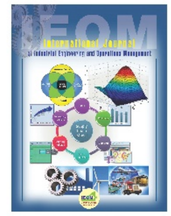
IEOM Society International

International Journal of Industrial Engineering and Operations Management (IJIEOM)

Volume 1, No. 1, May 2019
pp. 32 - 45



A Game Theoretic Approach to Multi-Period Newsvendor Problems with Censored Markovian Demand

Farzaneh Mansourifard
Department of Engineering & Technology
Alzahra University, Tehran, Iran

Parisa Mansourifard and Bhaskar Krishnamachari
Ming Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, California, 90089 USA

Primary Contact: Parisa Mansourifard. mailto:parisama@usc.edu

ABSTRACT

This paper studies the Newsvendor problem for a setting in which (i) the demand is temporally correlated, (ii) the demand is censored, (iii) the distribution of the demand is unknown. The correlation is modeled as a Markovian process. The censoring means that if the demand is larger than the action (selected inventory), only a lower bound on the demand can be revealed. The uncertainty set on the demand distribution is given by only the upper and lower bound on the amount of the change from a time to the next time. We propose a robust approach to minimize the worst-case total cost and model it as a min-max zero-sum repeated game. We prove that the worst-case distribution of the adversary at each time is a two-point distribution with non-zero probabilities at the extrema of the uncertainty set of the demand. And the optimal action of the decision-maker can have any of the following structures: (i) a randomized solution with a two-point distribution at the extrema, (ii) a deterministic solution at a convex combination of the extrema. Both above solutions balance over-utilization and under-utilization costs. Finally, we extend our results to uni-model cost functions and present numerical results to study the solution.

ARTICLE INFO

Received
September 14, 2018
Received in revised form
March 03, 2019
Accepted
May 14, 2019

KEYWORDS

Distribution-free newsvendor problem, Markovian process, Min-max, Game theory, Robust optimization

1. Introduction

Motivated by wide-ranging applications such as inventory management and retail, video delivery over networks using Multiple Description Codes (MDP), congestion control, rate adaptation, spectrum sharing, and provisioning of renewable energy, we study the state-tracking of a Markovian random process with an unknown transition matrix. All the above problems can be modeled as a Newsvendor problem or perishable inventory control problem which has been a research topic for many years (Arrow 1951). The newsvendor model relates by analogy to the situation faced by a newspaper vendor who must decide on how many newspapers to stock since he doesn't know how many demand (customer) he might have, and he knows that the leftover newspapers cannot be sold the next day (it is perishable in some sense). Since then, different solutions under different assumptions have been presented. One of the approaches to tackle such a challenge is formulating the problem as a robust optimization problem. For a complete literature review on robust optimization and its application in inventory control problems, we refer to Gabrel 2014 and Xin 2015a. Most of the works in the literature focus on the fully observable demand with ambiguous distribution (e.g. See 2010, Xin 2013). Among them, some of the works assume the demand is independent and identically distributed (i.i.d.) at different time periods (e.g. Ding 2002, See 2009, Solyali 2016). Some other works (such as Negoescu 2008, Besbes 2013) consider the case where the demand distribution is i.i.d but unknown, and to solve such a problem, they propose a learning process to estimate the distribution and make the decision. In recent years, it has been observed that the demand distribution is not necessarily i.i.d and it can have correlation over time (Xin 2015b, Carrizosa 2016, Natarajan 2017, Tai 2016). For example, In Xin 2015b, the Martingale demand is considered and the minimax optimal policy is explicitly computed in a closed form. Hu 2016 studied the inventory control problem with Markov-modulated demand. In Carrizosa 2016, a robust approach is proposed for the Newsvendor problem with auto regressive (AR) demand with an unknown distribution. Using numerical experiments, they show that the proposal usually outperforms the previous benchmarks in terms of robustness and the average revenue. The distributionally robust version of the inventory problems over the set of distributions satisfying the known information, which is usually mean and covariance of demand, is studied in Natarajan 2017. The authors show that a three-point distribution achieves the worst-case expected profit and derive a closed-form expression for the problem.

Note that in most of these studies, the demand is assumed to be fully observed. However, there are some research papers which study the inventory control problem with censored (partially observed) and temporally correlated demands (non i.i.d.) (e.g. Lu 2008, and Bisi 2011). In these works, a Bayesian scheme is employed to dynamically update the demand distribution for the newsvendor problem with a storable or perishable inventory. As another example, in Bensoussan 2007, a perishable inventory management problem with a memory (Markovian with known transition probabilities) and partially observable demand process is considered. Mansourifard 2013,2017 studied a Newsvendor problem with Markovian and censored demand, with the assumption that the transition probabilities are known, as well.

In practice, demands can be correlated over periods and knowing the exact transition probability is not possible and it needs availability of a lot of data. One approach to reduce this gap is to relax the assumption of knowing the transition probabilities and look at the problem from a lens of game theory. Thus, in this paper, we studied the case where the transition probability matrix is unknown and only the upper and lower bounds on the changes in demands are given.

The contribution of this paper is as follows:

- To our knowledge, this paper is the first work tackling the robust newsvendor problem with temporally correlated demand with censored demand, and uses a game theoretic approach in the solution. We model this problem as a zero-sum repeated game with incomplete information (Sorin 2002, Zamir 1992) and derive the solution in a closed-form.
- We prove that the worst-case distribution of the adversary at each time is a two-point distribution with non-zero probabilities at the lower and upper bound of the uncertainty set.
- Both the possible solutions balance the over-utilization and under-utilization costs. In other words, if the over-utilization cost is larger than the under-utilization cost, the decision-maker assigns a higher probability to the lower bound (for the solution (i)) or chooses a lower action (for the solution (ii)) to behave conservatively. Otherwise, he behaves more aggressively to increase the chance of getting full observation which can be useful in decreasing the future cost.
- We also show that similar results hold for a more general class of cost functions that are uni-model on the difference between the demand and the action.

Related Works

Here, we present the literature related to our work. The perishable inventory control problem in operations research management literature applies to the problem of optimal inventory control to meet uncertain demands for a perishable product (Lu 2005, Bensoussan 2005, Bensoussan 2008, Negoescu 2008, Besbes 2010, Chen 2010, Qin 2011, Bensoussan 2017, etc.). In these problems, the demand for some good is assumed to follow a stochastic process and at the beginning of each decision epoch the decision-maker decides on the inventory level (i.e. how many items to store) in order to satisfy the demand. Mapping the Markov demand to a hidden state and the inventory level to the selected action, the inventory control problem with perishable good is equivalent to that of tracking with asymmetric cost and information. Most of the works in the inventory control literature, e.g. Ding 2002, Bensoussan 2009, assume that the demand process is independent and identical distributed (i.i.d) at different time steps, with the exception of Bensoussan 2007. With this simplifying assumption, the optimal policy is easily shown to coincide with a myopic policy which minimizes the immediate expected cost. Please refer to in Khoudja 1999 and Qin 2011 for more literature reviews.

Most of the inventory control literature (e.g. Ding 2002, and Bensoussan 2009) assume that the demand process is independent and identical distributed (i.i.d) at different time periods. Some prior works (such as Negoescu 2008, Besbes

2013) consider the case where the demand distribution is i.i.d but unknown, so the learning plays an important role in estimating the distribution and making the decision. For instance, in Besbes 2013, the demand distribution is estimated from historical data. They show that the optimal policy has a percentile structure and characterize the implications of partial observations on the performance of the optimal policy in both discrete and continuous settings. However, in recent years, it has been observed that the demand distribution is not necessarily i.i.d and it can have correlation over time (Tai 2016). For example, Hu 2016 studied the inventory control problem with Markov-modulated demand. Note that in most of these papers, the demand is assumed to be fully observed.

In some other literature works such as Lu 2008, Chen 2010, and Bisi 2011, the inventory problem with partially observed (censored) i.i.d. demand has been studied. In Bisi 2011, a Bayesian scheme is employed to dynamically update the demand distribution for the problem with storable or perishable inventory. They show that the Weibull is the only newsvendor distribution for which the optimal solution can be expressed in scalable form. In Lu 2008, the perishable inventory control problem with censored demand is studied in which the demand distribution is assumed to be i.i.d. but unknown. They use Bayesian approach to update the distribution parameters periodically based on the censored historical sales data. Chen 2010 studied the non-perishable inventory control problem with censored and i.i.d. demand. They developed bounds and heuristics for such a problem.

Furthermore, there are some research papers which study the inventory control problem with censored and temporally correlated demands. In Bensoussan 2007, a perishable inventory management problem with memory (Markovian) demand process is considered. In their work, some structural properties of the optimal actions relative to the myopically optimal actions are obtained. And in Bensoussan 2008, they extended the work to the non-perishable inventory. In these papers, the existence of an optimal policy is shown. In their work, some structural properties of the optimal actions relative to the myopically optimal actions are obtained. In Mansourifard 2018, similar problem is studied but for non-perishable inventory control with correlated Markovian demand and known transition probabilities. In this work, we relax the assumption of knowing the transition probability and use robust optimization and game theory to tackle the problem.

In addition, many papers in literature dealt with the newsvendor problem with forecast updating, such as Bradford and Sugrue 1990, Fisher 1994, 2001, Choi 2006, Serel 2012, and Cheaitou 2014. For instance, in Cheaitou 2014, the proposed single-product, stochastic, two-period inventory control model combines demand forecast updating with the flexibility of two supply sources and demand is modeled by two independent, random variables over a two-period selling season. Their findings provide the structure of the second-period conditional optimal policy and analytical insights that characterize the first-period optimal policy. In our work, we study multi-period inventory control with correlated demands over periods.

Further, there are many papers studying multi-period newsvendor problems with non-stationary demands (e.g. Wang 2010, Kim 2015). For instance, in Kim, 2015, the authors proposed a multi-period newsvendor model with non-stationary demand. They formulated it as a multi-stage stochastic programming model and extended the progressive hedging algorithm to optimize the model efficiency. The experimental results show that the proposed multi-stage stochastic programming model performs better than the EOQ and single-period newsvendor models and the total cost can be reduced if we consider transshipments. In our work, we study correlated demand with known transitions.

To our knowledge, our work is the first to study multi-period inventory control problem with correlated demands over periods and unknown transition probabilities and model this problem as a game and derive the optimal min-max solution.

Problem Formulation

We consider a single-item multi-period Newsvendor problem. The newsvendor model is a mathematical model in operations management and applied economics that is used to decide about the optimal inventory level (action) and it typically assumes that the prices are fixed and the demand is uncertain for a perishable product. The decision-maker must select the action (e.g. inventory) r_t to satisfy the demand a_t where $t = 1, \dots, T$ is the time step with the finite horizon T. The goal of the decision maker is to minimize the total expected cost over the horizon.

In this paper, we assume the demand a_t is temporally correlated over time as a Markovian random process given by $a_t = a_{t-1} + \delta_t$ with δ_t as a linear transition of the demand from time step $t - 1$ to t. In general, we have no information about δ_t, however, we assume that is bounded as $\delta_t \in \{-\delta_t^l, \dots, 0, \dots, \delta_t^h\}$ where $-\delta_t^l$, δ_t^h are the lower and upper bounds on the transition, respectively.

As mentioned before, the demand is not necessarily fully observed at each time step. Thus, we consider the case of censored inventory problem in which at each time t, if $r_t > a_t$, the decision-maker gets a full observation about a_t, and if $r_t \leq a_t$, only partial observation about a_t reveals (i.e. a_t is censored).

For a given demand a_t and a selected action r_t, the decision maker faces an immediate cost as:

$$C(a_t, r_t) = \begin{cases} c_u(r_t - a_t) & if\ r_t > a_t \\ c_l(a_t - r_t) & if\ r_t \leq a_t \end{cases} \qquad (1)$$

where c_u and c_l are over-utilization and under-utilization cost coefficients, respectively. The goal is to minimize the total expected cost accumulated over the finite horizon. Since the demand is unknown, this goal could be formulated as a min-max optimization problem:

$$C_1^*\ (r_1^l\ , r_1^h) = \min_{p_{r_1},\dots,p_{r_T}} \quad \max_{p_{a_1},\dots,p_{a_1}} \sum_{t=1}^{T} \mathbb{E}_{r_t}\mathbb{E}_{a_t}[C(a_t, r_t)|F_t], \qquad (2)$$

where F_t is the information available to the player before time t, and p_{r_t} and p_{a_t} are probability distribution functions (PDFs) of the action and the demand at time t, respectively. Since the transition of the demand is bounded, the action would also be bounded in $[r_t^l\ , r_t^h]$. Let $C_t^*\ (r_t^l\ , r_t^h)$ indicate the min-max expected cost-to-go from time t onward where $C_1^*\ (r_1^l\ , r_1^h)$ is given by (2).

After taking the action r_t and the observation revealed about the demand, the bounds on the actions can be updated for the next time step:

$$r_{t+1}^l = \begin{cases} a_t - \delta_t^l & if\ r_t > a_t \\ r_t - \delta_t^l & if\ r_t \leq a_t \end{cases}$$

$$r_{t+1}^h = \begin{cases} a_t + \delta_t^h & if\ r_t > a_t \\ r_t^h + \delta_t^h & if\ r_t \leq a_t \end{cases}$$

This update is achieved as follows: Assume the bound on r_t, a_t is given as $r_t^l \leq r_t, a_t \leq r_t^h$. If no observation happens, the bound will be updated based on the bounds on the transitions, i.e. $r_t^l - \delta_t^l \leq r_{t+1} \leq r_t^h + \delta_t^h$. But, in the studied problem, depending on the relation between r_t and a_t, the decision-maker can have full or partial observations. If the selected action is higher than the demand, $r_t > a_t$, the full observation about the demand a_t will be revealed, so the decision-maker can narrow down the bounds to the uncertainty about the transitions, i.e. $a_t - \delta_t^l \leq r_{t+1} \leq a_t + \delta_t^h$. On the other hand, if the selected action is lower than the demand, $r_t \leq a_t$, only partial observation will be revealed, i.e. the demand is not lower than r_t, and this will only affect the lower bound and not the upper bound. Thus, the bound at $t+1$ will be $a_t - \delta_t^l \leq r_{t+1} \leq r_t^h + \delta_t^h$.

Now, the goal is to find the best actions r_t^* that achieves the min-max at (2). Fig. 1 shows an example of the demand path and the sequence of the taken actions with the corresponding costs and the bounds on the actions.

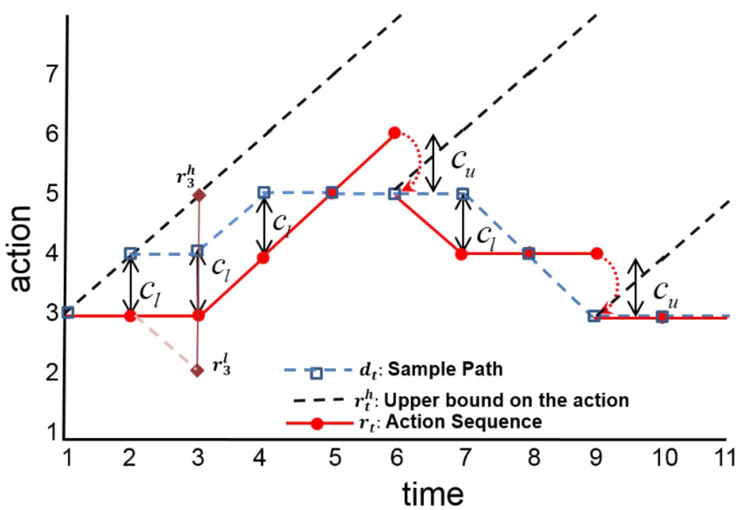


Figure 1. An example of the demand path and the sequence of the taken actions.

The Game Theoretic Approach

This can be modeled as a game between the adversary and the decision-maker. The sufficient statistic for F_t at each time step t, is the support $\{r_t^l\ , r_t^h\}$ and the adversary chooses the probability distribution of $a_t \in \{r_t^l\ , r_t^h\}$ to maximize the expected cost-to-go for the selected distribution of the action r_t. The solutions are given in the following theorem (and we prove them using induction):

Theorem 1-a) The worst-case distributions p_{at} are two-point distributions with non-zero probabilities at r_t^l and r_t^h , for all $t = 1, \dots, T$, i.e. $p_{a_t}(a)$=0 for $a \neq r_t^l, r_t^h$.. And there are two possible solution to the min-max problem:

1. $r_t^* = \frac{c_u r_t^l + y_t r_t^h}{c_u + y_t}$, $p_{r_t}(r_t^*) = 1$, and p_{a_t} can be any two-point distribution at extrema
2. $p_{r_t}(r_t^l) = 1 - p_{r_t}(r_t^h) = \frac{c_u}{c_u + c_l}$, and

$$p_{a_t}(r_t^l) = 1 - p_{a_t}(r_t^h) = \frac{y_t}{c_u + y_t},$$

where,

$$y_t = c_l + \frac{c_u y_{t+1}}{c_u + y_{t+1}} \;, \forall t = 1, \dots, T-1$$
$$y_T = c_l.$$

Theorem 1-b) The min-max cost-to-go at time t is obtained as:

$$C_t^*(r_t^l, r_t^h) = C_t^*(r_t^h - r_t^l) = \Delta_{t+1} + \frac{c_u y_t}{c_u + y_t}(r_t^h - r_t^l), \qquad (3)$$

where,

$$\Delta_t = \Delta_{t+1} + \frac{c_u y_t}{c_u + y_t}(\delta_t^h + \delta_t^l), \forall t = 1, \dots, T-1,$$
$$\Delta_T = \frac{c_u y_T}{c_u + y_T}(\delta_T^h + \delta_T^l).$$

If we assume that at $t = 0$ we had the full observation and $r_0^* = a_0$, then $r_1^h - r_1^l = \delta_1^h + \delta_1^l$. Therefore, the total min-max cost can be obtained as:

$$C_1^* = \Delta_1 = \sum_{\tau=1}^{T} \frac{c_u y_t}{c_u + y_t}(\delta_t^h + \delta_t^l), \qquad (4)$$

otherwise, the total min-max cost would be a functional of initial uncertainty range of $r_1^h - r_1^l$ as given in (3). The min-max optimal policy given in Theorem 1-a is illustrated in Algorithm 1 and the backward recursion to solve min-max cost given in Theorem 1-b is summarized in Algorithm 2.

Algorithm 1: Optimal min-max policy

1: **Given parameters** c_u, c_l, horizon T **and** δ_t^h, δ_t^l **for** $t = 1, \dots, T$

2: $y_T = c_l$

3: **for** $t = T-1, \dots, 1$ **do**

4: $y_t = c_l + \frac{c_u y_{t+1}}{c_u + y_{t+1}}$

5: **end for**

6: Start with the initial bound $[r_1^l, r_1^h]$,

7: **for** $t = 1, \dots, T$ **do**

8: Choose the optimal action in either of the following two ways:

1) $r_t^* = \frac{c_u r_t^l + y_t r_t^h}{c_u + y_t}$, or
2) $a_t = r_t^l$ w.p. $\frac{c_u}{c_u + c_l}$, and $a_t = r_t^h$ w.p. $\frac{c_l}{c_u + c_l}$

9: The worst-case demand is selected as:
$a_t = r_t^l$ w.p. $\frac{y_t}{c_u + y_t}$, and $a_t = r_t^h$ w.p. $\frac{c_u}{c_u + y_t}$

10: **If** $r_t^* > a_t$: Full observation

11: Update the bounds as $[r_{t+1}^l, r_{t+1}^h] \rightarrow [a_t - \delta_t^l, a_t + \delta_t^h]$

12: **Else If** $r_t^* \leq a_t$: Partial observation

13: Update the bounds as $[r_{t+1}^l, r_{t+1}^h] \rightarrow [r_t^* - \delta_t^l, r_t^h + \delta_t^h]$

Algorithm 2: Calculating the min-max cost-to-go

1: **Given parameters** c_u, c_l, horizon T **and** δ_t^h, δ_t^l **for** $t = 1, \dots, T$

2: $y_T = c_l$

3: **for** $t = T-1, \dots, 1$ **do**

4: $\quad y_t = c_l + \frac{c_u y_{t+1}}{c_u + y_{t+1}}$

5: **end for**

6: $\Delta_T = \frac{c_u y_T}{c_u + y_T}(\delta_T^h + \delta_T^l)$

7: **for** $t = T-1, \dots, 1$ **do**

8: $\quad \Delta_t = \Delta_{t+1} + \frac{c_u y_t}{c_u + y_t}(\delta_t^h + \delta_t^l)$

9: **end for**

10: $C_1^* = \Delta_1$

Proof:

We use induction to prove the both parts of the Theorem 1. First, at horizon T, we need to solve the single-period version of this problem:

$$\min_{p_{r_T}} \max_{p_{a_T}} \mathbb{E}_{r_T}\mathbb{E}_{a_T}[C(a_T, r_T)] = \mathbb{E}_{r_T}[\int_{x=r_T^l}^{r_T} c_u(r_T - x)p_{a_T}(x)dx + \int_{x=r_T}^{r_T^h} c_l(x - r_T)p_{a_T}(x)dx].$$

First, we will show that the PDF, p_{a_T}, maximizing the above equation is a two-point distribution where only $p_{a_T}(r_T^l)$ and $p_{a_T}(r_T^h)$ are non-zero. To show that, let us choose two random actions $r, r' \in [r_T^l, r_T^h]$ and see how the distribution of probability mass on these two points can affect the adversarial p_{a_T}. Without loss of generality we assume that $r < r'$. We assume that the probability mass of $p_{r_T}(r) + p_{r_T}(r')$ can be distributed between r, r' with the ratio of $X, 1 - X$, i.e. we assign $X = \frac{p_{r_T}(r')}{p_{r_T}(r) + p_{r_T}(r')}$ portion to r' and $1 - X$ portion to r and calculate the cost function for all possible $a_T = a$ and derive the worst-case p_{a_T} for different values of X that can maximize the cost function. The cost function can be written as:

$$\mathbb{E}_{r_T}[C(a, r_T)] = \left(p_{r_T}(r) + p_{r_T}(r')\right)\left(XC_T(a, r') + (1 - X)C_T(a, r)\right) + (1 - p_{r_T}(r) - p_{r_T}(r'))\mathbb{E}_{r_T \neq r, r'}[C(a, r_T)]$$

Since we only look at the two points of r, r', we only need the first part of the above equation. Thus, Fig. 2-a) shows the cost functions of $XC_T(a, r') + (1 - X)C_T(a, r)$ for all a and the two points of $r, r' \in [r_T^l, r_T^h]$, versus the ratio X. For instance, for $a = r_T^h$, the cost function equals to $XC_T(r_T^h, r') + (1 - X)C_T(r_T^h, r) = Xc_u(r_T^h - r') + (1 - X)c_u(r_T^h - r)$. The adversarial demand distribution tries to maximize the cost function, and as it is shown in Fig. 2-a), there is a point X^* that for all $X < X^*$ the worst-case demand is $a = r_T^h$ and for all $X > X^*$ the worst-case demand is $a = r_T^l$. For $X = X^*$ any of the $a = r_T^h$ or $a = r_T^l$ will achieve the maximum value of the cost, so the adversarial never selects any action except the bounds of r_T^h or r_T^l. This proves that the worst-case demand distribution at horizon p_{a_T} is a two-point distribution on the bounds. And next, we need to find the best distribution of p_{r_T} that tries to minimize the cost.

We assume that the PDF p_{a_T} is divided between two bounds by assigning Y portion to r_T^h and $1 - Y$ to r_T^l and we need to find the value of Y, such that $p_{a_T}(r_T^h) = Y$ and $p_{a_T}(r_T^l) = 1 - Y$. Fig. 2-b) shows the cost functions of $YC_T(r_T^h, r) + (1 - Y)C_T(r_T^l, r)$ for different actions of r to find the worst-case Y and p_{r_T}. For each value of Y, the decision-maker selects the action r that minimized the cost function. As it is shown in the figure, there is a point Y^* that for all $Y < Y^*$ the optimal action is $r = r_T^l$ and for all $Y > Y^*$ the optimal action is $r = r_T^h$. For $Y = Y^*$ any of the $r = r_T^h$ or $r = r_T^l$ will achieve the minimum value of the cost.

We can calculate Y^* as follows: Y^* is the point where the cost functions of $r = r_T^h$ and $r = r_T^l$ are equal.

$$Y^* C_T(r_T^h, r_T^h) + (1 - Y^*)C_T(r_T^l, r_T^h) = Y^* C_T(r_T^h, r_T^l) + (1 - Y^*)C_T(r_T^l, r_T^l)$$

$$(1-Y^*)c_u(r_T^h - r_T^l) = Y^* c_l(r_T^h - r_T^l)$$

$$Y^* = \frac{c_u}{c_l + c_u}$$

And the optimal cost function can be calculated as:

$$C_T^*(r_T^l, r_T^h) = Y^* c_l(r_T^h - r_T^l) = \frac{c_l c_u}{c_l + c_u}(r_T^h - r_T^l) \qquad (5)$$

Therefore, the worst-case distribution of demand equals to $p_{a_T}(r_T^h) = \frac{c_u}{c_l + c_u}$ and $p_{a_T}(r_T^l) = \frac{c_l}{c_l + c_u}$. The PDF p_{r_T} that can minimize the cost is either a two-point distribution on the bounds or a deterministic optimal action r_T^* which can achieve the same cost value regardless of chosen Y. The optimal action r_T^* is an action $r_T^l < r_T^* < r_T^h$ that can achieve the same cost function as given in (5):

$$c_l(r_T^h - r_T^*) = c_u(r_T^* - r_T^l) = \frac{c_l c_u}{c_l + c_u}(r_T^h - r_T^l)$$

$$r_T^* = \frac{c_l r_T^h + c_u r_T^l}{c_l + c_u}$$

An illustration of the proof is shown in figure 3. The cost function given in (1) will increase if the distance between demand and action is larger (the goal of adversary) and it will decrease if they are closer (the goal of decision-maker). Thus, when $X > X^*$ (which means higher probability for the larger action) the adversary will pick the smallest possible demand to increase the cost (fig. 3-left). When $X < X^*$, the adversary picks the largest possible demand (fig. 3-middle). And if $X = X^*$ the cost will be the same for demands being on the extreme values, so any two-point distribution will achieve the same cost (fig. 3-right). The optimal policy can be illustrated in the same way.

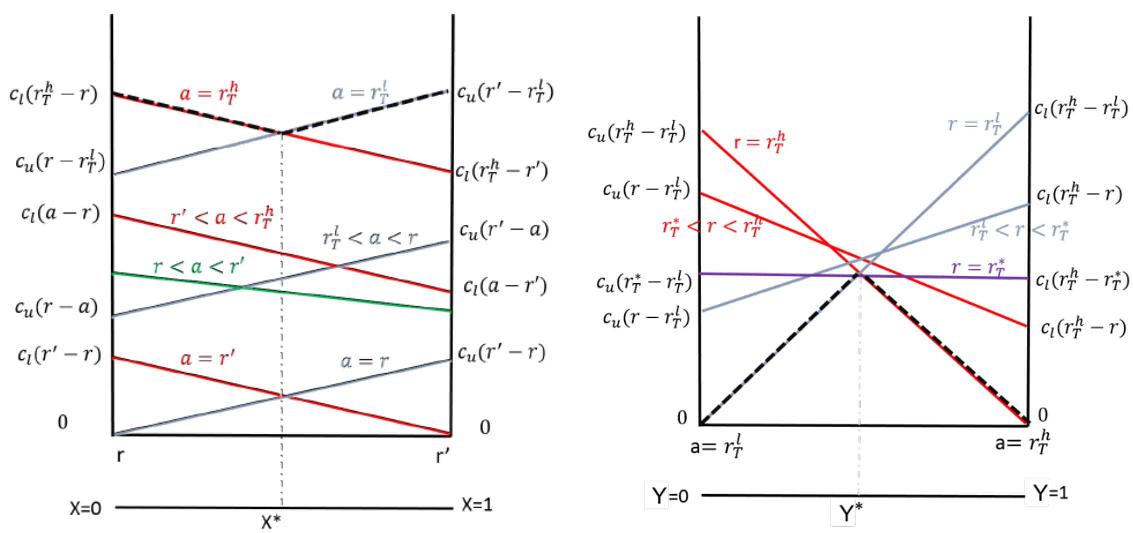


Figure 2. a) The cost functions for all a and any two pairs of r, r'; b) The cost functions of all $r \in \{r_T^l, r_T^*, r_T^h\}$ and any two-point distribution of p_{a_T} at r_T^l, r_T^h.

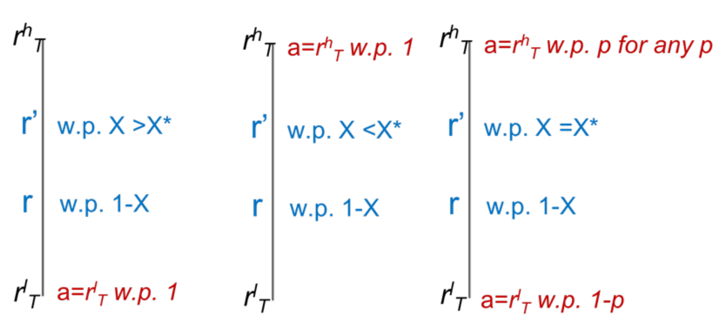


Figure 3. Illustration of the proof, left) when $X > X^*$ there is a higher chance of having a higher action, so the adversary will pick the smallest possible demand to increase the cost.

Now for time steps $t = 1, \ldots, T-1$, we use induction to find the best action distributions (or best deterministic action) and the worst-case adversary distributions. If the solutions are valid for $t + 1$, for time t we have:

$$\begin{aligned}\min_{\mathbb{P}_{r_T}} \max_{\mathbb{P}_{a_T}} \mathbb{E}_{r_t}\Big[\int_{x=r_t^l}^{r_t} c_u(r_t - x)p_{a_t}(x)dx + \int_{x=r_t}^{r_t^h} c_l(x - r_t)p_{a_t}(x)dx \\ + \int_{x=r_t^l}^{r_t} p_{a_t}(x)C_{t+1}^*(x - \delta_{t+1}^l, x + \delta_{t+1}^h)dx \\ + \int_{x=r_t}^{r_t^h} p_{a_t}(x)dx.\, C_{t+1}^*(r_t - \delta_{t+1}^l, r_t^h + \delta_{t+1}^h)\Big]\end{aligned}$$

Since $C_{t+1}^*(y,z) = C_{t+1}^*(z-y) = \Delta_{t+2} + \frac{c_u y_{t+1}}{c_u + y_{t+1}}(z-y)$,

$$\begin{aligned}&\min_{\mathbb{P}_{r_t}} \max_{\mathbb{P}_{a_t}} \mathbb{E}_{r_t}\Big[\int_{x=r_t^l}^{r_t} c_u(r_t - x)p_{a_t}(x)dx + \int_{x=r_t}^{r_t^h} c_l(x - r_t)p_{a_t}(x)dx \\ &\quad + \int_{x=r_t^l}^{r_t} p_{a_t}(x)dx.\left(\Delta_{t+2} + \frac{c_u y_{t+1}}{c_u + y_{t+1}}(\delta_{t+1}^l + \delta_{t+1}^h)\right) \\ &\quad + \int_{x=r_t}^{r_t^h} p_{a_t}(x)dx.\left(\Delta_{t+2} + \frac{c_u y_{t+1}}{c_u + y_{t+1}}(r_t^h - r_t + \delta_{t+1}^l + \delta_{t+1}^h)\right)\Big] \\ &= \Delta_{t+1} + \min_{\mathbb{P}_{r_t}} \max_{\mathbb{P}_{a_t}} \mathbb{E}_{r_t}\Big[\int_{x=r_t^l}^{r_t} c_u(r_t - x)p_{a_t}(x)dx + \int_{x=r_t}^{r_t^h} c_l(x - r_t)p_{a_t}(x)dx \\ &\quad + \int_{x=r_t}^{r_t^h} p_{a_t}(x)dx.\frac{c_u y_{t+1}}{c_u + y_{t+1}}(r_t^h - r_t)\Big]\end{aligned}$$

In other words,

$$C_t^*(r_t^l, r_t^h) = \Delta_{t+1} + \min_{\mathbb{P}_{r_t}} \max_{\mathbb{P}_{a_t}} \mathbb{E}_{r_t}\mathbb{E}_{a_t}[C_t'(a_t, r_t)],$$

where,

$$C_t'(a_t, r_t) = \begin{cases} c_u(r_t - a_t) & \text{if } r_t > a_t \\ c_l(a_t - r_t) + \frac{c_u y_{t+1}}{c_u + y_{t+1}}(r_t^h - r_t) & \text{if } r_t \le a_t \end{cases}$$

Figure 4-a) shows the cost functions $XC_t'(a,r) + (1-X)C_t'(a,r')$; WLOG we assume $r < r'$, $X = \frac{p_{r_t}(r')}{p_{r_t}(r) + p_{r_t}(r')}$. Note that to get the actual cost function we should add all of them by Δ_{t+1}. As it is shown in Fig. 4-a), we can ignore the adversary actions of $a \in (r_t^l, r_t^h)$ and the worst-case demand distribution is a two-point distribution on the bounds. Now to find the best distributions of the decision-maker, in Fig. 4-b) we plot the graph of the cost functions $YC_t'(\mathrm{r_T^h}, \mathrm{r}) + (1-\mathrm{Y})C_t'(\mathrm{r_T^l}, \mathrm{r})$ for different values of r and find the best probability density for $p_{a_t}(r_t^h) = Y$ and $p_{a_t}(r_t^l) = 1 - Y$.

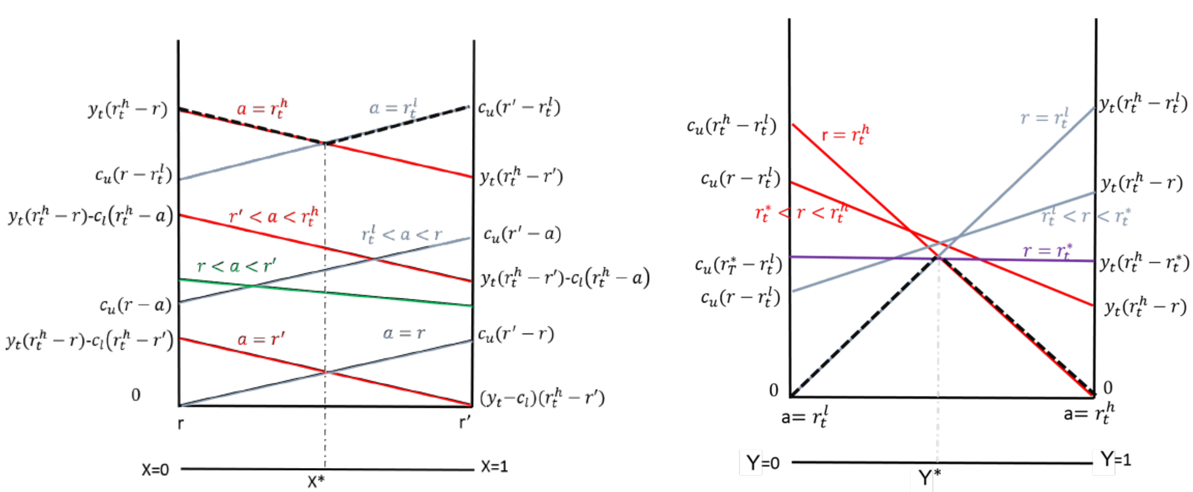


Fig. 4. a) The cost functions for all a and any two pairs of r, r'; b) The cost functions of all $r \in \{r_t^l, r_t^*, r_t^h\}$ and any two-point distribution of p_{a_T} at r_t^l, r_t^h.

And the worst-case distribution is a two-point distribution at r_t^l and r_t^h. Therefore:

$$\begin{aligned} C_t^*(r_t^l, r_t^h) &= p_{a_t}(r_t^l) c_u (r_t^* - r_t^l) + (1 - p_{a_t}(r_t^l)) c_l (r_t^h - r_t) \\ &+ p_{a_t}(r_t^l) \left[\Delta_{t+2} + \frac{c_u y_{t+1}}{c_u + y_{t+1}} (\delta_{t+1}^l + \delta_{t+1}^h) \right] \\ &+ \left(1 - p_{a_t}(r_t^l)\right) \left[\Delta_{t+2} + \frac{c_u y_{t+1}}{c_u + y_{t+1}} (r_t^h - r_t^* + \delta_{t+1}^l + \delta_{t+1}^h) \right] \\ &= p_{a_t}(r_t^l) c_u (r_t^* - r_t^l) + \left(1 - p_{a_t}(r_t^l)\right) \left(c_l + \frac{c_u y_{t+1}}{c_u + y_{t+1}} \right) (r_t^h - r_t) + \Delta_{t+2} \\ &= p_{a_t}(r_t^l) c_u (r_t^* - r_t^l) + \left(1 - p_{a_t}(r_t^l)\right) y_t (r_t^h - r_t) + \Delta_{t+2} \\ &= c_u (r_t^* - r_t^l) + \Delta_{t+2}. \end{aligned}$$

which results in the optimal actions given in Theorem 1-a and the min-max expected cost given in Theorem 1-b, thus completes the proof of Theorem 1.

Extension of Cost Function

We can get similar results for general form of uni-modal cost functions, given by:

$$C(a_t, r_t) = \begin{cases} C_u(r_t - a_t) & if\ r_t > a_t \\ C_l(a_t - r_t) & if\ r_t \leq a_t \end{cases}$$

where $C_u(y)$and $C_l(y)$ are increasing functions of y.

Lemma 1-a) The worst-case distribution at all time steps $t = 1, \dots, T$ are two-point distributions

$p_{a_t}(r_t^l) \neq 0$ and $p_{a_t}(r_t^h) \neq 0$.

Lemma 1-b) The min-max expected cost has the following property: $C_t^*\ (y + x, z + x) = C_t^*\ (y, z)$.

Proof: The min-max cost-to-go at time t is given by:

$$\begin{aligned} C_t^*(r_t^l, r_t^h) = \min_{r_t} \max_{\mathbb{P}_{a_t}} \int_{x=r_t^l}^{r_t} p_{a_t}(x) [c_u(r_t - x) + C_{t+1}^*(x - \delta_t^l, x + \delta_t^h)]\, dx \\ + \int_{x=r_t}^{r_t^h} p_{a_t}(x) [c_l(x - r_t) + C_{t+1}^*(r_t - \delta_t^l, r_t^h + \delta_t^h)]\, dx \end{aligned}$$

At horizon T the worst-case distribution is a two-point distribution: $r_T^* = \{r: C_u(r - r_T^l) = C_l(r_T^h - r)\}$

and the expected cost equals:

$$C_T^*(r_T^l, r_T^h) = C_l(r_T^h, r_T^*) = C_u(r_T^*, r_T^l).$$

This shows that Lemma 1 is true at $t = T$, now if it is true at $t + 1$, for time t we have:

$$\begin{aligned} C_t^*(r_t^l, r_t^h) = \min_{r_t} \max_{\mathbb{P}_{a_t}} \int_{x=r_t^l}^{r_t} p_{a_t}(x) [c_u(r_t - x) + C_{t+1}^*(r_t^l - \delta_t^l, r_t^l + \delta_t^h)]\, dx \\ + \int_{x=r_t}^{r_t^h} p_{a_t}(x) [c_l(x - r_t) + C_{t+1}^*(r_t - \delta_t^l, r_t^h + \delta_t^h)]\, dx \end{aligned}$$

where we replace x at $C_{t+1}^*(x - \delta_t^l, x + \delta_t^h)$ with r_t^l. From the above equation, it is obvious that: (i) the worst-case distribution is a two-point distribution, and (ii) if we add a fixed value to r_t^l and r_t^h the minimizing r_t will be added with the same amount and thus C_t^* $(r_t^l+x; r_t^h+x) = C_t^*$ (r_t^l, r_t^h) for any x.

And recursively: $r_t^* = \{r: C_l(r_t^h - r) + C_{t+1}^*(r - \delta_t^l, r_t^h + \delta_t^h) = C_u(r - r_t^l) + C_{t+1}^*(r_t^l - \delta_t^l, r_t^l + \delta_t^h)\}$

or a randomized solution as follows:

$$p_{r_t}(r_t^l) = 1 - p_{r_t}(r_t^h) = \frac{r_t^h - r_t^*}{r_t^h - r_t^l},$$

Or any combination of non-zero probabilities at $r_t \in \{r_t^l, r_t^*, r_t^h\}$ which proves Lemma 1-a. And the expected cost equals:

$$C_t^*(r_t^l, r_t^h) = C_l(r_t^h - r_t^*) + C_{t+1}^*(r_t^* - \delta_t^l, r_t^h + \delta_t^h) = C_u(r_t^* - r_t^l) + C_{t+1}^*(r_t^l - \delta_t^l, r_t^l + \delta_t^h).$$

This proofs Lemma 1-b.

Numerical Results

In this section, we present some numerical results for the min-max cost and the optimal solution given in Theorem 1 and study the effect of different parameters on the min-max cost and the optimal solution. The best action given in Theorem 1-a as r_t^* is a function of y_t. Choosing a larger y_t means that the under-utilization would be more important and a more aggressive action is preferable. At horizon $t = T$, $y_T = c_l$ is the under-utilization cost.

To see how the best solution behaves for different time steps, we plot the y_t versus time step t in the Fig. 5 for different over utilization coefficients c_u. As all plots show, y_t is larger for smaller time steps, i.e. the optimal min-max policy prefers to behave more aggressive in the earlier time steps and when it gets closer to the horizon it behaves more conservatively. For instance, for $c_u = 2$, the min-max policy chooses the under-utilization weight y_t equal to 2 (i.e. it puts twice weight on under-utilization compared to the over-utilization). For larger c_u the selected y_t is larger and it is higher for earlier time steps.

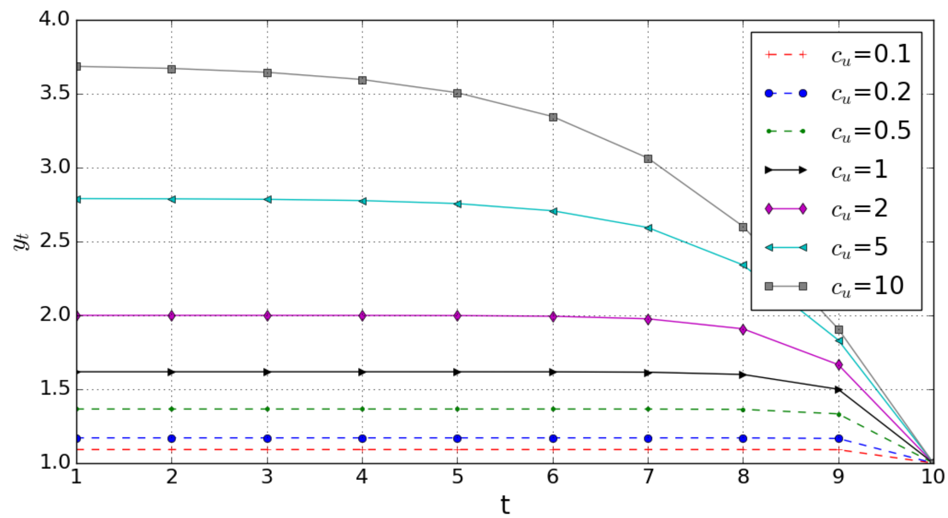


Figure 5. The under-utilization weight y_t versus time step t for different c_u, for $T = 10, c_l = 1, \delta_t^h = \delta_t^l = 1$.

Figure 6 shows Δ_t given in the min-max cost-to-go in (3) versus time step t for different c_u. For larger c_u, Δ_t is larger and the cost-to-go can be calculated given (3). The slope of the plots increases by c_u which means for larger c_u the difference between Δ_t and Δ_{t+1} is greater and the total min-max cost is higher for higher c_u.

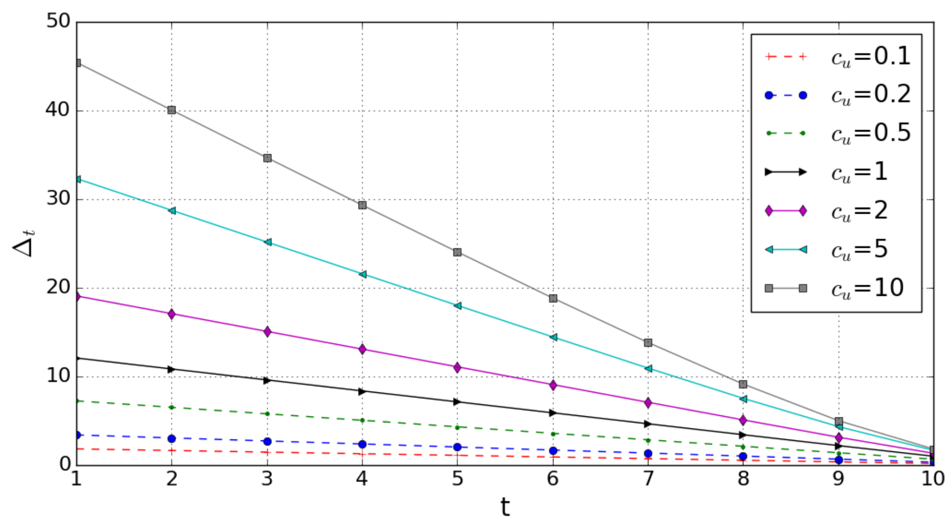


Figure 6. Δ_t given in the min-max cost-to-go in (3) versus time step t for different c_u, for $T = 10, c_l = 1, \delta_t^h = \delta_t^l = 1$. Note that total min-max cost-to-go is equivalent to Δ_1.

Further, the total min-max cost versus c_u is given in Fig. 7 for different time horizons. The total cost is larger for larger horizons and by increasing c_u the total cost will increase.

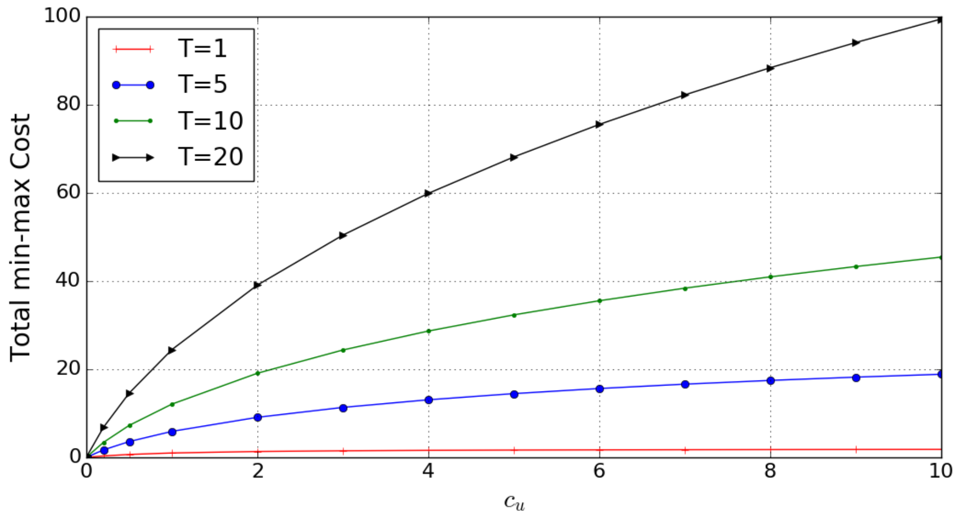


Figure 7. The total min-max cost-to-go which equivalent to Δ_1, given in (4), versus c_u, for different time horizon T, for $c_l = 1, \delta_t^h = \delta_t^l = 1$.

The total min-max cost versus $\delta_t^h = \delta^h$ is given in Fig. 8 for different time horizons. We fix $\delta_t^l = 1$ and assume that δ_t^h is the same for different time steps. By increasing δ^h, the total cost will increase linearly. The linear relation can also be confirmed by equation (4) for fixed bounds of $\delta_t^h + \delta_t^l = \delta^h + 1$.

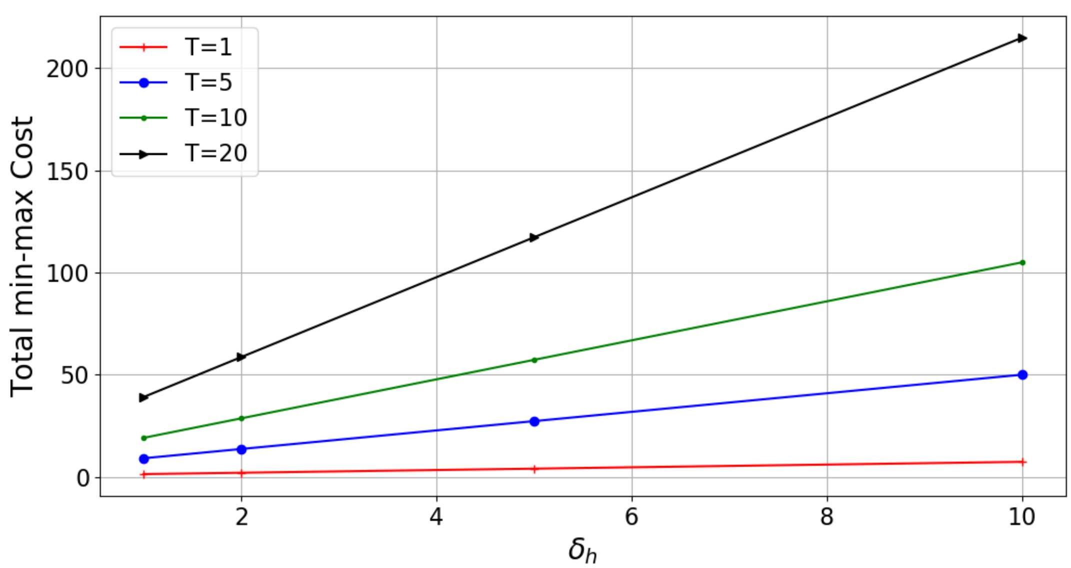


Figure 8. The total min-max cost-to-go which equivalent to Δ_1, given in (4), versus δ^h, for different time horizon T, for $c_l = 1, c_u = 1, \delta_t^l = 1$.

To study the effect of bounds on the cost-to-go function, Fig. 9 shows the changes of Δ_t versus t for different changes of δ_t^h. When δ_t^h is constant (for instance, $\delta_t^h = 2$ as shown in puple color in the figure), Δ_t is linearly decreasing over time. And when δ_t^h is exponentially increasing by time, i.e. $\delta_1^h = 1$, $\delta_2^h = 2$, $\delta_3^h = 4$, …, Δ_t will be exponentially decreasing and Δ_1, which is equivalent to the total min-max cost, will be larger.

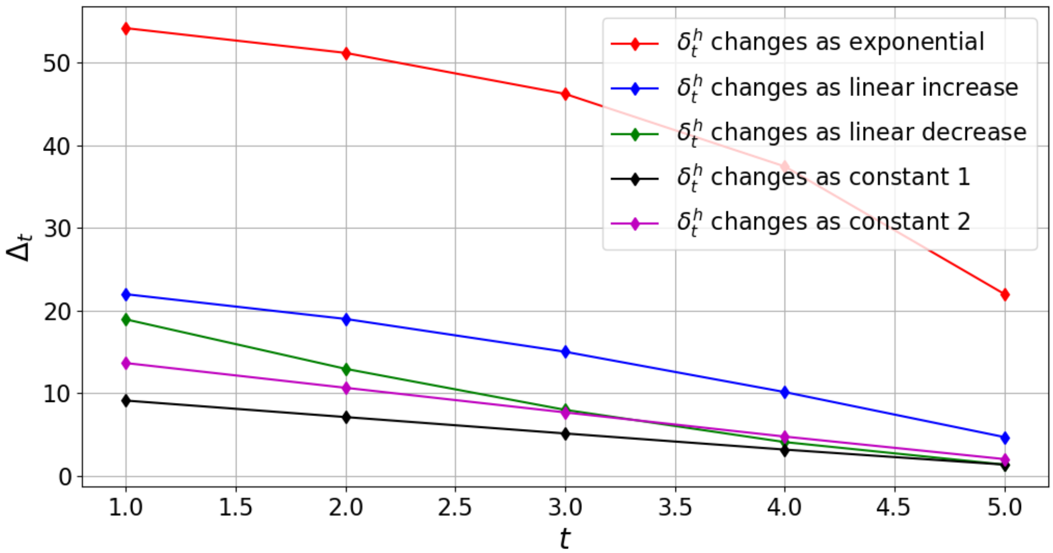


Figure 9. Δ_t given in the min-max cost-to-go in (3) versus time step t for different δ_t^h, for $T = 5, c_l = 1, c_u = 2, \delta_t^l = 1$.

In summary, the optimal solution aims to be more aggressive in earlier time steps and the total min-max is increasing with respect to c_u and time horizon T.

Conclusion

We have studied the Newsvendor problem with the following challenges: (i) the demand is temporally correlated as a Markovian process, (ii) the demand can only be censored (i.e. partially observable), (iii) the distribution of the demand and the transition probabilities of the Markovian process are unknown and only upper and lower bounds on the transitions are given. We modeled this problem as a min-max zero-sum repeated game. We have proved that the worst-case distribution of the adversary at each time is a two-point distribution with non-zero probabilities at the lower and upper bound of the uncertainty set. The optimal action to minimize the worst-case cost-to-go can have be any of the following two formats: (i) a randomized solution with a two-point distribution at the lower and upper bound of the uncertainty set. If the over-utilization cost is larger than the under-utilization cost, higher probability is assigned to the lower bound to behave conservatively. Otherwise, higher probability is assigned to the upper bound to behave more aggressively and increase the chance of full observation. (ii) a deterministic solution at a convex combination of the lower and upper bounds of the uncertainty set, which also balance the over-utilization and under-utilization costs. Finally, we showed that similar results hold for a more general class of cost functions that are uni-model on the difference between the demand and the action.

As a future work, it will be interesting to see how the minimax solution differs for the case in which only partial observation is revealed in both over-utilization and under-utilization. Also, this work can be extended to multi-player games where multiple demands need to be fulfilled from the same inventory. Further, studying the similar problem with unperishable products can be of interest.

References

Arrow, K. J., Harris, T., and Marschak, J., Optimal inventory policy, *Econo-metrica: Journal of the Econometric Society*, pp. 250–272, 1951.

Bradford, J.W., Sugrue, P.K., A bayesian approach to the two-period style-goods inventory problem with single replenishment and heterogeneous Poisson demands, *Journal of Operations Research Society*, vol. 41, pp. 211– 218, 1990.

Zamir, S., Repeated games of incomplete information: Zero-sum, *Handbook of Game Theory with Economic Applications*, vol. 1, pp. 109-154, 1992.

Fisher, M., Hammond, J.H., Obermeyer, W.R., Raman, A., Making supply meet demand in an uncertain world. *Harvard Business Review*, pp. 788–808, 1994.

Khouja, M., The single-period (newsvendor) problem: literature review and suggestions for future research. *Omega*, vol. 27, pp. 537–553, 1999.

Ding, X., Puterman, M. L., Bisi, A., The censored newsvendor and the optimal acquisition of information, *Operations Research*, vol. 50, no. 3, pp. 517–527, 2002.

Sorin, S., A first course on zero-sum repeated games, *Springer Science & Business Media*, vol. 37, 2002.

Bensoussan, A., Cakanyıldırım, M., and Sethi, S. P., On the optimal control of partially observed inventory systems, *Comptes Rendus Mathematique*, vol. 341, no. 7, pp. 419–426, 2005.

Choi, T.M., Li, D., Yan, H., Quick response policy with bayesian information updates, *European Journal of Operational Research*, vol. 170, pp. 788–808, 2006.

Bensoussan, A., Akanyıldırım, M. C¸ Sethi, S. P., A multiperiod newsvendor problem with partially observed demand, *Mathematics of Operations Research*, vol. 32, no. 2, pp. 322–344, 2007.

Negoescu, D., Frazier, P., Powell, W., Optimal learning policies for the newsvendor problem with censored demand and unobservable lost sales, URL: http://people.orie.cornell.edu/pfrazier/pub/learning-newsvendor.pdf, 2008.

Lu, X., Song, J., Zhu, K., Analysis of perishable-inventory systems with censored demand data, *Operations Research*, vol. 56 no. 4, pp. 1034-1038, 2008.

Bensoussan, A., ¸Akanyıldırım, M. C, Minjarez-Sosa, J. A., Royal, A., and Sethi, S. P., Inventory problems with partially observed demands and lost sales, *Journal of Optimization Theory and Applications*, vol. 136, no. 3, pp. 321–340, 2008.

Bensoussan, A., ¸Akanyıldırım, M. C, and Sethi, S. P., Technical note: a note on the censored newsvendor and the optimal acquisition of information, *Operations Research*, vol. 57, no. 3, pp. 791–794, 2009.

Besbes, O., and, Muharremoglu, A., On implications of demand censoring in the newsvendor problem, *Management Science*, pp. 12–7, 2010.

See, C.-T. and Sim, M., Robust approximation to multiperiod inventory management, *Operations Research*, vol. 58, no. 3, pp. 583-594, 2010.

Wang, H.F., Chen, B.C., and Yan, H.M., Optimal inventory decisions in a multiperiod newsvendor problem with partially observed Markovian supply capacities, *European Journal of Operational Research*, vol. 202, pp. 502–517, 2010.

Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., and Seref, M. M., The newsvendor problem: Review and directions for future research, *European Journal of Operational Research*, vol. 213, no. 2, pp. 361–374, 2011.

Bisi, A., Dada, M., Tokdar, S., A censored-data multi-period inventory problem with newsvendor demand distributions, Manufacturing & Service Operations Management, vol. 13, no. 4, pp. 525-533, 2011.

Serel, D.A., Multi-item quick response system with budget constraint, *International Journal of Production Economics*, vol. 137, pp. 235–249, 2012.

Mansourifard, P., Krishnamachari, B., Javidi, T., Bayesian congestion control over a Markovian network bandwidth process, in *Signals, Systems and Computers, 2013 Asilomar Conference on. IEEE*, pp. 332–336, 2013.

Cheaitou, A., van Delft, C., Jemai, Z., Dallery, Y., Optimal policy structure characterization for a two-period dual-sourcing inventory control model with forecast updating, *International Journal of Production Economics*, 157, 238–249, 2014.

Gabrel, V., and Murat, C., and Thiele, A., Recent advances in robust optimization: An overview, *European Journal of Operational Research*, vol. 235, no. 3, pp. 471-483, 2014.

Xin, L., New approaches to inventory control: algorithms, asymptotics and robustness, PhD. *Thesis*, Georgia Institute of Technology, 2015a.

Xin, L., and Goldberg, D. A., Distributionally robust inventory control when demand is a martingale, *arXiv preprint arXiv*:1511.09437, 2015b.

Mansourifard, P., Krishnamachari, B., Javidi, T., Tracking of real-valued Markovian random processes with asymmetric cost and observation, in *American Control Conference (ACC)*. IEEE, 2015.

Solyali, O., and Cordeau, J.-F., and Laporte, G., The impact of modeling on robust inventory management under demand uncertainty, *Management Science*, vol. 62, no. 4, pp. 1188-1201, 2015.

Park, K., Min-Max Stochastic Optimization with Applications to the Single-Period Inventory Control Problem, *Management Science & Financial Engineering*, vol. 21, no. 1, 2015.

Kim, G., Wu, K., & Huang, E., Optimal inventory control in a multi-period newsvendor problem with non-stationary demand, *Advanced Engineering Informatics*, vol. 29, no. 1, pp. 139–145, 2015.

Carrizosa, E., and Olivares-Nadal, A. V. and Ramrez-Cobo, P., Robust newsvendor problem with autoregressive demand, *Computers & Operations Research*, vol. 68, pp. 123-133, 2016.

Tai, A. H, Ching, WK, Recent advances on Markovian models for inventory research, *International Journal of Inventory Research*, vol. 3, no. 3, pp. 198-216, 2016.

Hu, J., Zhang, Ch., Zhu, Ch., (s, S) Inventory Systems with Correlated Demands, *INFORMS Journal on Computing*, vol. 28, no. 4, pp. 603-611, 2016.

Mansourifard, P., Javidi, T., Krishnamachari, B., Percentile Policies for Tracking of Markovian Random Processes with Asymmetric Cost and Observation, *arXiv preprint arXiv*:1703.01261, 2017.

Natarajan, K., and Sim, M., and Uichanco, J., Asymmetry and ambiguity in newsvendor models, *Management Science*, 2017.

Qiu, R., and Sun, M., and Lim, Y. F., Optimizing (s, S) policies for multi-period inventory models with demand distribution uncertainty: Robust dynamic programing approaches, *European Journal of Operational Research*, vol. 261, no. 3, pp. 880-892, 2017.

Mansourifard, F., Mansourifard, P., and Krishnamachari, B., Percentile Policies for Inventory Problems with Partially Observed Markovian Demands, *2nd IEOM European Conference on Industrial Engineering and Operations Management*, Paris, France, 2018.

Biographies

Farzaneh Mansourifard is a master student in industrial engineering at department of engineering and technology at Alzahra university, Tehran, Iran in 2019. She earned B.S. in industrial engineering from Tabriz university, Tabriz, Iran, 2016. She will join Oakland university to pursue her PhD in industrial engineering in Fall 2019. Her research interest is optimization with the applications in healthcare and supply chain.

Parisa Mansourifard received the B.S. and M.S. in electrical engineering from Sharif university of technology, Tehran, Iran, in 2008 and 2010 respectively. She also the M.S. in computer science and Ph.D. in electrical engineering from University of Southern California, Los Angeles, CA, USA, in 2015 and 2017, respectively. During her Ph.D. she held Viterbi Dean fellowships in 2011-2014 and AAUW dissertation completion fellowship in 2015-2016. She worked as a data scientist at Supplyframe Inc. and was a part-time lecturer at University of Southern California. Her research interest is decision-making under uncertainty, machine learning and stochastic optimization.

Bhaskar Krishnamachari received the B.E. degree in electrical engineering at The Cooper Union, New York, NY, USA, in 1998, and the M.S. and Ph.D. degrees from Cornell University, Ithaca, NY, USA, in 1999 and 2002, respectively. He is currently Professor and Ming Hsieh Faculty Fellow in the Department of Electrical Engineering at the Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA. He is also the Director of the USC Viterbi School of Engineering Center on Cyber-Physical Systems and the Internet of Things. His primary research interest is in the design and analysis of algorithms, protocols, and applications for next generation wireless networks.