Percentile Threshold Policies for Inventory Problems with Partially Observed Markovian Demands

Parisa Mansourifard
Joint work with:
Bhaskar Krishnamachari and Tara Javidi (UCSD)
University of Southern California
parisama @ usc.edu

Informs Annual Meeting, Philadelphia
Nev 04, 2015
Problem Formulation:

- Uncertain demand $\rightarrow$ Markovian Random process
- Inventory level $\rightarrow$ Action
- Supplier $\rightarrow$ Decision-maker

Asymmetries

- Holding cost vs. Shortage cost
- Full observation vs. partial observation (censoring)
Problem Formulation:

\( D_t \): the uncertain demand
\( r_t \): the new order
\( L_t \), the leftover from previous time steps.

The immediate cost:

\[
C(D_t, L_t; r_t) = c_o r_t + \begin{cases} 
  c_h (L_t + r_t - D_t) & \text{if } D_t \leq L_t + r_t \\
  c_s (D_t - L_t - r_t) & \text{if } D_t \geq L_t + r_t,
\end{cases}
\]

\( c_h \) and \( c_s \): holding and shortage cost units
\( c_o \): the ordering cost unit
\( I_t = r_t + L_t \)
\( L_t = (I_{t-1} - D_{t-1})^+ \)

- At each time if \( I_t > D_t \) the decision-maker gets full observation about \( D_t \)
- If \( I_t \leq D_t \) only partial observation about \( D_t \) reveals (i.e. \( D_t \) is censored).
Problem Formulation: Model

- A Discrete-time finite-state Markov process
- The finite horizon by \( T \) and the discrete time steps by \( t = 1, 2, \ldots, T \)
- A known transition matrix, \( P \)
- **Objective:** to decide how many to order at each time s.t. total expected cost is minimized.

Modeled as **Partially Observable Markov Decision Process (POMDP)**
Problem Formulation: Model

- **State:** $D_t \in \mathcal{D} = \{0, 1, ..., M\} \subseteq \mathbb{Z}$

- **State transition:** assumed to be known and stationary and indicated by a transition probability matrix, $P$:
  - an $|\mathcal{D}| \times |\mathcal{D}|$ matrix
  - with elements $P_{i,j} = Pr(D_{t+1} = j | D_t = i), i, j \in \mathcal{D}, \forall t$

- **Selected order:** $r_t \in \mathcal{D}$
**Problem Formulation: Belief**

- **Belief vector \((b_t)\):** the probability distribution of the actual demands
  - Given all past observations
  - \(b_t = [b_t(1), \ldots, b_t(M)]\), with elements of \(b_t(k) = Pr(D_t = k)\)

- **Belief updating:**
  (a) \(O_P\)
  (b) \(O_F\)

State: current demand, Action: Inventory level

---

(a) \(t\) \(t+1\)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(T_r\) \(\ast P\)

(b) \(t\) \(t+1\)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(T_r\) \(\ast P\)

- State, with zero belief
- State, with non-zero belief at \(t\)
- State, with non-zero belief at \(t=1\)
- Actual State
- Selected action at \(t\)
Problem Formulation: Model Summary

$D_t$: actual state (demand), $b_t$: belief vector, $r_t$: action (ordering), $o_t$: observation, $C_t$: immediate cost
Optimal Policy The policy $\pi$ specifies a sequence of functions $\pi_1, \ldots, \pi_T$, $r_t = \pi_t(b_t)$.

Goal: to minimize the total expected cost in the finite horizon $T$

$$\min_{\pi} J_T^\pi(b_1, L_1) = \min_{\pi} \mathbb{E}\left\{ \sum_{t=1}^{T} C(D_t, L_t; r_t) | b_1 \right\},$$

- $b_1$: the initial belief vector
- $L_1 = l_0$: the initial Inventory level
- The optimal policy $\pi^{opt}$: minimizing policy
- It exists since the number of admissible policies are finite.
Dynamic programming (DP)

\[
V_t(b_t, L_t) := \min_{r_t} V_t(b_t, L_t; r_t),
\]

\[
V_T(b_T, L_T; r_T) = \tilde{C}(b_T, L_T; r_T),
\]

\[
V_t(b_t, L_t; r_t) := \tilde{C}(b_t, L_t; r_t) + \mathbb{E}\{V_{t+1}(b_{t+1}, L_{t+1})|L_t, r_t, b_t\}, \quad t < T
\]

\[
r_t^{opt}(b_t, L_t) := \arg \min_{r \in \mathcal{D}} V_t(b_t, L_t; r).
\]

Computationally intractable
Related Works

- Newsvendor problem with i.i.d. demand assumption, e.g. [Ding et al. 2002], [Bensoussan et al. 2009]
- i.i.d with unknown distribution, e.g. [Besbes et al. 2010]
- A partially observed inventory model with the assumption that the demand process and the inventory level are temporally correlated: [Bensoussan et al. 2005]
- Multi-period newsvendor models in which the leftovers at each time step can be carried over to satisfy the future demand: e.g. [Bensoussan et al. 2008], [Chen et al. 2010]
- POMDP problem where the demand is a Markovian process, continuous state, only myopic policy (lower bound on optimal policy): [Bensoussan et al. 2007]
- Does exploitation-exploration trade-off matters? [Besbes et al. 2015]
Our Contributions

- Consider discrete states and actions
- Introducing new heuristics with **percentile threshold (PT)** structure
- Proposing a **heuristic policy with the percentile threshold structure** which outperforms the myopic policy, the best among all PT policies
- Performance bound: policy censoring cost
Percentile Threshold (PT) Policy

\[ r^{PT}(b_t, L_t) = \min \{ r \in D : \sum_{i=0}^{L_t+r} b_t(i) \geq h^{PT}(b_1, L_1) \} \]
A PT Policy: Myopic Policy

\[ r_t^{myopic}(b_t, L_t) := \arg \min_{r \in \mathcal{D}} \bar{C}(b_t, L_t; r) \]

\[ = \min \{ r \in \mathcal{D} : \sum_{i=0}^{L_t+r} b_t(i) \geq (c_s - c_o)/(c_h + c_s) \} \]

\[ J_T^{myopic}(b_1, L_1) = \mathbb{E}\left\{ \sum_{t=1}^{T} C(D_t, L_t; r_t^{myopic}(b_t, L_t)) \mid b_1 \right\} \]
Finite Resolution PT Policy: Heuristic Policy

\[ r^{FRPT}(b_t, L_t) = \min\{r \in \mathcal{D} : \sum_{i=0}^{L_t+r} b_t(i) \geq h^{FRPT}(b_1, L_1)\}, \]

such that

\[ h^{FRPT}(b_1, L_1) = \arg \min_{h^{PT} \in S} J_T^{PT}(b_1, L_1), \]

where \( S \subset [0, 1] \cup \{(c_s - c_o)/(c_s + c_h)\} \). Thus,

\[ J_T^{FRPT}(b_1, L_1) = \mathbb{E}\left\{ \sum_{t=1}^{T} C(D_t, L_t; r_t^{FRPT}(b_t, L_t)) \middle| b_1 \right\} \]
Performance Criteria:

\[ J_T^{\text{myopic}}(b_1, L_1) \geq J_T^{\text{FRPT}}(b_1, L_1) \geq J_T^{\text{opt}}(b_1, L_1) \geq J_T^{\text{uncensored}}(b_1, L_1) \]

The total cost of the uncensored problem:

\[ J_T^{\text{uncensored}}(b_1, L_1) = \mathbb{E}\left\{ \sum_{t=1}^{T} C(D_t, L_t; r_t^{\text{myopic}}(b'_t, L_t)) \mid b_1 \right\} \]

where \( b'_t = I_{D_{t-1}} P \) for \( t = 2, \ldots, T \).

A parametric bound on myopic optimality gap and FRPT optimality gap is their cost of censoring:

\[
\begin{align*}
MCC & := \frac{J_T^{\text{myopic}} - J_T^{\text{uncensored}}}{J_T^{\text{uncensored}}} \\
FCC & := \frac{J_T^{\text{FRPT}} - J_T^{\text{uncensored}}}{J_T^{\text{uncensored}}}
\end{align*}
\]
Simulation results:

Parameters: the holding cost unit $c_h = 0.5$, the ordering cost $c_o = 1$, the transition probabilities given by $P_{i,i-2} = .3; P_{i,i+j} = .1, j \in \{-1,1,3\}; P_{i,i+k} = .2, k \in \{0,2\}, i \in [0,9]$.

Figure: The threshold of PT policies versus shortage cost unit, $c_s$, for $T = 20$. 
Simulation results:

Figure: The threshold of PT policies versus horizon $T$, for $c_s = 3$. 
Simulation results:

Figure: Censoring cost versus shortage cost unit, $c_s$, for $T = 20$. 

Figure: Censoring cost versus shortage cost unit, $c_s$, for $T = 20$. 

Figure: Censoring cost versus shortage cost unit, $c_s$, for $T = 20$. 

Figure: Censoring cost versus shortage cost unit, $c_s$, for $T = 20$. 

Figure: Censoring cost versus shortage cost unit, $c_s$, for $T = 20$. 

Figure: Censoring cost versus shortage cost unit, $c_s$, for $T = 20$. 

Figure: Censoring cost versus shortage cost unit, $c_s$, for $T = 20$. 

Figure: Censoring cost versus shortage cost unit, $c_s$, for $T = 20$. 

Figure: Censoring cost versus shortage cost unit, $c_s$, for $T = 20$.
Simulation results:

Figure: Censoring cost versus horizon $T$, for $c_s = 3$. 
Summary:

- Multi-period Newsvendor problem with censored markovian Demand
- Trade-off between exploration and exploitation
- FRPT policy outperforms myopic policy

For unknown the transition probabilities:
- Data-driven estimator
- Robust Optimization
References


Thank you for your attention