Chapter 7

Solution of \((s, S)\) inventory problems: A Markov Decision Theory Approach

7.1 Introduction

This chapter deviates from the theme of previous chapters in that Markov decision approach to certain classes of inventory problems is discussed here. A large class of problems of sequential decision making under uncertainty can be modeled as stochastic dynamic programs, which, in general, is referred to as Markov Decision Problems. The Markov Decision model is a powerful tool for analyzing probabilistic sequential decision processes. It is a five tuple \((T, I, A, p, c)\) where \(T\) is a point of time known as decision epoch; \(I\) the state space; \(A\) the action space; \(p\) the state transition probability distribution function and \(c\) the instantaneous cost. Decisions or actions are made at certain event occurrence epochs. When we choose an action in a state,
then an immediate cost is incurred and the system moves to another state according to certain transition probability. A solution to a Markov Decision Process is a policy, which is a function from states to actions that minimizes the long-run average costs.

We proceed to model a few inventory models as Markov decision problems. First we formulate those problems. Then the Markov decision approach is employed to compute the optimal solution.

### 7.2 Model Description

Consider an \((s, S)\) inventory system, where demands follow a Bernoulli process with parameter \(p\). Order for replenishment is placed when inventory level drops to \(s\). The time between placing an order and its receipt is distributed geometrically with parameter \(r\). Assume that service time is negligible. At the time of replenishment, the following decisions or actions are made: Replenishment can take place when inventory level is in any one of the states \(i = s, s - 1, s - 2, \ldots, 1, 0\). We consider the model in which replenishment quantity varies according to the on hand inventory. In this situation we have to take decisions on how much to buy at the time of replenishment. We use Markov Decision Process for the solution.

Let \(Q, Q + 1, Q + 2, \ldots, Q + (s - i)\) be the possible replenishment quantities when the inventory level is \(i\) at the replenishment epoch \(0 \leq i \leq s\). Here \(Q = S - s\). When the inventory level is \(s\), the replenishment quantity is \(Q\) with probability \(p_Q^{(s)}\) which is equal to one. Assign probabilities \(p_Q^{(i)}, p_{Q+1}^{(i)}, \ldots, p_{Q+s-i}^{(i)}\) for the replenishment quantity to be \(Q, Q + 1, \ldots, Q + s - i, i = 0, 1, \ldots, s\) where replenishment occurs at inventory level \(i\). Note that \(\sum_{j=Q}^{Q+s-i} p_j^{(i)} = 1, i \in \{0, 1, \ldots s\}\).

The set of possible states of the inventory level process is denoted by \(I =\)
7.2. Model Description

\{0, 1, 2, \ldots, s, s+1, \ldots, S\}.

When the inventory level is \( \geq s+1 \), no action is taken. For each state in \( I \), a set of decisions can be made. Let \( A(s), A(s-1), \ldots, A(1), A(0) \) be the set of possible actions associated with the states \( s, s-1, \ldots, 1, 0 \) respectively. Then \( A(s) = a_{s,1} \), where the replenishment quantity (r.q) is \( Q \) having probability \( p_Q^{(s)} (=1) \); since there is only one choice for purchase quantity).

If replenishment takes place when inventory level is \( s-1 \), then the actions are:

\[
A(s-1) = \begin{cases} 
   a_{s-1,1}, & \text{where the r.q is } Q \text{ with probability } p_Q^{(s-1)} \\
   a_{s-1,2}, & \text{where the r.q is } Q+1 \text{ with probability } p_{Q+1}^{(s-1)}.
\end{cases}
\]

For state \( s-2 \) the actions are:

\[
A(s-2) = \begin{cases} 
   a_{s-2,1}, & \text{where the r.q is } Q \text{ with probability } p_Q^{(s-2)} \\
   a_{s-2,2}, & \text{where the r.q is } Q+1 \text{ with probability } p_{Q+1}^{(s-2)} \\
   a_{s-2,3}, & \text{where the r.q is } Q+2 \text{ with probability } p_{Q+2}^{(s-2)}.
\end{cases}
\]

\vdots

Finally for state 0 the possible actions and the corresponding probabilities are:

\[
A(0) = \begin{cases} 
   a_{0,1}, & \text{where the r.q is } Q \text{ with probability } p_Q^{(0)} \\
   a_{0,2}, & \text{where the r.q is } Q+1 \text{ with probability } p_{Q+1}^{(0)} \\
   \vdots \\
   a_{0,s+1}, & \text{where the r.q is } Q+s \text{ with probability } p_{Q+s}^{(0)}.
\end{cases}
\]

One step transition probabilities are given by

\[
p_{ij}^{(k)}(a_{k,l}) = p_{Q+l-1}^{(k)}, \text{ where } k = 0, 1, \ldots, s-1, s \text{ and } l = 1, 2, \ldots, s+1-k
\]
such that \( \sum_{l=1}^{s+1-k} p^{(k)}_{Q+l-1} = 1. \)

Let the stationary policies corresponding to states \( s, s-1, \ldots, 1, 0 \) be \( R_s, R_{s-1}, \ldots, R_1, R_0 \) respectively. Then \( R_j = \{a_{j,k} : k = 1, 2, \ldots, s+1-j\}; \quad j = s, s-1, \ldots, 0. \)

### 7.3 Description of the problem

Let \( X_n \) be the state of the system at time \( n \); and let \( D_n \) be the decision or action chosen. Then under a given policy \( R \), \( Y_n = (X_n, D_n) \) is a two-dimensional Markov chain with the transition probabilities

\[
P\{X_{n+1} = j, D_{n+1} = d' | X_n = i, D_n = d\} = p(j|i, d)p(d'|j) \quad (7.1)
\]

where \( p(j|i, d) \) is the conditional probability of the chain moving to the state \( j \) at time \( n + 1 \), given the current state is \( X_n = i \) and a decision \( D_n = d \) is taken and \( p(d'|j) \) is the probability of a decision \( D_{n+1} = d' \) being chosen at state \( X_{n+1} = j \). Suppose demand arrival is according to a geometric process with parameter \( p \) and lead time is geometric with parameter \( r \). Also assume that the service time is negligible. Then the one step transition probability matrix of the inventory level process is given by
7.4. The long run average cost per unit time

Since the state space and action sets are finite stationary policies exist. (See Tijms [57]). Among different stationary policies, we look for the optimal one, which minimizes the long run average cost per unit time. Suppose a cost \( C_{X_n,D_n} \) is incurred when the process is in state \( X_n \) and a decision \( D_n \) is made. Being a function of both \( X_n \in \{0,1,\ldots,s\} \) and \( D_n \in \{a,s+1,\ldots,a_{s+1}\} \), \( C_{X_n,D_n} \) is also a random variable. Its long-run average cost per unit time averaging over \( N \) periods is

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} E [C_{X_n,D_n}] = \sum_{i=0}^{s} \sum_{j \in \{Q,Q+1,\ldots,Q+s\}} \pi_{ij} c_{ij}
\]

where \( \pi_{ij} \) is the stationary probability distribution associated with the transition probabilities in (7.1). For an irreducible Markov chain, \( \pi_{ij} \geq 0, \forall i, j \) and \( \sum_{i=0}^{s} \sum_{j \in \{Q,Q+1,\ldots,Q+s\}} \pi_{ij} = 1 \). (see [56]).
Chapter 7. Solution of \((s,S)\) inventory problems: A Markov Decision Theory Approach

7.5 The Optimal Policy and the Policy Improvement Algorithm

Our objective is to find a policy that minimizes the long run average cost. For that purpose, we need to introduce the set of feasible policies and the associated Markov chains, the action sets associated with each state and the immediate cost associated with each state.

Assume that the Markov chain \(Y_n = (X_n, D_n)\) is irreducible. Then there exists a unique equilibrium distribution \(\{\pi_j(R), j \in I\}\). (see [57])

For any \(j \in I\),

\[
\lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} P_{ij}^{(n)}(R) = \pi_j(R),
\]

which is independent of initial state \(i\). The \(\pi_j(R)\) are the unique solution to the system of equilibrium equations

\[
\pi_j(R) = \sum_{i \in \{0,1,\ldots,s\}} p_{ij}(R_i) \pi_i(R), \quad j \in \{Q, Q+1, \ldots, Q+s\}
\]

with \(\sum_{j \in \{Q,Q+1,\ldots,Q+s\}} \pi_j(R) = 1\).

Let \(g(R)\) represent the long run expected average cost per unit time under any given policy \(R\). Then,

\[
g(R) = \sum_{j \in \{Q,Q+1,\ldots,Q+s\}} c_j(R_j) \pi_j(R).
\]

Let \(V^n(i, R)\) denote the total expected cost with \(i\) as the initial state, \(R\) as the stationary policy and evolving over a period of length \(n\). Then we have
7.5. The Optimal Policy and the Policy improvement Algorithm

the recursive formula

\[ V^n(i, R) = c_i(R_i) + \sum_{j \in \{Q, Q+1, \ldots, Q+s\}} p_{ij}(R_i) V^{n-1}(j, R) \]  

(7.2)

It follows that the total expected cost \( V^n(i, R) \) consists of the cost incurred when action \( a = R_i \) is taken in state \( i \) at the first decision epoch and the remaining \( n - 1 \) decision epochs, when the next state is \( j \).

Since the Markov chain under consideration is irreducible, the average cost function \( g_i(R) \) defined by \( g_i(R) = \lim_{n \to \infty} \frac{1}{n} \sum_{i} V^n(i, R) \) is equal to \( g(R) \), independently of the initial state \( i \in \{0, 1, \ldots, s\} \). This relation motivates the heuristic assumption that bias value \( v_i(R), i \in I \), exists such that, for each \( i \in I \),

\[ V^n(i, R) \approx ng(R) + v_i(R) \]  

for large values of \( n \)  

(7.3)

Substituting (7.3) in (7.2), we get

\[ ng(R) + v_i(R) = c_i(R_i) + \sum_{j \in \{Q, Q+1, \ldots, Q+s\}} p_{ij}(R_i) [(n - 1)g(R) + v_j(R)] \]

\[ = c_i(R_i) + (n - 1)g(R) + \sum_{j \in \{Q, Q+1, \ldots, Q+s\}} p_{ij}(R_i) v_j(R) \]

\[ = c_i(R_i) + (n - 1)g(R) + \sum_{j \in \{Q, Q+1, \ldots, Q+s\}} p_{ij}(R_i) v_j(R) \]

i.e.,

\[ g(R) = c_i(R_i) - v_i(R) + \sum_{j \in \{Q, Q+1, \ldots, Q+s\}} p_{ij}(R_i) v_j(R), \]

for \( i = 0, 1, \ldots, s \), with \( V^0(i, R) = 0 \).

Solving this system of equations, we get the long run average cost per unit time \( g(R) \) if policy \( R \) is used. An optimal policy is that of the lowest cost
To obtain the optimal policy, we use an iterative procedure, called policy-improvement algorithm (see [57]). This procedure begins by choosing an arbitrary stationary policy \( R \). Then compute the unique solution \( \{g(R), v_i(R)\} \) to the following system of linear equations:

\[
v_i = c_i(R_i) - g + \sum_{j \in \{Q, Q+1, \ldots, Q+s\}} p_{ij}(R_i)v_j, \quad i \in I
\]

with normalizing equation, \( v_k = 0 \), where \( k \) is an arbitrarily chosen state. In the second step, we can find an improved policy \( \overline{R} \). For that, determine an action \( a_i \), for each state \( i \in I \), which yields the minimum in

\[
\min_{a \in A(i)} \{c_i(a) - g(R) + \sum_{j \in \{Q, Q+1, \ldots, Q+s\}} p_{ij}(a)v_j(R)\}.
\]

Then \( \overline{R} \) is obtained by choosing \( \overline{R}_i = a_i, \forall i \in I \) with the convention that \( \overline{R}_i \) is chosen equal to the old action \( R_i \) when this action minimizes the policy-improvement quantity.

In the third step, if \( \overline{R} = R \), then the algorithm is stopped with policy \( R \). Otherwise, go to the beginning step with \( R \) replaced by \( \overline{R} \).

Since the state space is finite, there are only a finite number of possible stationary policies. Hence after a finite number of iterations, we will be able to reach the optimal policy.

### 7.5.1 Performance measures

We have then the following measures for evaluating performance of the system.

1. Expected replenishment quantity when the inventory level is \( i \)

\[
= \frac{1}{s-i+1}[Q + (Q + 1) + \ldots + (Q + s - i)]
\]
7.5. The Optimal Policy and the Policy improvement Algorithm

= \( Q + \frac{s-i}{2} \)

where \( \frac{1}{s-i+1} \) is the uniform probability that the replenishment quantity is \( Q, Q+1, Q+2, \ldots, Q+(s-i) \), \( i \in \{0,1,\ldots,s\} \).

2. Expected replenishment quantity is given by

\[
ERQ = (Q + \frac{s}{2})(\frac{pr}{1-pr})^s + r \sum_{i=1}^{s} (Q + \frac{s-i}{2}) \frac{(pr)^{s-i}}{(1-pr)^{s-i+1}}.
\]

3. Mean time required for an arrival = \( \sum_{k=1}^{\infty} k(1-p)^{k-1}p \).

4. Mean number of demands lost EL, when the inventory level is zero is given by

\[
EL = \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \sum_{l=n}^{\infty} k nC_k (1-p)^{n-k} p^k q^l,
\]

\[
= \frac{pq}{(1-q)^2}, \quad 0 < p, q < 1.
\]

### 7.5.2 Cost analysis

We define the following costs:
- \( c_0 \) - fixed cost for order placement
- \( c_1 \) - cost per unit item of inventory
- \( c_2 \) - revenue loss due to unit customer lost when inventory is empty.
- \( \alpha, 0 < \alpha < 1 \) - is the discount factor.

For calculating the Expected Total Cost (ETC) at different states and the respective actions, we need expected cycle length from replenishment to replenishment.

Let \( E_{i,j} \) be the expected duration of the cycle with replenishment at state \( i \) and replenishment quantity \( j \), where \( i = \{0,1,\ldots,s-1,s\} \) and \( j = \{Q,Q+1,\ldots,Q+s\} \).

Then Expected Total Cost (ETC) can be calculated as:
### Chapter 7. Solution of \((s, S)\) inventory problems: A Markov Decision Theory Approach

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>ETC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s)</td>
<td>(a_{s,1})</td>
<td>(\frac{c_0}{E_{i,j}} + c_1 Q P_{s}^{(s)}), (i = s, j = Q)</td>
</tr>
<tr>
<td>(s - 1)</td>
<td>(a_{s-1,1})</td>
<td>(\frac{c_0}{E_{i,j}} + c_1 Q P_{s-1}^{(s-1)}), (i = s - 1, j = Q)</td>
</tr>
<tr>
<td>(a_{s-1,2})</td>
<td></td>
<td>(\frac{c_0}{E_{i,j}} + c_1 Q P_{Q+1}^{(s-1)}[Q + \alpha]), (i = s - 1, j = Q + 1)</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>(0)</td>
<td>(a_{0,1})</td>
<td>(\frac{c_0}{E_{i,j}} + c_1 Q P_{Q}^{(0)} + c_2 E L), (i = 0, j = Q)</td>
</tr>
<tr>
<td>(a_{0,2})</td>
<td></td>
<td>(\frac{c_0}{E_{i,j}} + c_1 Q P_{Q+1}^{(0)}[Q + \alpha] + c_2 E L), (i = 0, j = Q + 1)</td>
</tr>
<tr>
<td>(a_{0,3})</td>
<td></td>
<td>(\frac{c_0}{E_{i,j}} + c_1 Q P_{Q+2}^{(0)}[Q + 2\alpha] + c_2 E L), (i = 0, j = Q + 2)</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>(a_{0,s+1})</td>
<td></td>
<td>(\frac{c_0}{E_{i,j}} + c_1 Q P_{Q+s}^{(0)}[Q + s\alpha] + c_2 E L), (i = 0, j = Q + s)</td>
</tr>
</tbody>
</table>

#### 7.6 Numerical Illustration

Let us consider \(s = 3\) and \(S = 7\) so that \(Q = 4\). The states are 0, 1, 2 and 3. The set of all possible actions or decisions on the states are defined as:

\(\{A(i) : i = 0, 1, 2, 3\}\), where \(A(i) = \{a_{i,l} : l = 1, \ldots, 4 - i\}\). Here replenishment quantity is \(3 + l\) having probability \(p_{3+l}^{(i)}\).
Table 7.1: $c_0 = 100$, $c_1 = 10$, $c_2 = 1$, $p = \frac{1}{2}$, $q = \frac{2}{3}$, $\alpha = \frac{1}{3}$, $p_4^{(3)} = 1$, $p_4^{(2)} = p_5^{(2)} = \frac{1}{2}$, $p_4^{(1)} = p_5^{(1)} = p_6^{(1)} = \frac{1}{3}$, $p_4^{(0)} = p_5^{(0)} = p_6^{(0)} = p_7^{(0)} = \frac{1}{3}$

<table>
<thead>
<tr>
<th>States</th>
<th>Actions with costs</th>
<th>Minimum cost</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$a_{0,1}$</td>
<td>37.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{0,2}$</td>
<td>29.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{0,3}$</td>
<td>29.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{0,4}$</td>
<td>28.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>28.64</td>
<td>$a_{0,4}$</td>
</tr>
<tr>
<td>1</td>
<td>$a_{1,1}$</td>
<td>25.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{1,2}$</td>
<td>24.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{1,3}$</td>
<td>23.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>23.80</td>
<td>$a_{1,3}$</td>
</tr>
<tr>
<td>2</td>
<td>$a_{2,1}$</td>
<td>32.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{2,2}$</td>
<td>31.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>31.67</td>
<td>$a_{2,2}$</td>
</tr>
<tr>
<td>3</td>
<td>$a_{3,1}$</td>
<td>52.50</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.2: $c_0 = 100$, $c_1 = 10$, $c_2 = 1$, $p = \frac{1}{2}$, $q = \frac{2}{3}$, $\alpha = \frac{1}{3}$, $p_4^{(3)} = 1$, $p_4^{(2)} = \frac{2}{3}$, $p_5^{(2)} = \frac{1}{3}$, $p_4^{(1)} = \frac{1}{2}$, $p_5^{(1)} = p_6^{(1)} = \frac{1}{3}$, $p_4^{(0)} = p_5^{(0)} = p_6^{(0)} = p_7^{(0)} = \frac{1}{3}$

<table>
<thead>
<tr>
<th>States</th>
<th>Actions with costs</th>
<th>Minimum cost</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$a_{0,1}$</td>
<td>36.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{0,2}$</td>
<td>29.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{0,3}$</td>
<td>29.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{0,4}$</td>
<td>22.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>22.39</td>
<td>$a_{0,4}$</td>
</tr>
<tr>
<td>1</td>
<td>$a_{1,1}$</td>
<td>32.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{1,2}$</td>
<td>20.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{1,3}$</td>
<td>20.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20.00</td>
<td>$a_{1,3}$</td>
</tr>
<tr>
<td>2</td>
<td>$a_{2,1}$</td>
<td>39.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{2,2}$</td>
<td>24.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24.40</td>
<td>$a_{2,2}$</td>
</tr>
<tr>
<td>3</td>
<td>$a_{3,1}$</td>
<td>52.50</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>States</th>
<th>Actions with costs</th>
<th>Minimum cost</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$a_{0,1}$</td>
<td>37.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{0,2}$</td>
<td>29.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{0,3}$</td>
<td>29.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$a_{0,4}$</td>
<td>28.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>28.64</td>
<td>$a_{0,4}$</td>
</tr>
</tbody>
</table>
Chapter 7. Solution of \((s,S)\) inventory problems: A Markov Decision Theory Approach

Table 7.3: \(c_0 = 100, c_1 = 10, c_2 = 1, p = \frac{1}{2}, q = \frac{2}{3}, \alpha = \frac{1}{3}, p_4^{(3)} = 1, p_4^{(2)} = \frac{1}{2}, p_5^{(2)} = \frac{2}{3}, p_4^{(1)} = p_5^{(1)} = \frac{1}{2}, p_6^{(1)} = \frac{1}{2}, p_4^{(0)} = \frac{1}{2}, p_5^{(0)} = p_6^{(0)} = \frac{1}{4}, p_7^{(0)} = \frac{3}{8}\).

<table>
<thead>
<tr>
<th>States</th>
<th>Actions with costs</th>
<th>Minimum cost</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(a_{0,1}) 26.50</td>
<td>(a_{0,2}) 29.82</td>
<td>(a_{0,3}) 29.00</td>
</tr>
<tr>
<td>1</td>
<td>(a_{1,1}) 22.50</td>
<td>(a_{1,2}) 20.83</td>
<td>(a_{1,3}) 31.67</td>
</tr>
<tr>
<td>2</td>
<td>(a_{2,1}) 25.83</td>
<td>(a_{2,2}) 38.89</td>
<td>25.83</td>
</tr>
<tr>
<td>3</td>
<td>(a_{3,1}) 52.50</td>
<td>52.5</td>
<td>(a_{3,1})</td>
</tr>
</tbody>
</table>