Markov Decision Process in Supply Chain Inventory Systems

Mathematics

KEYWORDS :

S.Priscilla Jane	Department of Mathematics, Bethlahem Institute of Engineering ,Karungal.
Dr.C.Elango	Department of Mathematical Sciences, Cardamom Planters' Association College, Bodinayakanur.
Dr.A.Nagarajan	Reader in Mathematics, V.O.C. College, Thoothukudi.

ABSTRACT

In this article we have considered a two-echelon inventory control in a supply chain having one ware-house and multiple retailers. Batch reordering is assumed at each node with variable ordering quantity depending on state of the inventory level at the time of review(periodic review), that is an ordering upto S policy is adopted. By imposing proper cost structure, an optimal policy parameter, reorder level for the inventory control is found by policy iteration procedure. Numerical examples are provided to illustrate the model.

1. Introduction

Most manufacturing enterprises are organized into networks of manufacturing and distribution sites that procure raw material, process them into finished goods, and distribute the finished goods to customers. The terms "multi-echelon" or "multi-level" production / distribution networks are also synonymous with such networks (or supply chains) when an item moves through more than one step before reaching the final customer.

Inventories exist throughout the supply chain in various forms for various reasons. They exists at the distribution warehouses, and they exist 'in-transit', or 'in the pipeline', on each path linking these facilities. All these are related in the sense that :

- The downstream sites create demands on the upstream inventories.
- These uncertain demands, combined with uncertain production and / or transit times largely determine the inventory at a given site.

The central premises here is that the lowest inventories result when the entire supply chain is considered as a single system. Such co-ordinated decisions have produced spectacular results at Xerox [16], and at Hewlett Packard [9], which were able to reduce their respective inventory levels by over 25%, consequently reduced the cost by 33%.

2. Motivation and problem setting

Maintaining inventory efficiently in a supply chain (multi echelon inventory) is a tedious task. Many researchers working in this direction for the last two decades to get a widely accepted inventory control model to solve the problem. Yet a comprehensive model for multi-echelon inventory control is lacking. Most of the systems considered assume that the demand rate and the lead time are constants or follows a well known probability distribution which is a more restricted condition. The authors are motivated by the previous researchers in relaxing conditions on demand and replenishment time.

This paper is organized in the following way. Section 1 deals the introductory part of the problem. In section 2, motivation for the proposed problem is presented in a lucid manner. Section 3, highlights the relevant literature. Section 4, introduces the model and the solution procedure. Section 5, deals with the policy iteration algorithm to find optimal policy parameter and verify the model proposed in Section 4. Section 6 depicts the numerical examples with optimal solution procedure.

3. Literature review

The model in this paper borrows key ideas from two streams of literature: the multi-echelon inventory systems and Markov decision process for inventory control system.

3.1. *Multi-echelon inventory systems*

Since a supply chain consists of various levels of echelons, a key question that needs to be addressed is how partners at these various levels interact with each other. The literature bears evidence of two major but contrasting philosophies: the "push" and the "pull" system. In a distribution system such as the one described in this paper, a push philosophy would mean that there is a central decision maker, say a warehouse manager, who has access to information about inventory levels at all the concerned facilities ; all inventory decisions are then made centrally based on this information. In the pull system, however, the inventory decisions are made by local managers based on their local conditions [6]. In both the push and pull systems, the decision variables (order quantity, reorder point, etc.) are determined so that the overall system costs are minimized. Since this paper deals with a continuous review pull system, we will only review the literature concerning such systems. For a comprehensive discussion of the push system, the reader is directed to Federgruen [5].

One of the earliest continuous review multi-echelon inventory model was due to Sherbrooke [14]. He considers a two-echelon system with several retailers at the lowest echelon and a warehouse that supplies to these retailers. In determining the optimal level of inventory in the system, he introduced the now classic METRIC approximation. When the demand distribution is Poisson, the METRIC approximation essentially characterizes the outstanding retailer orders as Poisson. Graves [6] extends the METRIC approximation by estimating two parameters, the mean and the variance, to describe the outstanding retailer orders. He fits the negative Binomial distribution to these parameters to determine the optimal inventory policy. Axs a ter [1] later provides an exact solution to the problem and shows that the METRIC approximation underestimates whereas the Graves two-parameter approximation overestimates the retailer backorders. Muckstadt [11] and later Lee [10] also provide simple extensions to the basic Sherbrooke model.

All the above studies use a one-for-one ordering policy, i.e., an order is placed as soon
as a demand has occurred. Axsater[1] showed how the methods for one-for-one ordering policy can be extended when there is only one retailer. Analysis of batch-ordering policies in arborescent systems (when number of retailers > 1) can be done similar in spirit to Sherbrooke [14]. Deuermeyer and Schwarz [4] were perhaps the first to analyze such a system. They estimated the mean and variance of the lead time demand to obtain average inventory levels and backorders at the warehouse assuming that lead-time demand is normally distributed. In addition to reviewing the literature in this area, Moinzadeh and Lee [12] , Lee and Moinzadeh [8], and Svoronos and Zipkin [17] also provide several extensions to the Deuermeyer and Schwarz [4] model. The reader is directed to Axsater [1] for a review.

3.2. *Markov decision process*

Markov decision process is a versatile and powerful tool for analyzing probabilistic sequential decision processes with an infinite planning horizon. This process is an outgrowth of Markov process and dynamic programming. Bellman [2] developed the dynamic programming principle in early 1950s. The basic ideas of dynamic programming are the states, the principle of optimality and functional equations. Howard (1960)[7] used basic principle from Markov chain and dynamic programming to develop a policy iteration algorithm for solving probabilistic sequential decision processes with infinite planning horizon. The Markov decision model has many potential applications in inventory control, maintenance, manufacturing and telecommunication systems. Berman, O., and Sapna, K. P. [3] analyzed optimal control of service rate in a service facility system using inventory for service completion.

4. Model development and solutions

The purpose of our model is to provide a near optimal ordering policy and order quantity (reorder point). i.e., (s, Q) – type for retailers and (0,S)-type for warehouse that minimizes the total logistic and maintenance costs subject to customer service constraints.

Three subsystems need to be analyzed.

(1) Inventory at each retailer.
(2) The demand process of the warehouse.
(3) The inventory at the warehouse.
A synthesis of these subsystems is used to arrive at the final cost-minimizing model.

4.1. Assumptions and Notations

In this model we assume that a single product with constant unit price u, weighing a kilogram(for convenience) is stocked in a distribution system that consists of N identical retailers, one central warehouse and a unique supplier or manufacturer.

Assumptions:

- The periodic (day or week or month, etc.) review strategy to monitor the system and the demand at retailer r (r = 1,2 ,……n) is stochastic with certain probability distribution.
- The demand process is assumed so that the transition from state i to state j is given by the matrix (i=1, 2).
- Whenever the inventory level depletes to , the retailer places an order for Q_r to the warehouse .
- Also we assume that the order quantity is adjusted at the time of replenishment so that the inventory level reaches immediately after replenishment.
- The lead times of each retailer preventive reorder and emergency reorder at level zero are assumed to be single period. This lead time has three components; a constant ordering time of each retailer, waiting time at the warehouse for processing and the transit time. Assume that all retailers behavior are identical (in demand and supply).
- The warehouse's inventory, mean while depletes with the retailer orders and whenever inventory level reaches zero level, instantaneous replenishment is done from the supplier or manufacturer. That is (0, S) policy is assumed at warehouse.

The objective is to minimize the total inventory-logistic costs subject to consumer service constraints.

The following summarizes the notations used in this paper.

- N - the number of retailers.
- s_r - reorder point of the rth retailer () , r = 1,2,3,….N.
- Q_r - variable order quantity of the retailer.
- - transition probability from state i to state j .
- $E_1 = \{ s_0 , s_0-1, \ldots\ldots,1 \}$
- $E_2 = \{ s_r , S_r -1 , \ldots..,2, 1 ,0\}$
- E =
- = the preventive order cost
- = the enforced order cost
- = the holding cost

4.2. Inventory analysis

Let $I_r(t)$, $I_0(t)$ denote the inventory level at retailer r (r = 1,2,….N) and of the warehouse (central) at time 't'(each retailer behaves identical suffix r will do). By the assumption we made for the system considered, the two dimensional stochastic process

$\{(I_r(t) , I_0(t)) ; t \geq 0 \}$ is a Markov process with state space E = $E_1 \times E_2 = \{ s_0 , s_0-1, \ldots, 2, 1 \} \times \{ S_r , S_r -1 , \ldots,2, 1, 0\}$

The Markov process $\{(I_r(t) , I_0(t)) ; t \geq 0 \}$ have finite state space E and the embedded

Markov chain $\{(I_r^{(n)} , I_0^{(n)}); n \geq 0 \}$ where $I_r^{(n)}, I_0^{(n)}$ denote the inventory level (on hand- backorder) at nth decision epoch, the discrete time unit (day or week).

The transition probability for each action state 'a' is defined as

$$p_{(i,q)}^{(j,r)} = \Pr \{ (I_r^{(n)} , I_0^{(n)})=(j,r) / (I_r^{(n)} , I_0^{(n)}) = (i,q) \}$$

The reordering time and the quantity to be ordered are decided at each state of the system say (j, r) $\in$ E. The long-run average cost is used as an economic criterion to find the near optimal policy for the system.

Since E = { (i, q) : i = s_0 , s_0-1, ……,1 ; q = S_r , S_r -1 , …,2, 1, 0} is the state space for the given system, which is a finite set, this problem can be put in the framework of a Discrete-Time Markov Decision Process (DTMDP) . This problem can also be considered a parallel one to system maintenance, in which a piece of equipment is inspected at equal interval of time (unit time : day, week, month or year).

The preventive purchase orders (reorders) are placed at $s_r - 1$, S_r -2 , …..,2, 1 for 1, 2,…, $s_r - 1$ items repectively and an enforced reorder is placed at level '0' for S_r items. For the enforced order the replenishment time is unit (day or week).

Define the actions a =

$$a = \begin{cases} 0 & \text{if no order is placed} \\ 1 & \text{if a preventive order is placed} \\ 2 & \text{if an enforced order is placed for } S_r \text{ items} \end{cases}$$

The set of possible actions on the state space E are given as :

A (s_r ,q) = 0

A (i , r) = { 0,1 } , for 1 $\leq$ i $\leq$ s_r – 1

A (0, q) = { 2 }

The one step transition probabilities $p_{(i,q)}^{(j,r)}(a)$, denote the transition probability for (i,q) to (j,r) when action 'a' is implemented at state i , then

$$p_{(i,q)}^{(j,r)}(0) = t_j \quad for \quad 1 \leq i < s_r - 1$$

$$p_{(i,q)}^{(s_r,q)}(1) = 1 \quad for \quad 1 \leq i \leq s_r - 1$$

$$p_{(0,q)}^{(s_r,q)}(2) = 1 \quad and$$

$$p_{(i,q)}^{(j,r)}(a) = 0 \quad for \quad all \quad other \quad states.$$

The one step costs are given by

$$c_{(i,q)}^{(0)} = ic_h , c_{(i,q)}^{(1)} = c_{p_i} + ic_h , \; c_{(s_r,q)}^{(2)} = c_f , \text{ for } i = S, S\text{-}1, \ldots, 2, 1.$$

The corresponding one step inventory depletion (due to demand) transition matrix is of the form

$$P = (\, p_{(i,q)}^{(j,r)} \,), (i, q), (j, r) \in E \, .$$

5. Policy- iteration procedure :

Fix a stationary policy R, which assigns to each state i a fixed action a= Ri. Under policy R each time the action a = R_i is taken whenever the system is in state (i, q) at a decision epoch. The process {($I_r^{(n)}$, $I_0^{(n)}$) ; n ≥ 0} describing the state of the system at the decision epochs is a Markov chain with one-step transition probabilities

$p_{st}(R_s)$; s , t ∈ E , s = (i, q) , t = (j, r), whenever the policy R is used. The n-step transition probability of this Markov Chain is defined as

$$p_{st}^{(n)}(R) = \Pr\left\{\left(I_r^{(n)}, I_0^{(n)}\right) = (j,r) / \left(I_r^{(0)}, I_0^{(0)}\right) = (i,q)\right\}$$

where $p_{st}^{(1)}(R) = p_{st}(R)$.

By Chapman- Kolmogoroff equations

$$p_{st}^{(n)}(R) = \sum_{u \in E} p_{st}^{(n-1)}(R) p_{ut}(R_u) \quad , n = 1,2,3,......$$

where $p_{st}^{(0)}(R) = 1 \quad for \quad s = t$

and $p_{st}^{(0)}(R) = 0 \quad for \quad all\ s \neq t$

The expected cost function for the system we considered is given by

V_n (s,r) = the total expected cost over the first n decision epochs when the initial state is s = (i, q) and policy R is used.

According to the definition $V_n(s,R) = \sum_{k=0}^{n-1} \sum_{t \in E} p_{st}^{(k)}(R) c_t(R_t)$

where s = (i, q) , t = (j, r) are all in state space E.

We define the average cost function $g_s(R) = \lim_{n \to \infty} \frac{1}{n} V_n(s,R) \quad , s \in E$

The limit defined in the above function exists because of the following theorem:

Theorem A:

For all $s,t \in E$, $\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p_{st}^{(k)}$ always exists. For any t ∈ E ,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p_{tt}^{(k)} = \begin{cases} \frac{1}{\mu_{tt}} & if\ state\ t\ is\ recurrent \\ 0 & if\ state\ t\ is\ transient \end{cases}$$

where μ_{tt} denotes the mean recurrent time from state t to itself.

Here $g_s(R)$ represent the long run average expected cost per time unit when the system is controlled by policy R and initial state s.

The policy improvement strategy we try to apply in our model is described by the following theorem.

Theorem B :

Let g and $v_s, s \in E$ be given numbers. Suppose that the stationary policy $\overline{R}$ has the property $C_s(\overline{R_s}) - g + \sum_{t \in E} p_{st}(\overline{R_s}) v_t \leq v_s$

for each t ∈ E , where s = (i,q) ; t = (j,r)

Then the long-run average cost of policy $\overline{R}$ satisfies

$$g_s(\overline{R}) \leq g \ , s \in E$$

Policy- iteration algorithm :

Step 0: (initialization) Choose a stationary policy R.

Step 1: (value-determination step) For the current rule R, compute the unique solution $\{g(R), v_s(R)\}$ to the following system of

linear equations:

where s = (i, q) ; t = (j, r) and $v_x = 0$ where x is an arbitrary chosen state.

Step 2: (policy-improvement step) For each state $s \in E$, determine an action a_s

yielding the minimum in . $\min_{a \in A(s)} \left\{ c_s(a) - g(R) + \sum_{t \in E} p_{st}(a) v_t(R) \right\}$

The new stationary policy $\overline{R}$ is obtained by choosing $\overline{R} = a_s$ for all $s \in E$ with the conversion that $\overline{R}_s$ is chosen equal to the old action R_s when this action minimizes the policy-improvement quantity.

Step 3: (convergence test) If the new policy $\overline{R}$ = R , then the algorithms is stopped with policy R. Otherwise, go to step 1 with R replaced by $\overline{R}$.

6. Numerical illustrations :

Consider a two-echelon supply chain with one central warehouse and 6 identical retailer. We consider only one retailer r having the following system parameters:

N= 6, Sr =5 , $c_{h=}0.1$, $c_f = 10$, $c_{p1} = 6$, $c_{p2} = 6$, $c_{p3} = 5$, $c_{p4} = 5$.

The transition probability matrix due to demand is given by

$$P = \begin{bmatrix} 0.6 & 0.2 & 0.1 & .05 & .05 & 0.0 \\ 0.7 & 0.2 & .05 & .05 & 0.0 & 0.0 \\ 0.8 & 0.1 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.9 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

Initiate the policy iteration algorithm with the arbitrary policy:

$R^{(1)} = (0,0,0,0,0,2)$, which descirbes the reorder for Sr items only at inventory level (total depletion), with c_f= 10.0.

The policy improvement cost is given by

$$T_i(a,R) = c_i(a) - g(R) + \sum_{(j,r)} p_{(i,q)}^{(j,r)} v_{(j,r)}(R).$$

when current policy is R, $T_i(a,R) = v_i(R)$, for a= R.

Step 1: The' average cost' and relative value for policy , are given by the system of equations:

$$v_5 = 0 - g + 0.6v_5 + 0.2v_4 + 0.1v_3 + 0.1v_2,$$

$$v_4 = 0 - g + 0.7v_4 + 0.2v_3 + 0.05v_2 + 0.05v_1,$$

$$v_3 = 0 - g + 0.8v_3 + 0.1v_2 + 0.05v_1 + 0.05v_0,$$

$$v_2 = 0 - g + 0.9v_2 + 0.05v_1 + 0.05v_0,$$

$$v_1 = 0 - g + 0.9v_1 + 0.1v_0$$

$$v_0 = 10 - g + v_5,$$

$$v_5 = 0.0,$$

Solving the above equations we get the values of

$g(R^{(1)}) = 0.50632$, $v_5(R^{(1)}) = 0$, $v_4(R^{(1)}) = 0.632911$, $v_3(R^{(1)}) = 1.898734$

$v_2(R^{(1)}) = 1.898734$, $v_1(R^{(1)}) = 4.43038$, $v_0(R^{(1)}) = 9.49367$

The policy improvement can be done using the next step of the algorithm and finaly we get an optimal policy $R^{(3)}$=(0,0,0,1,0,0), which suggest that it is optimal to reorder at inventory level s_r =2.

7. Conclusion and future research :

Two echelon inventory in a supply chain has been studied by many researchers and is a fairly recent research work. Most of the earlier work in this direction has been the determination of ordering policies given on order quantity. We approach the problem in a different way. Using Markov Decision Process we use policy iteration algorithm to fix optimal ordering policy by searching on the policy space completely.

We made one important observation that the policy iteration procedure converges quickly to get optimal policy.

In future the sc structure may be changed as single ware-house with multiple non-identical retailers. This will increase the complexity of the problem.

REFERENCE

1. Axs ter , Continuous review policies for multi-level inventory systems with stochastic demand in : S. C. Graves, A. Rinnooy Kan, P. Zipkin (Eds.), Handbook in Operations Research and Management Science, vol.4, Logistics of Production and Inventory, 1993, pp. 175-197. | 2. R. Bellman, (1970), Introduction to Matrix analysis, McGraw Hill, New York. | 3. O. Berman, and Sapna, K. P., Optimal control of service facilities for systems with arbitrarily distributed service times. | 4. B. Deuermeyer, L. B. Schwarz, A model for the analysis of the system service level in warehouse/retailer distribution systems: the identical retailer case in; L. B. Schwarz(Ed.) | Studies in Management Sciences(16), pp-163-193. | 5. A. Federgruen, Centralized planning models, in ; L. B. Schwarz (Ed.), Studies In Management Sciences, volume.16, Multi-Level Production/Inventory Control Systems, North-Holland, Amsterdam, 1993, pp. 133-173. | 6. S. C. Graves, A multi-echelon inventory model for a repairable item with one-for-one replenishment, Management Science 31 (1985) 1247-1256. | 7. R. A. Howard (1960) , Dynamic programming and Markov processes, John Wiley and Sons Inc., New York. | 8. H. L. Lee, K. Moinzadeh, Two-parameter approximations for multi-echelon repairable inventory models with batch ordering policy, IIE Tansactions. 19(1987) 140-147. | 9. H. L .Lee, C. Billington, The evolution of supply-chain management models and practice at Hewlett-Packard, Interfaces 25 (5) (1995) 42-63. | 10. H. L. Lee, A multi-echelon inventory model for repairable items with emergency lateral shipments, Management Science 33 (1987) 1302-1316. | 11. J. A. Muckstadt, A model for a multi-item, multi-echelon, multi-indenture inventory system, Management Science 20 (1973) 472-481. | 12. K. Moinzadeh, H. L. Lee, Batch-size and stocking levels in multi-echelon repairable systems, Management Science 32 (1986) 1567-1581. | 13. L. B. Schwarz, Introduction in : L.B.Schwarz (Ed.), Studies in Management Sciences, vol.16, Multi-Level Production / in Inventory Control Systems, North-Holland, Amsterdam, 1981, pp. 163-193. | 14. C. C. Sherbrooke, METRIC: A multi-echelon technique for recoverable item control, Operations Research 16 (1968) 122-141. | 15. A. J. Stenger, Inventory Decision Framework, in: J. F. Robeson, W .C. Copucino (Eds.). The Logistics Handbook, The Free Press, New York. 1994, 391-409. | 16. F.M. Stenross, G. J. Sweet, Implementing an Integrated Supply Chain in Xerox, Annual Conference Proceedings, Oak Brook, volume 2, III: Council of Logistics Management, 1991, pp.341-351. | 17. A. Svoronos, P. Zipkin, Estimating the performance of multi-level inventory systems, Operations Research 36 (1988) 57-72.