n(y-1)^n$$

in this expression and simplify, we obtain

$$\pi'(y) = \frac{L}{DE(\tau)} G(y) , \tag{11}$$

with $\pi(1) = 1$. This is the result described in the introduction (equation 2). It is possible, using these techniques to compute all of the limiting probabilities associated with the inventory process. A simple calculation, which we shall not perform here, permits us to determine

$$\pi_j(y) = \sum_0^\infty \pi_{m,j} y^m .$$

3. The Finite Model with an Arbitrary Interarrival Distribution and a Negative Exponential Distribution for the Time Lag

In this section we shall discuss the case in which the inventory on hand is required to be nonnegative, and determine the average time spent in the states of the system by a modification of the technique used in Section 2. Let the inventory policy be defined by the numbers (s, S) with $D = S - s$, and let $S = nD + r$, where $0 \leqq r < D$. The number of shipments on the books is an integer $m = (0, 1, \cdots, n)$.

Let us sample the system, as before, whenever a new order is placed on the books. It is important to notice that the lengths of time between successive samplings are not identically distributed. More precisely, if the system is in state m, with $m < n$, at one sampled time, then the interval of time until the next order is placed on the books is an observation from the distribution $\Theta_D(\tau)$. On the other hand, if the system is in state n at a sampled time (necessarily the number of items on hand will be r), it may happen that r demands arrive before a shipment is delivered. This reduces the number of items on hand to zero. No sales will be made, even though demands arrive, until a shipment is delivered to replenish the inventory on hand, and therefore the number of demands that arrive before the next order is placed may be greater than D.

We may therefore say that if $m < n$, the transition probability $p_{m,k+1}$ is given by

$$p_{m,k+1} = \binom{m}{k} \int_0^\infty e^{-(\tau k/L)} (1 - e^{-(\tau/L)})^{m-k} \, d\Theta_D(\tau) , \tag{12}$$

which is the same as equation 7. However if $m = n$, the following argument is required. Assume for the moment that the time necessary for the first r demands to be satisfied is τ_r. (τ_r has a distribution $\Theta_r(\tau)$, an r-fold convolution of $\Theta(\tau)$.) At the time when the rth demand has just been satisfied, the number of outstanding shipments is q, with probability

$$\binom{n}{q}e^{-(\tau_r q/L)}(1 - e^{-(\tau_r/L)})^{n-q}.$$

As long as $q < n$, there will be enough items in current inventory to satisfy all demands until the next order is placed τ_{D-r} time units later. During this later time period the number of shipments on the books changes from q to $k+1$ with probability

$$\binom{q}{k}\int_0^\infty e^{-(\tau k/L)}(1 - e^{-(\tau/L)})^{q-k}\,d\Theta_{D-r}(\tau).$$

Therefore the probability that in one cycle the number of shipments outstanding changes from n to $k+1$ without there being an intervening period with no items on hand is

$$\sum_{q=k}^{n-1}\binom{n}{q}e^{-(\tau_r q/L)}(1 - e^{-(\tau_r/L)})^{n-q}\binom{q}{k}\int_0^\infty e^{-(\tau k/L)}(1 - e^{-(\tau/L)})^{q-k}\,d\Theta_{D-r}(\tau).$$

On the other hand, with probability $e^{-(\tau_r n/L)}$ no shipments will have been delivered at time τ_r and therefore there will be no items on hand. The length of time until the first of the n outstanding shipments is delivered is a negative exponential with mean L/n, and we let z be that random

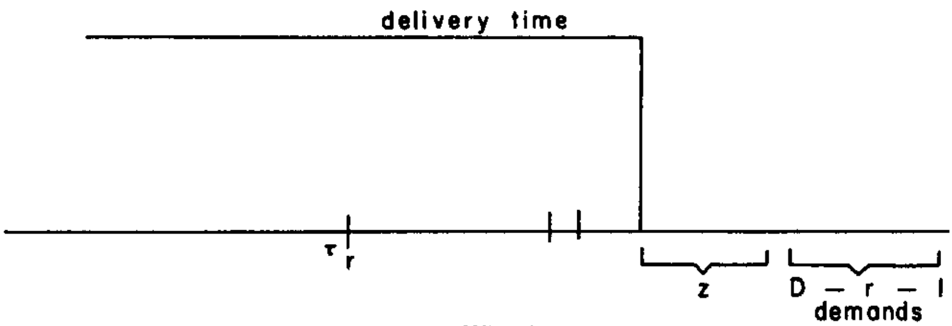


FIGURE 2

variable which represents the time between delivery of the first outstanding order and the next demand. If we denote the distribution function of z by $F(z)$ and its Laplace transform by

$$\tilde{F}(\alpha) = \int_0^\infty e^{-\alpha z}\,dF(z),$$

then it may be shown that

$$\tilde{F}(\alpha) = \frac{\tilde{\Theta}(\alpha) - \tilde{\Theta}\left(\frac{n}{L}\right)}{\left(1 - \frac{L\alpha}{n}\right)\left[1 - \tilde{\Theta}\left(\frac{n}{L}\right)\right]}. \tag{13}$$

Therefore the probability of a transition from n to $k+1$ in one cycle, there being a period in the cycle with no items on hand, is given by

$$e^{-(\tau_r n/L)}\binom{n-1}{k}\int_0^\infty e^{-(\tau k/L)}(1 - e^{-(\tau/L)})^{n-1-k}\,d(F^*\Theta_{D-r-1})(\tau).$$

Adding this expression to the above, integrating with respect to the distribution of τ_r, and simplifying, we obtain

$$(14)\qquad p_{n,k+1} = \binom{n}{k}\int_0^\infty e^{-(\tau k/L)}(1 - e^{-(\tau/L)})^{n-k}\, d\Theta_D(\tau) + \frac{\tilde{\Theta}_{r+1}\left(\frac{n}{L}\right)}{1-\tilde{\Theta}\left(\frac{n}{L}\right)}\binom{n}{r}\int_0^\infty e^{-(\tau k/L)}(1 - e^{-(\tau/L)})^{n-k}\, d(\Theta_{D-r} - \Theta_{D-r-1})(\tau)$$

for $k = 0, 1, \cdots, n-1$. The reader may verify that

$$\sum_{k=0}^{n-1} p_{n,k+1} = 1\ .$$

Let us denote the stationary distribution for this Markov process by $(p_1, \cdots, p_n)$. Let us define, as we did in the infinite model, $\hat{a}_{m,k}$ to be the average time spent with k undelivered shipments during one cycle, assuming the cycle to start off in state m. If $m < n$, we may apply the calculation of the preceding section (equation 10) and obtain

$$(15)\qquad \hat{a}_{m,k} = \int_0^\infty\int_0^\tau \binom{m}{k} e^{-\xi k/L}(1 - e^{-\xi/L})^{m-k}\, d\xi\, d\Theta_D(\tau)\ .$$

However, if $m = n$ a special argument is required.

Reasoning along those lines used in obtaining $p_{n,k+1}$, we may obtain an expression for $\hat{a}_{n,k}$. It is a rather complex expression, and is of no use to us directly. However, a fairly elaborate manipulation demonstrates the following important result:

$$(16)\qquad \frac{d}{dy}\sum_{k=0}^{m} \hat{a}_{m,k} y^k = \frac{-L}{y-1}\sum_{k=1}^{m} p_{m,k+1} y^k + \frac{Ly^m}{y-1}\ ,$$

for all $m = 1, 2, \cdots, n$ (for $m = n$ the term $p_{n,n+1}$ is understood to be zero).

We shall now relate π_m, the average time spent with m undelivered orders, to the quantities $\hat{a}_{m,k}$ and p_m. It is clear that π_m is equal to the average time in state m per cycle, divided by the average length of each cycle. Therefore

$$(17)\qquad \pi_m = \frac{\sum_{v=1}^{n} p_v \hat{a}_{vm}}{\text{average cycle length}}\ .$$

The average cycle length may be determined as follows. With probability $1 - p_n$ a cycle will begin with less than n outstanding shipments, and consequently the average cycle length will be $DE(\tau)$. On the other hand, with probability p_n a cycle will begin with n outstanding shipments.

Let τ_r be the time for the first r demands of the cycle to arrive. With probability

$$1 - e^{-(\tau_r n/L)}$$

a shipment will have been delivered before this time, and the expected cycle length will be $\tau_r + (D - r)E(\tau)$. With probability

$$e^{-(\tau_r n/L)}$$

a shipment will not have been delivered and the expected cycle length will be $\tau_r + L/n + E(z) + (D - r - 1)E(\tau)$. It may be shown, from the formula given above for the Laplace transform of the distribution of z (equation 13), that

$$E(z) = \frac{E(\tau)}{1 - \tilde{\Theta}\left(\frac{n}{L}\right)} - \frac{L}{n}.$$

Combining this, and integrating out τ_r, we see that the average cycle length is

$$C = DE(\tau) + p_n E(\tau) \frac{\tilde{\Theta}_{r+1}\left(\frac{n}{L}\right)}{1 - \tilde{\Theta}\left(\frac{n}{L}\right)}. \tag{18}$$

This enables us to write

$$\pi_m = \frac{\sum_{v=1}^{n} p_v \hat{a}_{vm}}{C}. \tag{19}$$

We are now able to obtain an explicit expression for the generating function

$$\pi(y) = \sum_{0}^{n} \pi_m y^m .$$

We substitute (19), differentiate, and make use of (16), obtaining

$$\begin{aligned}
\pi'(y) &= \frac{L}{C} \sum_{k=0}^{n-1} p_{k+1} y^k \\
&= \frac{L}{C} \sum_{k=0}^{n-1} y^k \sum_{m=1}^{n} p_m p_{m,\,k+1} \\
&= \frac{L}{C} \sum_{m=1}^{n} p_m \int_0^\infty [1 + (y-1)e^{-(\tau/L)}]^m \, d\Theta_D(\tau) \\
&\quad + \frac{Lp_n}{C} \frac{\tilde{\Theta}_{r+1}\left(\frac{n}{L}\right)}{1 - \tilde{\Theta}\left(\frac{n}{L}\right)} \int_0^\infty [1 + (y-1)e^{-(\tau/L)}]^n \, d(\Theta_{D-r} - \Theta_{D-r-1})(\tau)
\end{aligned} \tag{20}$$

by (14), and therefore

$$\pi'(y) = \int_0^\infty [1 + (y-1)e^{-(\tau/L)}]\pi'[1 + (y-1)e^{-(\tau/L)}]d\Theta_D(\tau) \tag{21}$$

$$+ \frac{Lp_n}{C} \frac{\tilde{\Theta}_{r+1}\left(\frac{n}{L}\right)}{1 - \tilde{\Theta}\left(\frac{n}{L}\right)} \sum_{k=1}^{n} \binom{n}{k} (y-1)^k \left[\tilde{\Theta}_{D-r}\left(\frac{k}{L}\right) - \tilde{\Theta}_{D-n-1}\left(\frac{k}{L}\right)\right].$$

Let us denote the transformation that carries $f(y)$ into

$$\int_0^\infty [1 + (y-1)e^{-(\tau/L)}] f[1 + (y-1)e^{-(\tau/L)}]\, d\Theta_D(\tau)$$

by Tf. If we apply the operator T to the function

$$G_k(y) = \sum_{j=0}^{k} a_j (y-1)^j ,$$

defined in the introduction, we obtain

$$TG_k = G_k + a_k \tilde{\Theta}_D\left(\frac{k+1}{L}\right)(y-1)^{k+1} .$$

Therefore the solution of the equation

$$\pi' = T\pi' + (y-1) \sum_{j=0}^{\infty} b_j (y-1)^j \tag{22}$$

is given by

$$\pi' = - \sum_{k=0}^{\infty} \frac{b_k}{a_k \tilde{\Theta}_D\left(\frac{k+1}{L}\right)} G_k(y) . \tag{23}$$

In our case

$$b_k = \frac{Lp_n}{C} \frac{\tilde{\Theta}_{r+1}\left(\frac{n}{L}\right)}{1 - \tilde{\Theta}\left(\frac{n}{L}\right)} \left[\tilde{\Theta}_{D-r}\left(\frac{k+1}{L}\right) - \tilde{\Theta}_{D-r-1}\left(\frac{k+1}{L}\right)\right]\binom{n}{k+1} ,$$

for $k \leqq n-1$, and zero otherwise. Therefore, after some simplification, we obtain

$$\pi'(y) = \frac{L}{C} p_n \frac{\tilde{\Theta}_{r+1}\left(\frac{n}{L}\right)}{1 - \tilde{\Theta}\left(\frac{n}{L}\right)} \sum_{k=0}^{n-1} \frac{\binom{n}{k+1} G_k(y)\left[1 - \tilde{\Theta}\left(\frac{k+1}{L}\right)\right]}{a_k \tilde{\Theta}_{r+1}\left(\frac{k+1}{L}\right)} . \tag{24}$$

It remains only to determine the value of the constant p_n, and this is obtained by virtue of the fact that $\pi'(1) = L/C$ (see equation 20). Therefore

$$\frac{1}{p_n} = \frac{\tilde{\Theta}_{r+1}\left(\frac{n}{L}\right)}{1 - \tilde{\Theta}\left(\frac{n}{L}\right)} \sum_{k=0}^{n-1} \frac{\binom{n}{k+1}}{a_k} \frac{1 - \tilde{\Theta}\left(\frac{k+1}{L}\right)}{\tilde{\Theta}_{r+1}\left(\frac{k+1}{L}\right)} .$$

If we recall that

$$C = DE(\tau) + p_s E(\tau) \frac{\tilde{\Theta}_{r+1}\left(\frac{n}{L}\right)}{1 - \tilde{\Theta}\left(\frac{n}{L}\right)},$$

we see that

$$\pi'(y) = \frac{\frac{L}{E(\tau)} \sum_{k=0}^{n-1} \frac{\binom{n}{k+1}}{a_k} \frac{1 - \tilde{\Theta}\left(\frac{k+1}{L}\right)}{\tilde{\Theta}_{r+1}\left(\frac{k+1}{L}\right)} G_k(y)}{1 + D \sum_{k=0}^{n-1} \frac{\binom{n}{k+1}}{a_k} \frac{1 - \tilde{\Theta}\left(\frac{k+1}{L}\right)}{\tilde{\Theta}_{r+1}\left(\frac{k+1}{L}\right)}}. \tag{25}$$

If $D = 1$, and consequently $r = 0$, this reduces to

$$\pi'(y) = \frac{\frac{L}{E(\tau)} \sum_{k=1}^{n} \frac{\binom{n}{k}}{a_k} G_{k-1}(y)}{\sum_{k=0}^{n} \frac{\binom{n}{k}}{a_k}}. \tag{26}$$

4. A Determination of Certain Averages

A very important figure of merit associated with any inventory policy is the average number of sales per unit time, which we denote by R. As we shall show, R is related to the average number of items on the books (this latter quantity is equal to $D\pi'(1)$) in the following way: $R = D\pi'(1)/L$. Since $\pi'(1) = L/C$, this is equivalent to $R = D/C$. This latter equation is correct, since D is the number of sales per cycle and C the average length of a cycle. The fraction of all demands which are not satisfied is

$$1 - E(\tau)R = \frac{1}{1 + D \sum_{k=0}^{n-1} \frac{\binom{n}{k+1}}{a_k} \frac{1 - \tilde{\Theta}\left(\frac{k+1}{L}\right)}{\tilde{\Theta}_{r+1}\left(\frac{k+1}{L}\right)}}.$$

Another measure of the merit of an inventory policy is the average inventory on hand. For example, the average cost per unit time due to a holding cost which is proportional to the amount of inventory on hand, per unit time, is equal to the proportionality constant times the average inventory on hand. If the inventory on hand is x, then $x = S - mD - j$, and $E(x) = S - D\pi'(1) - E(j)$. We shall evaluate $E(j)$ by examining the time spent in any state $j = 0, 1, \cdots, D - 1$, during a given cycle. If $j \neq r$, then the average time in j is $E(\tau)$. Therefore

$(D-1)E(\tau)$ + average time in r = average cycle length = C. If we divide by the length of the cycle, we see that frequency of time spent in any state other than r is given by $E(\tau)/C$ and the frequency in state r is

$$1-(D-1)\frac{E(\tau)}{C}.$$

Therefore

$$E(j)=r+E(\tau)R\left[\frac{D-1}{2}-r\right],$$

and

$$E(x)=S-LR-r+E(\tau)R\left[r-\frac{D-1}{2}\right].$$

Another important figure of merit, especially if set-up costs are involved, is the frequency of ordering, which is clearly given by R/D.

We shall now present some computations of these figures of merit under the assumption that demands arrive according to a Poisson process with mean μ per unit time. If we define $R' = R/\mu$, i.e., the average fraction of incoming demands which are actually satisfied, then it may be shown that as a consequence of the Poisson assumption, both R' and $E(x)$ are functions of μL, which we shall denote by α.

In the following computations we assume that $\alpha = 30$.

$S = 40$

s	R'	$E(x)$
0	.57	11.7
10	.58	9.9
20	.72	11.6
30	.83	11.2

$S = 60$

s	R'	$E(x)$
0	.67	20.3
10	.70	18.9
20	.72	18.8
30	.84	22.5
40	.91	23.9
45	.95	24.9

$S = 80$

s	R'	$E(x)$
0	.73	29.4
10	.76	28.4
20	.79	28.6
30	.82	30.0
40	.91	35.0
50	.94	36.9
60	.99	41.1
70	.99	45.6

$S = 100$

s	R'	$E(x)$
0	.77	38.8
10	.81	38.0
20	.84	38.6
30	.86	40.3
40	.88	42.8
50	.94	48.5
60	.97	51.4
70	.99	55.8
80	.99	60.6

Let us consider a problem in which it is required to satisfy 95 per cent of the sales, and keep the average inventory below 40. If the ordering cost is concave, we would like to maximize D, the size of each individual order. From an examination of these computations we see that $(60, 45)$ and $(80, 50)$ satisfy the constraints, and that $(80, 50)$ is preferable to $(60, 45)$ on the ground that it yields a larger value of D.

REFERENCES

[1] E. Brockmeyer, H. L. Holström, and Orne Jensen, "The Life and Works of A. K. Erlang," *Transactions of the Danish Academy of Technical Sciences*, Copenhagen, 1948.

[2] W. Feller, *Probability Theory and Its Applications*, Vol. I. New York, 1950.

[3] D. G. Kendall, "Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of the Imbedded Markov Chain," *Ann. Math. Stat.*, Vol. XXIV, No. 3 (1953).

[4] L. Takács, "On Secondary Stochastic Processes Generated by Recurrent Processes," *Acta Math. Acad. Sci. Hungary*, VII (1956), 17-29.

[5] L. Takács, "On the Generalization of Erlang's Formula," *Acta Math. Acad. Sci. Hungary*, VIII (1957), 419-33.