A Non-Parametric Approach to Stochastic Inventory Planning with Lost Sales and Censored Demand

Woonghee Tim Huh,* Columbia University

Paat Rusmevichientong,† Cornell University

January 6, 2006

Abstract

We study stochastic inventory planning systems with lost sales and censored demand under stationary and non-stationary settings. Contrary to classical inventory theory, we assume that no knowledge of demand is initially available, and lost sales in each period are unobservable. We take a non-parametric approach and propose adaptive inventory policies that generate a sequence of ordering decisions over time. The decision in each period depends only on historical sales data of the past. In the stationary demand setting, any excess inventory in each period is either scrapped (perishable) or carried over to the next period (non-perishable). We also consider non-stationary inventory systems with seasonal demand – allowing for a cyclic pattern of demand distributions – with product updates at the beginning of each season.

To assess the quality of our inventory policies, we use as a benchmark the optimal expected cost that would have incurred if the true distribution were known. Our adaptive algorithms are easy to implement and converge to the optimal solution. Furthermore, for any $T \geq 1$, the average cost during the first T periods under our inventory policies differs from the optimal cost by at most $O\left(1/\sqrt{T}\right)$. Extensive computation shows that our adaptive policies perform well.

*Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, USA. huh@ieor.columbia.edu.

†School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, USA. paatrus@cornell.edu.

1 Introduction

The problem of inventory control and planning has received much interest from practitioners and academics from the early years of operations research. The early literature in this area modeled demand as deterministic and having known quantities, but it soon became apparent that deterministic modeling was inadequate, and uncertainty needed to be incorporated in modeling future demand. As a result, a majority of the papers on inventory theory during the past fifty years employ stochastic demand models. In these models, future demand is given by a specific exogenous random variable, and the inventory decisions are made with full knowledge of the future demand distribution. In many applications, however, the demand distribution is not known *a priori*. Even when past data have been collected, the selection of the most appropriate distribution and its parameters remains ambiguous. In the case that excess demand is lost, the information available to the inventory manager is further limited since she does not observe the realized demand but only the sales quantity (often referred to as censored demand), which is the smaller of the stocking level and the realized demand.

Motivated by these realistic constraints, we develop a non-parametric approach to stochastic inventory planning in the presence of lost sales and censored demand. We consider both stationary and non-stationary demand. When demand is stationary, excess inventory is either scrapped (perishable) or carried over to the next period (non-perishable). We also consider non-stationary seasonal demand, where inventory is scrapped only at the end of each season.

We describe our model in the case of stationary demand. Time periods are indexed forward by $t = 1, \ldots, T$. We develop an adaptive inventory policy $\phi = (y_t \mid t \geq 1)$, where y_t is a decision variable, representing the order-up-to level in period t. We allow y_t to depend *only on the observed historical sales (or censored demand) during the previous* $t-1$ *periods*, with neither assuming any prior knowledge of the demand distribution nor observing any lost sales quantity. We assume that the inventory decision is made at the beginning of each period, and that the replenishment lead-time is instantaneous. For any order-up-to level y, let $C(y)$ denote the expected per-period cost, and let C^* denote the minimum expected cost that we would incur had we known the true demand distribution. To assess the quality of an inventory policy $\phi = (y_t : t \geq 1)$, we use the optimal cost

C^* as a benchmark, and compare it to the average cost over time under ϕ, i.e., for any $T \geq 1$, let

$$\Delta_T(\phi) \equiv E\left[\frac{1}{T}\sum_{t=1}^{T} C(y_t)\right] - C^*.$$

Note that Δ_T is nonnegative by the definition of C^*. In the non-stationary inventory system with seasonal demand, we instead let t correspond to each season. A season consists of a fixed number of, say L, periods. Here, each y_t is a vector of length L, where each component indicates the order-up-to level of the corresponding period in the t'th season.

A major result of this paper is to propose an adaptive inventory policy ϕ^* for stationary systems with both perishable and non-perishable products, and establish that its average expected cost converges to the optimal expected cost at the rate of $1/\sqrt{T}$, i.e., $\Delta_T(\phi^*) = O\left(1/\sqrt{T}\right)$. We establish an analogous result for the non-stationary system with seasonal demand under additional assumptions. To our knowledge, there is no proven upper bound on the rate of convergence prior to our paper for the case of the stationary system with perishable inventory. Furthermore, for the other systems (the stationary system with non-perishable products and the non-stationary system), our algorithm is the first algorithm that is asymptotically optimal.

We briefly outline the ideas behind our algorithm for the stationary setting with perishable products. The key insight is the observation that, given the order-up-to level y_t in period t, we can compute an unbiased estimate of a subgradient of C at y_t using only the sales quantity at time t. This result enables us to leverage the stochastic gradient descent method for minimizing the convex function C, by adjusting the order-up-to level in the next period based on the subgradient of C. In other words, we let

$$y_{t+1} \quad = \quad P_{[0,\bar{y}]}\left(y_t - \epsilon_t H_t(y_t)\right),$$

for any $t \geq 1$, where ϵ_t denotes the step size in period t and $H_t(y_t)$ denotes an unbiased estimate of the subgradient of C at y_t. To ensure that the order-up-to levels remain bounded, we use the projection operator $P_{[0,\bar{y}]}(\cdot)$ onto a bounded interval $[0, \bar{y}]$ where $\bar{y}$ denotes an upper bound on the optimal order-up-to level. By choosing $\epsilon_t = O\left(1/\sqrt{t}\right)$, we prove in Theorem 1 (Section 2.1) that the average expected cost converges to the optimal at the rate of $1/\sqrt{T}$. Through probabilistic rounding, Theorem 3 (Section 2.3) extends the result to handle batch ordering.

In stationary systems with non-perishable products, the target order-up-to levels computed using the above stochastic gradient descent method may not be feasible because left-over inventory is carried over to the next period and the target order-up-to level may be less than the starting on-hand inventory. By choosing an appropriate sequence of step sizes, however, we prove in Theorem 5 (Section 3.1) that, over T periods, the average expected inventory in excess of our target order-up-to levels is at most $O\left(1/\sqrt{T}\right)$. Thus, the average incremental holding cost is at most $O(1/\sqrt{T})$, still giving us the desired convergence result.

In non-stationary systems with seasonal demand, we generalize the stochastic gradient descent method to allow *biased* estimates of subgradients in each iteration. We then relate the quality of our adaptive inventory policy to the average expected magnitude of the bias associated with our subgradient estimators. Under some assumptions on the demand distributions, we show that the average expected bias converges to zero at the rate of $1/\sqrt{T}$, leading to the desired convergence rate. This result is given in Theorem 8 of Section 4.2.

The non-parametric approach taken in this paper contrasts with conventional approaches that exist in the inventory literature. The classical stochastic inventory theory assumes that while the inventory manager does not know the realization of future demand, she has full access to its distribution by the time she makes inventory ordering decisions. The most well-known stochastic inventory problem is the newsvendor problem, whose objective is to minimize the expected overage and underage costs in a single period. The optimal solution for this problem is a fractile of the demand distribution corresponding to a ratio involving per-unit overage and underage costs. In the multi-period setting with stationary demand, the newsvendor-based base-stock policy is optimal provided that the replenishment lead-time is instantaneous. (See, for example, Karlin and Scarf (1958).) With non-stationary demands, Karlin (1960) and Veinott (1965b) have shown the optimality of a base-stock policy. In particular, when demand distributions follow a cyclic pattern or, more generally, an exogenous Markov chain, the optimal policies can be computed using the results of Iglehart and Karlin (1962) and Song and Zipkin (1993). In stochastic inventory problems with instantaneous replenishment lead-times, optimal inventory policies can be computed tractably for any given sequence of demand distributions. In this paper, unlike the classical stochastic inventory literature, we assume that the manager has no prior information regarding future demand distributions, and observes only the sales data.

When the information on the demand distribution is not available, the most common approach in the literature is the use of Bayesian updates. Under this approach, the inventory manager has limited access to demand information; in particular, she knows the family of distributions to which true demand belongs, but she is uncertain about its parameters. Initially, she has a prior belief regarding the uncertainty of the parameter values, and this belief is continually being updated based on historical realized demands. Early papers such as Scarf (1959, 1960), Karlin (1960) and Iglehart (1964) consider cases where the demand distribution belongs to the exponential and range families. Other papers that incorporate the Bayesian approach into stochastic inventory models include Murray and Silver (1966), Chang and Fyffe (1971), and Azoury (1985). Lovejoy (1990) shows that a simple myopic inventory policy based on critical fractile is optimal or near-optimal. In all these above references to Bayesian updates, in contrast to our approach, the manager observes the realized demand, regardless of whether it is higher or lower than the inventory level.

In many applications, excess demand is lost when stock-out occurs, making it impossible for the manager to observe the realized demand; she observes only the sales (or censored demand) information. The contrast between demand and sales quantities was pointed out by Conrad (1976), who shows the effect of censoring in estimating the parameter of the Poisson demand distribution. In this literature with unobservable lost sales, demand is assumed to be stationary, and the replenishment lead-time is instantaneous. Excess inventory is either perishable or non-perishable. In the former case of perishable inventory, the inventory decision in each period is not constrained by the ending inventory level of the previous period. Here, the main result for any general parametric demand model is that the optimal stocking quantity is higher than the myopic solution. The intuition behind this result is that by stocking higher, it is more likely that we can obtain more accurate, uncensored demand information, which is useful for future decisions. This result is due to Harpaz et al. (1982) and Ding et al. (2002). A recent paper by Lu et al. (2004) provides an alternate proof of this result using the first order condition of the optimality equation. In the latter case of non-perishable inventory, however, the inventory level of a period is constrained below by the ending inventory of the previous period. Thus, the impact of overstocking may last longer than a single period, and the above *stock-higher* result no longer holds, i.e., the optimal inventory level may be higher or lower than the myopic solution. Lariviere and Porteus (1999) study this case with a particular distribution called the "newsvendor distribution" (Braden and Freimer (1991)), and

provide sufficient conditions for the stock-higher result to hold. Using a sample-path argument, Lu et al. (2005) prove that, in general, the stock-higher result does not hold. In addition, we mention that Chen and Plambeck (2005) consider the Bayesian learning of product substitution.

When the manager knows the distribution family to which demand belongs, but does not know either its parameters or their priors, Liyanage and Shanthikumar (2005) propose an approach called operational statistics, which integrates the tasks of parameter estimation and expected profit optimization. They consider the stationary models with perishable inventory. Subsequently, Chu et al. (2005) show how to find the optimal mapping from data to the decision variable.

All the current literature on unobservable lost sales and censored demand focus primarily on the Bayesian framework, where the posterior distribution of the demand is updated based on observed sales data. In the Bayesian approach, it is difficult to parsimoniously update the prior distribution as pointed out by Nahmias (1994). Additionally, in many applications, it is unclear which particular prior distribution one should be using.

In this paper, however, we take a non-parametric approach, assuming that the inventory manager knows neither the demand distribution nor the distribution family to which demand belongs. There are several non-parametric approaches available in the literature. One possible approach is the min-max approach. Given that a newsvendor has limited access to the demand distribution (such as mean and standard deviation), we can compute the optimal stocking quantity that will provide the maximum expected profit against the worst possible demand for that stocking quantity. See Scarf (1958), Jagannathan (1977), and Gallego and Moon (1993). Perakis and Roels (2005) present an algorithm for minimizing regrets from not ordering the optimal quantity. Another approach is the bootstrap method, as shown in Bookbinder and Lordahl (1989), to estimate the fractile of demand distribution.

Another non-parametric approach is based on a variant of a stochastic approximation algorithm that approximates the critical fractile of the demand distribution directly based on censored demand samples. Using this approach, Burnetas and Smith (2000) develop an adaptive algorithm for ordering and pricing perishable products. They show that the average profit converges to the optimal, but do not establish the rate of convergence. We note that our approach is fundamentally different from the stochastic approximation algorithm. Our algorithms exploit the convexity of the

cost function and make use of the gradient information in each iteration, enabling us to establish the convergence rate and to extend our result to non-perishable products and non-stationary systems.

Yet another non-parametric approach that utilizes historical data to estimate the newsvendor cost function by recognizing its convexity is the Concave, Adaptive Value Estimation (CAVE) procedure. This algorithm successively approximates the cost function with a sequence of piecewise linear functions. For the stationary inventory problem with perishable goods, Godfrey and Powell (2001) show that the CAVE algorithm has good numerical performance, but does not provide any provable convergence. Powell et al. (2004) extend this line of research, and propose a modified algorithm that produces an asymptotically optimal solution. They assume both perishable inventory and stationary demand (the classical newsvendor setting). In this paper, we propose the first algorithm that is asymptotically optimal with neither the perishability of inventory nor the stationarity of demand.

Furthermore, while there exists no performance guarantee in the unobservable lost sales literature using a non-parametric approach, we prove the square-root convergence rate of the average cost for our adaptive inventory policies. Our performance guarantee is benchmarked against the full knowledge of the demand distribution. As a corollary, this square-root convergence rate is also an upper bound on the value of having full information on the demand distribution. This value of information is the difference between the optimal performance of non-parametric systems with unobservable lost sales (given available information) against the same benchmark. The optimal solutions of these systems are studied in Ding et al. (2002), Lu et al. (2004), and Lu et al. (2005).

Recently, Levi et al. (2005) have studied a multi-period inventory system without any knowledge of the demand distribution. They assume that uncensored samples from the demand distributions are available and compute the sample size required to achieve a certain level of accuracy with high probability. The demand distributions may be non-stationary. In contrast to this paper, they do not explicitly consider the cost associated with sampling from the demand distributions.

When the demand distribution is unknown and only uncensored samples are available, Chang et al. (2005) propose an adaptive sampling algorithm for solving a finite horizon inventory model with lost sales using results from multi-armed bandit problems (see Lai and Robbins (1985) and Auer et al. (2002) for more details). We note that the multi-armed bandit framework is not

appropriate in our setting because we have neither uncensored demand samples nor do we observe any lost sales, making it impossible to compute the overage and underage cost associated with each ordering decision.

The algorithms developed in this paper are based on recent developments in computer science and artificial intelligence. The aim of online convex optimization, as in regular convex optimization, is to minimize a convex function defined over a convex set. However, it is "online" since the optimizer does not know the objective function at the beginning of the algorithm, and at each iteration, he chooses a feasible solution based on the information available to him thus far. He incurs a cost associated with his decision for that period, and obtains some pertinent information regarding the problem. When this information is the gradient of the objective function at the current solution, Zinkevich (2003) has shown that the average T-period cost converges to the optimal cost at the rate of $O(1/\sqrt{T})$. This result was extended by Flaxman et al. (2004) to the case where the optimizer instead obtains an unbiased estimator of the gradient. If the available information is an unbiased estimator of the objective function, then Flaxman et al. (2004) and Kleinberg (2004) have independently proposed asymptotically optimal algorithms with a convergence rate slower than $O(1/\sqrt{T})$.

In this paper, we consider online convex programming with stochastic derivatives. We extend the existing results by removing the assumption of unbiasedness in the estimator of the derivative at the current solution. Such an extension may prove useful in applications beyond the context of this paper. Furthermore, existing results in the literature assume that the decision in each period is independent of all the other periods (see Zinkevich (2003), Flaxman et al. (2004), and Kleinberg (2004)). We extend the result to include the case where the decision of one period is constrained by the decision (or state) of previous periods. (See our models with non-perishable inventory or non-stationary demand.)

The paper is organized as follows. In Section 2, we study the stationary system with perishable inventory, establishing an adaptive inventory policy whose average expected cost over time converges to the optimal at a rate of $O(1/\sqrt{T})$. Next, in Section 3, we extend our result to the case of the non-perishable inventory. Section 4 contains our analysis of the non-stationary inventory systems with

seasonal demand where a product update occurs between consecutive seasons. Section 5 contains experimental results, comparing the performance of our proposed algorithms to other alternatives.

2 Stationary System with Perishable Inventory

In this section, we develop an adaptive inventory policy for the stationary system with perishable inventory. In Section 2.1, we describe the model in detail, and introduce an adaptive inventory management (AIM) algorithm called AIM-Perishable. Section 2.2 contains a proof that the average expected cost under the AIM-Perishable algorithm converges to the optimal expected cost, as well as establishing its rate of convergence. Section 2.3 provides an extension of AIM-Perishable to the case that the ordering quantities are constrained to be integral multiples of a fixed batch size.

2.1 Model and Algorithm

We consider a multi-period inventory system with stationary demand. At the end of each period, any excess demand is lost, and excess inventory is scrapped. Both overage and underage costs are linear. This problem is commonly known as a multi-period newsvendor problem. We assume that while the manager knows that the demand is independent and identically distributed in each period, she has no *a priori* information on its distribution. Over time, she learns about the distribution by observing sales quantity in each period; however, she does not observe the quantity of lost sales.

In each period $t \geq 1$, we assume that the following sequence of events occur.

1. At the beginning of each period, there is no on-hand inventory.
2. The manager makes a replenishment decision to order y_t units, where y_t is any nonnegative real number. (In Section 2.3, we will consider the setting where y_t is restricted to discrete quantities). The ordering cost is linear with respect to the order quantity, where $c \geq 0$ is the per-unit purchase cost. Delivery lead-time is instantaneous.
3. Then, a random demand $D_t \geq 0$ for period t is realized. We assume that the sequence of demand random variables $(D_1, D_2, \ldots)$ are independent and identically distributed with a

common distribution function F. Let d_t denote the realized demand in period t. The manager does not observe d_t, but observes the sales quantity $\min\{d_t, y_t\}$.

4. Any excess inventory is scrapped at the salvage value of $s \in [0, c]$ per unit. Let $h = c - s$ be the per-unit overage cost. While the manager does not observe the quantity of lost sales, we assume that she incurs the goodwill loss of b per unit.

For any $y \geq 0$, let $Q(y)$ denote the expected one-period cost when the inventory level is y, i.e.

$$Q(y) \quad = \quad h \cdot E[y - D]^+ + b \cdot E[D - y]^+ \ ,$$

where the random variable D denotes demand. It is well-known that this single-period cost function $Q(\cdot)$ is convex, and achieves its minimum at the newsvendor quantity given by $y^{NV} = \inf\{y \geq 0 \mid F(y) \geq b/(b+h)\}$. Since the manager does not know the distribution function F in advance, we aim to find a sequence of inventory levels $y_1, y_2, \ldots$ whose average expected cost converges to the minimum cost $Q\left(y^{NV}\right)$. We assume that the manager knows an upper bound $\bar{y}$ on y^{NV}. This condition is satisfied, for example, when the demand distribution has a finite support, known to the manager.

We define an AIM algorithm called AIM-Perishable for the stationary system with perishable inventory. This algorithm defines a sequence of order-up-to levels $y_1, y_2, \ldots$ as follows. Let y_1 be any number in $[0, \bar{y}]$. For any $t \geq 1$, define

$$y_{t+1} \quad = \quad P_{[0,\bar{y}]}\left(y_t - \epsilon_t H_t\left(y_t\right)\right),$$

where $\epsilon_t = \bar{y} / \left\{\max\{b, h\}\sqrt{t}\right\}$ and

$$H_t(y_t) \quad = \quad \begin{cases} h, & \text{if } D_t < y_t; \\ -b, & \text{if } D_t \geq y_t. \end{cases} \tag{1}$$

We note that the random variable $H\left(y_t\right)$ only depends on the sales quantity at time t. The event $D_t \geq y_t$ corresponds to zero ending inventory (i.e. sales equals inventory); the event $D_t < y_t$ corresponds to strictly positive ending inventory. The function $P_{[0,\bar{y}]}\left(\cdot\right)$ denote the projection operator onto the set $[0, \bar{y}]$, mapping any point z to its closest point in the interval $[0, \bar{y}]$, i.e., $P_{[0,\bar{y}]}(z) = \max\{\min\{z, \bar{y}\},\ 0\}$.

One of the main results of this section is the following theorem, which states that the expected running average cost of AIM-Perishable converges to the optimal newsvendor cost. This theorem is proven in Section 2.2.

Theorem 1. *In the stationary system with perishable inventory, AIM-Perishable produces a sequence of order-up-to levels $y_1, y_2, \ldots$ such that for any $T \geq 1$,*

$$E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right] - Q(y^{NV}) \quad \leq \quad \frac{2\,\bar{y}\,\max\{b,h\}}{\sqrt{T}}\ .$$

As mentioned in Section 1, the AIM-Perishable algorithm corresponds to the classical stochastic gradient descent algorithm for minimizing convex functions. It is a well-known result that for any $y \geq 0$, the left derivative of Q at y is given by $h\mathcal{P}\{D_1 < y\} - b\mathcal{P}\{D_1 \geq y\}$. Thus, $H_t(y_t)$ denotes an unbiased estimate of a subgradient of Q at y_t.

We also remark that in the statement of the above theorem, the performance of AIM-Perishable is not compared to the optimal algorithm in the unobservable lost sales setting; rather it is benchmarked against the newsvendor cost, which assumes the full knowledge of the demand distribution. While we do not compute the optimal algorithm for the unobservable lost sales setting, its performance lies between AIM-Perishable and the newsvendor benchmark. Therefore, the above bound is also applicable to the performance gap between the optimal algorithm in this setting and the newsvendor benchmark. It implies that the value of having access to the exact demand distribution in the unobservable lost sales setting diminishes at the rate of $O(1/\sqrt{T})$.

2.2 Online Convex Programming and Proof of Theorem 1

In this section, we provide an extension to recent advances in online convex optimization to prove Theorem 1.

In an online convex optimization problem, the objective function is not known *a priori*, and an iterative selection of a feasible solution yields some pertinent information. When this information is the exact gradient at each step, Zinkevich (2003) has proposed the first asymptotically optimal algorithm, where the expected running average converges to the optimal at the rate of $O(1/\sqrt{t})$. This algorithm is extended to the case of the stochastic gradient by Flaxman et al. (2004). Although

we can establish Theorem 1 from existing results, Lemma 2 below provides a generalization of Flaxman et al. (2004) to the case of biased stochastic gradients. In addition, we also allow for the non-differentiability of the objective function. In Section 4, we will need to use the result of Lemma 2 in its full generality.

Let S be a compact and convex set in $\mathbb{R}^n$. We denote by $diam(S)$ the diameter of S, i.e.

$$diam(S) \quad = \quad \max\{\|u - v\| \mid u, v \in S\},$$

where $\|\cdot\|$ denotes the standard Euclidean norm. Let $P_S : \mathbb{R}^n \to S$ denote the projection operator onto the set S. For any real-valued convex function $\Phi : S \to \mathbb{R}$ defined on S, let $\bigtriangledown\Phi(z)$ denote the set of subgradients of Φ at $z \in S$. The proof of Lemma 2 is contained in Appendix A.

Lemma 2. *Let $\Phi : S \to \mathbb{R}$ be a convex function defined on a compact convex set $S \in \mathbb{R}^n$. For any $z \in S$, let $g(z)$ be any subgradient of Φ at z, i.e., $g(z) \in \bigtriangledown\Phi(z)$. Suppose that there exists $\bar{B}$ such that $\|g(z)\| \leq \bar{B}$ for all $z \in S$. For any $z \in S$, let $H(z)$ be an n-dimensional random vector defined on S, and define $\delta(z) \;=\; E[H(z) \mid z] - g(z)$.*

Let w_1 be any point in S. For any $t \geq 1$, recursively define

$$w_{t+1} \quad = \quad P_S(w_t - \epsilon_t H(w_t)),$$

where $\epsilon_t = \gamma\, diam(S)/\left\{\bar{B}\sqrt{t}\right\}$ for some $\gamma > 0$. Then, for all $T \geq 1$,

$$E\left[\frac{1}{T}\sum_{t=1}^{T}\Phi(w_t)\right] - \Phi(w^*) \quad \leq \quad \left(\gamma + \frac{1}{\gamma}\right)\left(\frac{diam(S)\,\bar{B}}{\sqrt{T}}\right) \;+\; diam(S)\cdot E\left[\frac{1}{T}\sum_{t=1}^{T}\|\delta(w_t)\|\right],$$

where $w^* = \arg\min_{w\in S}\Phi(w)$.

In Lemma 2, the variable $\delta(w_t)$ denotes the bias associated with our estimate of the subgradient of Φ at w_t. Thus, the above result shows that the difference between the average expected cost over T periods and the minimum cost depends on a term that converges to zero at the rate of $1/\sqrt{T}$ and the average expected magnitude of the bias of the subgradient estimators over T periods.

Now we prove Theorem 1. In the stationary system with perishable inventory, the expected single period cost Q has a left-derivative given by

$$h \cdot \mathcal{P}\{D_1 < y\} - b \cdot \mathcal{P}\{D_1 \geq y\} \quad = \quad -b + (b+h)\mathcal{P}\{D < y\} \;.$$

From the definition of $H_t(y_t)$ (Equation 1), the above expression is the expected value of $H_t(y_t)$ conditioned on the value of y_t. Moreover, it is easy to verify that $|H(\cdot)| \leq \max\{b, h\}$. Theorem 1 follows from Lemma 2 with $S = [0, \bar{y}]$, $\bar{B} = \max\{b, h\}$, $\gamma = 1$ and $\delta(z) = 0$ for all $z \in S$.

We remark that for the application of Lemma 2 in this section, stochastic gradients are unbiased, i.e., $\delta(w_t) = 0$. Lemma 2 is stated in its full generality as biased gradients are used in Section 4.

2.3 Extension: Batch Ordering

In Section 2.1, we assume that the order quantity in each period is any nonnegative real number, and show that the AIM-Perishable algorithm produces a sequence of inventory levels whose expected running average cost converges to the optimal cost. The distribution is either continuous or discrete. However, in practice, the set of possible ordering quantities may be constrained to a discrete set. If this set is the integer multiples of a fixed quantity, the problem is known as *batch ordering* in the literature (See Veinott (1965a), Chen (2000), and Gallego and Toktay (2004)). In this section, we consider the stationary system with perishable inventory under batch ordering. Without loss of generality, we assume the ordering quantity in each period is restricted to be nonnegative integers.

To address the integrality constraint, we introduce the following variant of the AIM-Perishable algorithm, which we will call AIM-Batch. The AIM-Batch algorithm maintains an auxiliary sequence $(z_t \in \mathbb{R} : t \geq 1)$ and a sequence of *integer* stocking levels $(y_t \in \mathbb{Z}_+ \cup \{0\} : t \geq 1)$. We set $z_1 = y_1$ to be any integer in $[0, \bar{y}]$, where $\bar{y}$ is an upper bound on the newsvendor quantity y^{NV}. For any $t \geq 1$, the auxiliary sequence is defined by

$$z_{t+1} \quad = \quad P_{[0,\bar{y}]}\left(z_t - \epsilon_t H_t\left(y_t\right)\right) \ ,$$

where $\epsilon_t = \bar{y}/\left(\max\{b, h\}\sqrt{t}\right)$ as defined before and $H_t\left(y_t\right)$ is defined as in Equation (1), i.e. $H_t\left(y_t\right) = -b + (b + h) \cdot \mathbf{1}\left[D_t < y_t\right]$. We obtain y_{t+1} from z_{t+1} by probabilistic rounding, i.e.,

$$y_{t+1} \quad = \quad \begin{cases} \lceil z_{t+1} \rceil, & \text{with probability } z_{t+1} - \lfloor z_{t+1} \rfloor; \\ \lfloor z_{t+1} \rfloor, & \text{with probability } 1 - (z_{t+1} - \lfloor z_{t+1} \rfloor). \end{cases}$$

Although we maintain the auxiliary sequence, we implement the integral stocking level y_t, incurring the expected cost of $Q\left(y_t\right)$. We then compute the estimate $H_t(y_t)$ of the subgradient of Q at y_t and use this information to update the value of z_{t+1}.

The following theorem states the performance guarantee of AIM-Batch.

Theorem 3. *Suppose that the set of points where the demand distribution has a positive mass is finite. In the stationary system with perishable inventory under batch ordering, AIM-Batch produces a sequence of integer stocking levels $y_1, y_2, \ldots$ such that for any $T \geq 1$,*

$$E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right] - Q(y^{NV}) \;\; \leq \;\; \frac{4\ \bar{y}\ \max\{b,h\}}{\sqrt{T}} + 2\ \max\{b,h\}\ .$$

While the above theorem does not show that AIM-Batch is asymptotically optimal, the asymptotical error term is at most twice the overage or underage cost. This asymptotical error exists because we do not have access to the function values. Even in the case of the newsvendor problem with full information on the demand distribution, if ordering quantities are discrete while demand is continuous, the fractional newsvendor solution is rounded up or down by comparing costs evaluated at these integer quantities; derivative-only approaches (without evaluating costs) do not specify how it should be rounded. We remark that the above bound is on the difference between the performance of the AIM-Batch algorithm and the optimal fractional newsvendor quantity y^{NV}, without being restricted to integer values.

The proof of Theorem 3 follows immediately from the following lemma. The proof of Lemma 4 appears in Appendix B.

Lemma 4. *Let $\Phi : [0, \bar{S}] \to \mathbb{R}$ be a convex function defined on a closed interval $[0, \bar{S}]$ in $\mathbb{R}$ such that Φ is continuously differentiable except at finitely many points. Suppose there exists $\bar{B}$ such that for any subgradient $\phi(z)$ of Φ at $z \in [0, \bar{S}]$, $\|\phi(z)\| \leq \bar{B}$ holds. Let $H(z)$ be a random variable such that $E[H(z) \mid z] \in \bigtriangledown\Phi(z)$ holds for any integer $z \in [0, \bar{S}]$.*

Let $\hat{w}_1$ be any integer in $\{0, 1, \ldots, \bar{S}\}$. For any $t \geq 1$, recursively define

$$\hat{w}_{t+1} \;\; = \;\; P_{[0,\bar{S}]}\left(\hat{w}_t - \epsilon_t H(\bar{w}_t)\right)\ ,$$

where $\epsilon_t = \bar{S}/\left\{\bar{B}\sqrt{t}\right\}$ and $\bar{w}_t$ is obtained by probabilistically rounding $\hat{w}_t$, i.e.

$$\bar{w}_t \;\; = \;\; \begin{cases} \lceil \hat{w}_t \rceil, & \textit{with probability } \hat{w}_t - \lfloor \hat{w}_t \rfloor; \\ \lfloor \hat{w}_t \rfloor, & \textit{with probability } 1 - (\hat{w}_t - \lfloor \hat{w}_t \rfloor). \end{cases}$$

Then, for all $T \geq 1$,

$$E\left[\frac{1}{T}\sum_{t=1}^{T}\Phi(\bar{w}_t)\right] - \min_{w\in[0,\bar{S}]}\Phi(w) \quad \leq \quad \frac{4\ \bar{S}\ \bar{B}}{\sqrt{T}} + 2\bar{B}\ .$$

3 Stationary System with Non-Perishable Inventory

We continue to study a multi-period inventory system with stationary demand and lost sales. While any excess inventory at the end of each period is scrapped in Section 2, it is carried over to the next period in this section. Inventories in this section are non-perishable (or durable). In Section 3.1, we describe the model and present an appropriate adaptive inventory policy. We also state the main convergence result. The proof of this result, which uses properties of a stochastic storage process, appears in Section 3.2.

3.1 Model and Algorithm

We describe the stationary system with non-perishable products in further detail as follows.

1. At the beginning of each period t, the manager observes the initial on-hand inventory level $x_t \geq 0$. Without loss of generality, we assume that $x_1 = 0$.

2. She makes a replenishment decision to order $u_t \in \mathbb{R}^+$ units, incurring the ordering cost of $c \cdot u_t$. We assume instantaneous replenishment. Let $y_t = x_t + u_t$ denote the inventory level after the replenishment decision.

3. The demand d_t is then realized from the distribution D_t. The manager does not know this distribution, but knows that $D_1, D_2, \ldots$ are independent and identically distributed. While she does not observe d_t, she observes the sales quantity $\min\{d_t, y_t\}$ instead. For the demand distribution D_t, we assume that there exists a bound $\bar{D}$ such that for all t, $D_t \leq \bar{D}$ with probability one. Without loss of generality, we assume that $E[D_t] > 1$ for all t. (In Section 3.3, we extend our results to the case when $E[D_t] \leq 1$.)

4. The overage and underage cost associated with this period is $h \cdot [y_t - d_t]^+ + b \cdot [d_t - y_t]^+$. The inventory at the beginning of the next period is given by $x_{t+1} = [y_t - d_t]^+$.

The planning horizon is either finite or infinite. By making an appropriate salvage value assumption at the end of the planning horizon, we suppose that the purchase cost is zero, i.e., $c = 0$; interested readers are referred to Veinott and Wagner (1965) and Janakiraman and Muckstadt (2004) for details. As in Section 2, we denote by $Q(y_t)$ the expected single-period cost as a function of the order-up-to level y_t.

In the classical inventory model where the manager knows the demand distribution, the stationarity of demand implies that a myopic solution is optimal. Thus, the stationary multi-period inventory model is analytically equivalent to the newsvendor model, and ordering up to y^{NV} in each period is also optimal for this problem. Under this optimal policy, the constraint $y_{t+1} \geq [y_t - d_t]^+$ never becomes binding. However, when the demand distribution is unknown, the manager makes a decision based on the collection of observed sales quantities, and as a result, the order-up-to levels may change. Thus, her decision in a period may be *tightly* constrained by the carry-over inventory from the previous period.

We propose a version of the AIM algorithm called AIM-Durable for the non-perishable case. This algorithm is similar to AIM-Perishable; it mimics AIM-Perishable unless the starting inventory level already exceeds the target inventory level. AIM-Durable maintains a pair of sequences $(\hat{y}_t : t \geq 1)$ and $(y_t : t \geq 1)$. The auxiliary sequence $(\hat{y}_t : t \geq 1)$ represents the target inventory levels while the second sequence $(y_t : t \geq 1)$ represents the *actual implemented* inventory levels after ordering. These sequences are recursively defined as follows. Set $y_1 = \hat{y}_1$ to any value in $[0, \bar{y}]$, where $\bar{y}$ denotes an upper bound on the newsvendor quantity y^{NV} as before. For each $t \geq 1$, set

$$\begin{aligned} \hat{y}_{t+1} &= P_{[0,\bar{y}]}\left(\hat{y}_t - \epsilon_t H_t(\hat{y}_t)\right) \ , \\ y_{t+1} &= \max\left\{\hat{y}_{t+1},\ x_{t+1}\right\} \ , \end{aligned}$$

where $H_t(\hat{y}_t) = h \cdot \mathbf{1}[D_t < \hat{y}_t] - b \cdot \mathbf{1}[D_t \geq \hat{y}_t] = -b + (b+h) \cdot \mathbf{1}[D_t < \hat{y}_t]$, and for any $t \geq 1$, the step size ϵ_t is given by

$$\epsilon_t = \frac{1}{h\sqrt{t}},$$

where h is the per-unit overage cost.

We note that while the implemented inventory level is y_t (not $\hat{y}_t$), the event $D_t < \hat{y}_t$ is observable. Since $\hat{y}_t \leq y_t$, the event $D_t < \hat{y}_t$ occurs exactly when the ending inventory level exceeds $y_t - \hat{y}_t$. Thus, we can compute $H_t(\hat{y}_t)$ based on the observed sale quantity in period t.

The main result of this section is the following theorem, which shows that the expected running average of AIM-Durable converges to the optimal at the rate of $O(1/\sqrt{T})$. The proof of Theorem 5 appears in Section 3.2.

Theorem 5. *In the stationary system with non-perishable inventory, the AIM-Durable algorithm produces the sequence of after-ordering inventory levels $y_1, y_2, \ldots$ such that for any $T \geq 1$,*

$$E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right] - Q(y^{NV}) \quad \leq \quad \left(\frac{\max\{b,h\}^2}{h} + h\bar{y}^2 + \frac{4\alpha h}{(1-\alpha)^2}\right)\frac{1}{\sqrt{T}}\ ,$$

holds with $\alpha = \exp\left\{-2\cdot(E[D_1]-1)^2/\bar{D}\right\}$.

We note that on the right hand side of Theorem 5, the first two terms

$$\left(\frac{\max\{b,h\}^2}{h} + h\bar{y}^2\right)\frac{1}{\sqrt{T}} \quad = \quad \left(\frac{\max\{b,h\}}{h\bar{y}} + \frac{h\bar{y}}{\max\{b,h\}}\right)\frac{\bar{y}\max\{b,h\}}{\sqrt{T}},$$

are attributed to the application of Lemma 2 to the auxiliary sequence $(\hat{y}_t : t \geq 1)$ with $\gamma = \max\{b,h\}/\{h\bar{y}\}$. This expression represents the difference between the optimal cost $Q\left(y^{NV}\right)$ and the average cost under the auxiliary sequence if the inventory is scrapped in each period. The last term on the right hand side of Theorem 5 represents an upper bound on the T-period average of the expected excess inventory above the target inventory level $\hat{y}_t$. Our choice of the step size ϵ_t enables us to establish that the average excess inventory also decreases to zero at the rate of $1/\sqrt{T}$.

3.2 A Stochastic Storage Process and Proof of Theorem 5

To establish an upper bound on the average expected excess inventory above our target levels, we need the following result on a stochastic storage process.

Lemma 6. *Let $D_1, D_2, \ldots$ be independent and identically distributed nonnegative random variables such that for all t, $D_t \leq \bar{D}$ with probability one and $E[D_t] > 1$. Consider a sequence of random variables $(Z_t \mid t \geq 1)$ defined by*

$$Z_{t+1} \quad = \quad \left[Z_t + \frac{1}{\sqrt{t}} - D_t\right]^+ \ ,$$

where $Z_0 = 0$. *Then, for any* $T \geq 1$,

$$E\left[\frac{1}{T}\sum_{t=1}^{T} Z_t\right] \leq \frac{4\alpha}{(1-\alpha)^2}\cdot\frac{1}{\sqrt{T}}$$

where $\alpha = \exp\left\{-2\cdot\left(E\left[D_1\right]-1\right)^2/\bar{D}\right\}$.

Proof. The proof of Lemma 6 combines several non-trivial ideas. We briefly outline its sketch here, and refer the reader to Appendix C for the complete proof. The difficulty of working with the Z_t process is its non-stationarity. Thus, we introduce an auxiliary process W_t by $W_{t+1} = [W_t + 1 - D_t]^+$, which stochastically dominates the original process. This auxiliary process is stationary, and we find, using the well-known Hoeffding inequality, an upper bound on the expected length of the "busy" interval (consecutive periods in which the auxiliary process is strictly positive). We use this bound to derive an upper bound for the expected average value of another stochastic process that also stochastically dominates the original Z_t process. □

We remark that in the proof of Lemma 6, the auxiliary process W_t is the waiting time of the t^{th} customer in the $GI/D/1$ queuing system, where the inter-arrival time between the t^{th} and $t+1^{th}$ customers is distributed as D_t, and the service time is deterministically 1 . The proof establishes an upper bound on the expected length of a busy period.

Now we proceed to the proof of Theorem 5. The main idea in this proof is that the auxiliary sequence $(\hat{y}_t : t \geq 1)$ in AIM-Durable corresponds to the decisions of AIM-Perishable in Section 2. Thus, we know that the running average error (in cost) with respect to $\hat{y}_t$ diminishes at the rate of $O(1/\sqrt{T})$. However, instead of $\hat{y}_t$, we want to prove the same result with respect to y_t. To achieve this, we show the running average of the difference $Q(y_t) - Q(\hat{y}_t)$ also diminishes at the rate of $O(1/\sqrt{T})$ using a stochastic storage process defined in Lemma 6.

We apply Lemma 2 to the auxiliary sequence $(\hat{y}_t : t \geq 1)$. Using $\gamma = (\max\{b,h\})/(h\bar{y})$, $S = [0,\bar{y}]$ and $B = \max\{b,h\}$, we obtain

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T} Q(\hat{y}_t)\right] - Q(y^{NV}) &\leq \left(\frac{\max\{b,h\}}{h\bar{y}} + \frac{h\bar{y}}{\max\{b,h\}}\right)\frac{\bar{y}\max\{b,h\}}{\sqrt{T}} \\ &= \left(\frac{\max\{b,h\}^2}{h} + h\bar{y}^2\right)\frac{1}{\sqrt{T}}\,. \end{aligned}$$

Thus, it remains to prove

$$E\left[\frac{1}{T}\sum_{t=1}^{T}\left(Q(y_t)-Q(\hat{y}_t)\right)\right] \;\leq\; \frac{4\alpha h}{(1-\alpha)^2\sqrt{T}} \;\;.$$

Since $y_t = \max\{\hat{y}_t, x_t\}$ for each t, it follows

$$\begin{aligned} Q(y_t)-Q(\hat{y}_t) &= h\cdot E[y_t-D_t]^+ + b\cdot E[D_t-y_t]^+ - h\cdot E[\hat{y}_t-D_t]^+ - b\cdot E[D_t-\hat{y}_t]^+ \\ &= h\cdot E[y_t-\max\{\hat{y}_t,D_t\}]^+ - b\cdot E[\min\{y,D_t\}-\hat{y}_t]^+ \\ &\leq h\cdot(y_t-\hat{y}_t)\;. \end{aligned}$$

This difference $y_t - \hat{y}_t$ is always nonnegative. We claim that it satisfies the following recursive relation: for any $t \geq 1$,

$$(y_{t+1}-\hat{y}_{t+1}) \;\leq\; [(y_t-\hat{y}_t)+h\epsilon_t-d_t]^+\;,$$

where d_t denotes the realized demand in period t. If $x_{t+1} \leq \hat{y}_{t+1}$, then by the definition of AIM-Durable, we have $y_{t+1} - \hat{y}_{t+1} = 0$, and the above claim holds. Otherwise, we have $x_{t+1} > \hat{y}_{t+1}$, in which case $y_{t+1} = x_{t+1} = y_t - d_t$ holds. Since the AIM-Durable algorithm starting at $x_1 = 0$ does not let any target inventory level $\hat{y}_t$ exceed $\bar{y}$, we have

$$y_{t+1}-\hat{y}_{t+1} \;=\; y_{t+1}-P_{[0,\bar{y}]}\left(\hat{y}_t-\epsilon_t H_t(y_t)\right) \;\leq\; y_{t+1}-(\hat{y}_t-\epsilon_t H_t(y_t))\,.$$

From $y_{t+1} = y_t - d_t$, it follows

$$y_{t+1}-\hat{y}_{t+1} \;\leq\; y_t-\hat{y}_t+\epsilon_t H_t(y_t)-d_t \;\leq\; y_t-\hat{y}_t+h\epsilon_t-d_t\;,$$

where the last inequality follows from the fact that $H_t(y_t) \leq h$.

Consider the stochastic process $(Z_t \mid t \geq 1)$ defined by

$$Z_{t+1} \;=\; [Z_t+h\epsilon_t-D_t]^+ \;=\; \left[Z_t+\frac{1}{\sqrt{t}}-D_t\right]^+$$

for each $t \geq 0$ and $Z_0 = 0$. (The last equality above follows from the definition of ϵ_t.) Clearly, Z_t is an upper bound on $y_t - \hat{y}_t$ for each sample path. It follows that for any $T \geq 1$,

$$E\left[\frac{1}{T}\sum_{t=1}^{T}Q(y_t)-Q(\hat{y}_t)\right] \;\leq\; h\cdot E\left[\frac{1}{T}\sum_{t=1}^{T}(y_t-\hat{y}_t)\right] \;\leq\; h\cdot E\left[\frac{1}{T}\sum_{t=1}^{T}Z_t\right] \;\leq\; \frac{4\alpha h}{(1-\alpha)^2\sqrt{T}}$$

where the last inequality follows from Lemma 6. This completes the proof of Theorem 5.

3.3 Extensions

We can extend Theorem 5 to the case when $E[D_t] \leq 1$. In this case, let $\kappa > 0$ denote a lower bound on the expected demand, i.e. $E[D_t] > \kappa$ for all t. Note that if no such κ exists, then the demand is zero almost surely and the problem is trivial. By changing the step size to $\epsilon_t = \kappa/\{h\sqrt{t}\}$ for all t, we can show that

$$E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right] - Q(y^{NV}) \quad \leq \quad \left(\frac{\kappa \max\{b,h\}^2}{h} + \frac{h\bar{y}^2}{\kappa} + \frac{4\alpha\kappa h}{(1-\alpha)^2}\right)\frac{1}{\sqrt{T}},$$

where $\alpha = \exp\left\{-2\left(E[D_1] - \kappa\right)^2/\bar{D}\right\}$. The proof of this result is similar to the one given in Section 3.2.

4 Seasonal Demand with Product Updates

In this section, we consider a non-stationary inventory system where demand distribution follows a cyclic pattern, in contrast to the stationary demand in Sections 2 and 3. Each season (or cycle) consists of a fixed number of periods. We assume that excess inventory is carried over from one period to the next, except at the end of a season. This assumption is applicable if the product is updated at the beginning of each season, or the holding cost from the end of one season to the beginning of the next is prohibitively expensive. In Section 4.1, we describe our model in detail. In Section 4.2, we propose an AIM algorithm for this setting, and state its convergence result. Section 4.3 contains the proof of this result.

4.1 Model

In the non-stationary inventory system with seasonal demand, we assume the cyclic demand of length L, where L is the number of periods in each selling season. We index each season by $s = 1, 2, \ldots$, and each period within a season by $r = 1, 2, \ldots, L$. The manager does not know the exact demand distribution, but knows that demands are independent and cyclic, i.e., for each r, the sequence of demands in the r^{th} period of each season $D_{1,r}, D_{2,r}, \ldots$ are independent and identically distributed. We denote this common distribution by D_r. We assume that for any $1 \leq r \leq L$,

the random variable D_r has a continuous density function with bounded support and the density function is bounded above by M. Thus, there exists an upper bound $\bar{D}$ such that $\sum_{r=1}^{L} D_r \leq \bar{D}$ with probability one.

The following sequence of events take place.

1. At the beginning of each period (s,r), the manager observes the initial inventory level $x_{s,r}$. We assume $x_{s,r} = 0$ if $r = 1$.

2. The manager orders sufficient inventory to raise it up to $y_{s,r}$ units, where $y_{s,r} \geq x_{s,r}$.

3. Then, demand $d_{s,r}$ is realized.

4. An appropriate overage and underage cost is realized depending on the quantity of excess demand or excess inventory. Its expected cost, as a function of the inventory level after ordering, is given by

$$Q_r(y) \quad = \quad h \cdot E[y - D_r]^+ + b \cdot E[D_r - y]^+ \ .$$

   We refer to this cost as the *myopic* cost. As before, the manager does not observe the realized demand $d_{s,r}$, but sales quantity $\min\{d_{s,r}, y_{s,r}\}$. Excess demand is lost. Thus, the beginning inventory level of the next period satisfies $x_{s,r+1} = [y_{s,r} - d_{s,r}]^+$ for $r = 1, 2, \ldots, L-1$, except that the first period of each season has no carry-over inventory, i.e., $x_{s,1} = 0$ for each season s.

In the benchmark system where the manager knows the true demand distributions $D_1, \ldots, D_L$ (as in the classical stochastic inventory theory), it is well-known that an optimal policy belongs to the class of order-up-to policies. In these policies, for given order-up-to levels $R_1, \ldots, R_L$, , the after-ordering inventory level in each period (s,r) is set to $y_{s,r} = \max\{x_{s,r}, R_r\}$. Let $C(R_1, \ldots, R_L)$ denote the total expected cost in a season as a function of the order-up-to levels $R_1, R_2, \ldots, R_L$. The expression for $C(R_1, \ldots, R_L)$ is given by the following dynamic program. For any r, define

$$U_r(R_r \mid R_{r+1}, \ldots, R_L) \quad = \quad Q_r(R_r) + E_{D_r}\left[U_{r+1}(\max\{R_r - D_r, R_{r+1}\} \mid R_{r+2}, \ldots, R_L)\right] \ ,$$

where $U_{L+1}(\cdot) = 0$. We also define the backward incremental cost in period r by

$$C_r(R_r \mid R_{r+1}, \ldots, R_L) \quad = \quad U_r(R_r \mid R_{r+1}, \ldots, R_L) - U_{r+1}(R_{r+1} \mid R_{r+2}, \ldots, R_L) \ .$$

Clearly, $C(R_1,\ldots,R_L)=\sum_{r=1}^{L} C_r(R_r \mid R_{r+1},\ldots,R_L)$. Denote the minimizer of $C(\cdot)$ by $(R_1^*,\ldots,R_L^*)$, which are the optimal order-up-to levels. The following lemma follows from the definition of the above dynamic program and the convexity of Q_r's.

Lemma 7. *In the non-stationary inventory system with seasonal demand and product updates, the optimal order-up-to levels* $(R_1^*,\ldots,R_L^*)$ *satisfy* $R_r^* = \arg\min_{R\geq 0} C_r(R \mid R_{r+1}^*,\ldots,R_L^*)$, *and each* $C_r\left(\cdot\middle| R_{r+1}^*,\ldots,R_L^*\right)$ *is convex.*

4.2 Algorithm

In this section, we develop an adaptive algorithm for the non-stationary system with seasonal demand. This algorithm, called AIM-Seasonal, is an order-up-to policy, where the target inventory levels of each season are determined based on observed sales quantities in the past. For each season $s\geq 1$, let $(\hat{y}_{s,1},\hat{y}_{s,2},\ldots,\hat{y}_{s,L})$ denote the target inventory levels. We note that all target inventory levels for a season are generated simultaneously at the beginning of that season.

As in AIM-Durable of Section 3.1, the AIM-Seasonal algorithm also maintains a pair of sequences $\{(\hat{y}_{s,1},\ldots,\hat{y}_{s,L}) \mid s\geq 1\}$ and $\{(y_{s,1},\ldots,y_{s,L}) \mid s\geq 1\}$. The first sequence is an auxiliary sequence representing the target inventory levels, and the second sequence represents the *actual implemented* inventory levels. Initially, we set $(\hat{y}_{1,1},\ldots,\hat{y}_{1,L})$ to any vector in $[0,\bar{D}]^L$, where $\bar{D}$ is an upper bound on $\sum_{r=1}^{L} D_r$. For each s and r, define

$$\hat{y}_{s+1,r} \;=\; P_{[0,\bar{D}]}\left(\hat{y}_{s,r}-\epsilon_s H_{s,r}\right),$$

where ϵ_s and $H_{s,r}$ are defined below. Also, let $y_{s+1,r}=\max\{\hat{y}_{s+1,r},\ x_{s+1,r}\}$.

We now provide the definitions of ϵ_s and $H_{s,r}$. Let

$$\epsilon_s \;=\; \frac{\bar{D}}{L\ \max\{b,h\}}\cdot\frac{1}{\sqrt{s}}\ .$$

We denote by $D_{[(s,r_1),(s,r_2)]}=\sum_{i=r_1}^{r_2} D_{(s,i)}$ the cumulative demand from period r_1 to r_2 in season s. Let

$$\begin{aligned} H_{s,r} \;=\;& \left\{-b+(b+h)\cdot\mathbf{1}\left[D_{(s,r)}<\hat{y}_{s,r}\right]\right\} \\ &+\sum_{r'=r+1}^{L}\left\{-b+(b+h)\cdot\mathbf{1}[D_{[(s,r),(s,r')]}<\hat{y}_{s,r}]\right\}\cdot\prod_{i=r+1}^{r'}\mathbf{1}\left[\hat{y}_{s,r}-D_{[(s,r),(s,i-1)]}>\hat{y}_{s,i}\right]\ . \end{aligned}$$

For the simplicity of explanation, suppose that the after-ordering inventory level in period r is $\hat{y}_{s,r}$, i.e., $y_{s,r} = \hat{y}_{s,r}$. In the above expression, the first term represents the derivative of the myopic cost in period r. For $r' \geq r+1$, the second indicator function represents whether the beginning inventory level is higher than the target inventory level in period r', i.e., $y_{s,r'} > \hat{y}_{s,r'}$. This event corresponds to the event that the order quantity in each of periods $r+1, \ldots, r'$ is zero. When this event happens, the summand is either h or b, corresponding to whether any positive quantity of inventory remains at the end of period r' or not, respectively. It is straightforward to check that each $H_{s,r}$ can be computed at the end of season s prior to the beginning of the next season $s+1$ using only the sales data from season s.

An important property of $H_{s,r}$ is that for any $\hat{y}_{r+1}, \hat{y}_{r+2}, \ldots, \hat{y}_L$, the expected value of $H_{s,r}$ is a subgradient of $C_r(\hat{y}_r \mid \hat{y}_{r+1}, \ldots, \hat{y}_L)$ at $\hat{y}_r$, i.e.,

$$E\left[H_{s,r} \mid \hat{y}_{s,r}, \ldots, \hat{y}_{s,L}\right] \;=\; C_r'\left(\hat{y}_{s,r} \mid \hat{y}_{s,r+1}, \ldots, \hat{y}_{s,L}\right) \;.$$

The proof of this property follows from the definition of C_r. When Q_r's are differentiable, for any order-up-to levels $\hat{y}_r, \hat{y}_{r+1}, \ldots, \hat{y}_L$, we have

$$C_r'(\hat{y}_r \mid \hat{y}_{r+1}, \ldots, \hat{y}_L) \;=\; Q_r'(\hat{y}_r) + \sum_{r'=r+1}^{L} E\left[Q_{r'}'(\hat{y}_r - D_{[r,r'-1]}) \cdot \prod_{i=r+1}^{r'} \mathbf{1}[\hat{y}_r - D_{[r,i-1]} > \hat{y}_i]\right] \;,$$

where $D_{[r_1,r_2]} = \sum_{i=r_1}^{r_2} D_i$ is the distribution of the cumulative demand in periods $\{r_1, \ldots, r_2\}$. (See, for example, Levi et al. (2005) for details). We note that for $r' \geq r+1$,

$$Q_{r'}'\left(\hat{y}_r - D_{[r,r'-1]}\right) = -b + (b+h)\mathcal{P}\left\{D_{r'} < \hat{y}_r - D_{[r,r'-1]}\right\} = -b + (b+h)\mathcal{P}\left\{D_{[r,r']} < \hat{y}_r\right\},$$

and it is easy to show that $H_{s,r}$ is an unbiased estimator of $C_r'\left(\hat{y}_{s,r} \mid \hat{y}_{s,r+1}, \ldots, \hat{y}_{s,L}\right)$.

We recall from Lemma 7 that R_r^* is the minimizer of $C_r(\cdot \mid R_{r+1}^*, \ldots, R_L^*)$. We caution that $H_{s,r}$ is the unbiased estimator of $C_r'(\cdot \mid \hat{y}_{r+1}, \ldots, \hat{y}_L)$, and not of $C_r'(\cdot \mid R_{r+1}^*, \ldots, R_L^*)$. Despite this fact, the following theorem states that AIM-Seasonal converges to the optimal solution. The proof of this result appears in Section 4.3.

Theorem 8. *In the inventory system with seasonal demand and product updates, the AIM-Seasonal algorithm produces a sequence of target inventory levels* $\{(\hat{y}_{s,1}, \ldots, \hat{y}_{s,L}) : s \geq 1\}$ *such that for any* $T \geq 1$,

$$E\left[\frac{1}{T}\sum_{s=1}^{T} C\left(\hat{y}_{s,1}, \ldots, \hat{y}_{s,L}\right)\right] - C\left(R_1^*, \ldots, R_L^*\right) \;\leq\; \frac{2L^2 \max\{b,h\}\left(1+\bar{D}M\right)^L}{M} \cdot \frac{1}{\sqrt{T}} \;.$$

4.3 Biased Derivatives and Proof of Theorem 8

In this section, we present the proof of Theorem 8. A key idea in this proof is to use our earlier result (Lemma 2) on the performance of adaptive algorithms with biased stochastic derivatives. For the simplicity of exposition, we assume that all demand distributions have continuous probability density functions, which guarantees the differentiability of all involved functions.

The proof of Theorem 8 also makes use of the following result that establishes an upper bound on the difference of C_r for any $1 \leq r \leq L$. The proof of the following lemma appears in Appendix D.

Lemma 9. *Let $R_1^*, \ldots, R_L^*$ denote the optimal order-up-to levels. For any $r = 1, \ldots, L$, and order-up-to levels $R_r, \ldots, R_L$, we have*

$$\begin{aligned} & C_r\left(R_r \middle| R_{r+1}, \ldots, R_L\right) - C_r\left(R_r \middle| R_{r+1}^*, \ldots, R_L^*\right) \\ \leq & \sum_{\ell=r+1}^{L} C_\ell\left(R_\ell \middle| R_{\ell+1}^*, \ldots, R_L^*\right) - C_\ell\left(R_\ell^* \middle| R_{\ell+1}^*, \ldots, R_L^*\right) \ . \end{aligned}$$

Let f_{D_r} be the continuous probability density function of demand distribution D_r in period r. We recall from Section 4.1 that each f_{D_r} is uniformly bounded, i.e., there exists $M > 0$ such that $f_{D_r}(x) \leq M$ for any $x \in [0, \bar{D}]$ and $1 \leq r \leq L$. The following lemma, whose proof appears in Appendix E, establishes an upper bound on the difference in the derivatives of C_r.

Lemma 10. *Let $R_1^*, \ldots, R_L^*$ denote the optimal order-up-to levels. Let $r \in \{1, \ldots, L\}$. For each $l - r, \ldots, L$, suppose that the demand distribution D_l has a continuous density, which is bounded above by $M > 0$. Then, for any order-up-to levels, $R_r, \ldots, R_L$, we have*

$$\begin{aligned} 0 \ \leq \ & C_r'\left(R_r \middle| R_{r+1}^*, \ldots, R_L^*\right) - C_r'\left(R_r \middle| R_{r+1}, \ldots, R_L\right) \\ \leq \ & M \sum_{\ell=r+1}^{L} C_\ell\left(R_\ell \middle| R_{\ell+1}^*, \ldots, R_L^*\right) - C_\ell\left(R_\ell^* \middle| R_{\ell+1}^*, \ldots, R_L^*\right) \ . \end{aligned}$$

We now proceed with the proof of Theorem 8. Let

$$\Delta_r(T) \ = \ \frac{1}{T} \sum_{s=1}^{T} E\left[C_r(\hat{y}_{s,r} \mid R_{r+1}^*, \ldots, R_L^*) - C_r(R_r^* \mid R_{r+1}^*, \ldots, R_L^*)\right] \ ,$$

for $1 \leq r \leq L$. Note that by definition $\Delta_r(T) \geq 0$ for all r. We will provide an expression involving $\Delta_r(T)$'s such that it is both an upper bound on the left hand side expression in the statement of Theorem 8, and a lower bound on the right hand side expression.

From $C\left(\hat{y}_{s,1}, \ldots, \hat{y}_{s,L}\right) = \sum_{r=1}^{L} C_s\left(\hat{y}_{s,r} \middle| \hat{y}_{s,r+1}, \ldots, \hat{y}_{s,L}\right)$, it follows

$$\begin{aligned}
& C\left(\hat{y}_{s,1}, \ldots, \hat{y}_{s,L}\right) - C\left(R_1^*, \ldots, R_L^*\right) \\
&= \sum_{r=1}^{L} C_s\left(\hat{y}_{s,r} \middle| \hat{y}_{s,r+1}, \ldots, \hat{y}_{s,L}\right) - \sum_{r=1}^{L} C_s\left(R_r^* \middle| R_{r+1}^*, \ldots, R_L^*\right) \\
&= \sum_{r=1}^{L} \left[C_s\left(\hat{y}_{s,r} \middle| R_{r+1}^*, \ldots, R_L^*\right) - C_s\left(R_r^* \middle| R_{r+1}^*, \ldots, R_L^*\right)\right] \\
&\quad + \sum_{r=1}^{L} \left[C_s\left(\hat{y}_{s,r} \middle| \hat{y}_{s,r+1}, \ldots, \hat{y}_{s,L}\right) - C_s\left(\hat{y}_{s,r} \middle| R_{r+1}^*, \ldots, R_L^*\right)\right] \ .
\end{aligned}$$

We take expected values on both sides, and take an average over time. Consider each of the two summations separately. From the definition of $\Delta_r(T)$, it follows

$$\frac{1}{T}\sum_{s=1}^{T}\sum_{r=1}^{L} E\left[C_s\left(\hat{y}_{s,r} \middle| R_{r+1}^*, \ldots, R_L^*\right) - C_s\left(R_r^* \middle| R_{r+1}^*, \ldots, R_L^*\right)\right] \ = \ \sum_{r=1}^{L} \Delta_r(T) \ .$$

By Lemma 9,

$$\begin{aligned}
& \frac{1}{T}\sum_{s=1}^{T}\sum_{r=1}^{L} \left[C_s\left(\hat{y}_{s,r} \middle| \hat{y}_{s,r+1}, \ldots, \hat{y}_{s,L}\right) - C_s\left(\hat{y}_{s,r} \middle| R_{r+1}^*, \ldots, R_L^*\right)\right] \\
&\leq \frac{1}{T}\sum_{s=1}^{T}\sum_{r=1}^{L}\sum_{\ell=r+1}^{L} C_\ell\left(\hat{y}_{s,\ell} \middle| R_{\ell+1}^*, \ldots, R_L^*\right) - C_\ell\left(R_\ell^* \middle| R_{\ell+1}^*, \ldots, R_L^*\right) \\
&= \sum_{r=1}^{L}\sum_{\ell=r+1}^{L} \frac{1}{T}\sum_{s=1}^{T} C_\ell\left(\hat{y}_{s,\ell} \middle| R_{\ell+1}^*, \ldots, R_L^*\right) - C_\ell\left(R_\ell^* \middle| R_{\ell+1}^*, \ldots, R_L^*\right) \\
&= \sum_{r=1}^{L}\sum_{\ell=r+1}^{L} \Delta_\ell(T) \ = \ \sum_{\ell=2}^{L}\sum_{r=1}^{\ell-1} \Delta_\ell(T) \ \leq \ (L-1)\sum_{\ell=1}^{L} \Delta_\ell(T) \ .
\end{aligned}$$

Therefore,

$$\begin{aligned}
E\left[\frac{1}{T}\sum_{s=1}^{T} C\left(\hat{y}_{s,1}, \ldots, \hat{y}_{s,L}\right)\right] - C\left(R_1^*, \ldots, R_L^*\right) &= \frac{1}{T}\sum_{s=1}^{T} E\left[C\left(\hat{y}_{s,1}, \ldots, \hat{y}_{s,L}\right) - C\left(R_1^*, \ldots, R_L^*\right)\right] \\
&\leq L\sum_{r=1}^{L} \Delta_r(T) \ .
\end{aligned}$$

Now we show that $L\sum_{r=1}^{L}\Delta_r(T)$ is bounded above by the expression on the righthand side of Theorem 8. By the choice of $H_{s,r}$,

$$\begin{aligned} E\left[H_{s,r}\middle|\hat{y}_{s,r},\ldots,\hat{y}_{s,L}\right] &= C_r'\left(\hat{y}_{s,r}\middle|\hat{y}_{s,r+1},\ldots,\hat{y}_{s,L}\right) \\ &= C_r'\left(\hat{y}_{s,r}\middle|R_{r+1}^*,\ldots,R_L^*\right)+\delta_r\left(\hat{y}_{s,r},\ldots,\hat{y}_{r,L}\right), \end{aligned}$$

where we define $\delta_r\left(\hat{y}_{s,r},\ldots,\hat{y}_{s,L}\right) = C_r'\left(\hat{y}_{s,r}\middle|\hat{y}_{s,r+1},\ldots,\hat{y}_{s,L}\right) - C_r'\left(\hat{y}_{s,r}\middle|R_{r+1}^*,\ldots,R_L^*\right)$. Note that δ_r represents the bias associated with our estimate of the gradient $C_r'\left(\hat{y}_{s,r}\middle|R_{r+1}^*,\ldots,R_L^*\right)$. Also, note that $|H_{s,r}| \le L\max\{b,h\}$. We apply Lemma 2 with $\gamma=1$, $S=[0,\bar{D}]$, $\bar{B}=L\max\{b,h\}$, and obtain

$$\begin{aligned} \Delta_r(T) &\le \left(\frac{2\bar{D}L\max\{b,h\}}{\sqrt{T}}\right)+\frac{\bar{D}}{T}\sum_{s=1}^{T}E\left[\left|\delta_r\left(\hat{y}_{s,r},\ldots,\hat{y}_{s,L}\right)\right|\right] \\ &= \left(\frac{2\bar{D}L\max\{b,h\}}{\sqrt{T}}\right)+\frac{\bar{D}}{T}\sum_{s=1}^{T}E\left[C_r'\left(\hat{y}_{s,r}\middle|R_{r+1}^*,\ldots,R_L^*\right)-C_r'\left(\hat{y}_{s,r}\middle|\hat{y}_{s,r+1},\ldots,\hat{y}_{s,L}\right)\right], \end{aligned}$$

where the last equality follows since $\delta_r\left(\hat{y}_{s,r},\ldots,\hat{y}_{s,L}\right)$ is non-positive from the lower bound in Lemma 10. By the upper bound in Lemma 10,

$$\begin{aligned} \Delta_r(T) &\le \frac{2\bar{D}L\max\{b,h\}}{\sqrt{T}}+\bar{D}M\sum_{\ell=r+1}^{L}\frac{1}{T}\sum_{s=1}^{T}E\left[C_\ell\left(R_{s,\ell}\middle|R_{\ell+1}^*,\ldots,R_L^*\right)-C_\ell\left(R_\ell^*\middle|R_{\ell+1}^*,\ldots,R_L^*\right)\right] \\ &= \frac{2\bar{D}L\max\{b,h\}}{\sqrt{T}}+\bar{D}M\sum_{\ell=r+1}^{L}\Delta_\ell(T)\ . \end{aligned}$$

Thus, we obtain a recursive bound on $\Delta_r(T)$'s for each $r=1,\ldots,L$. Let

$$A = 2\bar{D}L\max\{b,h\}/\sqrt{T}\ , \qquad \text{and} \qquad B = \bar{D}M\ .$$

Then, the above bound is $\Delta_r(T) \le A + B\sum_{\ell=r+1}^{L}\Delta_\ell(T)$, where $\Delta_L(T)=A$. A simple backward induction shows

$$\sum_{\ell=r}^{L}\Delta_l(T) \le A\cdot\left[1+(B+1)+(B+1)^2+\cdots+(B+1)^{L-r}\right].$$

Therefore,

$$\begin{aligned} L\sum_{\ell=1}^{L}\Delta_\ell(T) &\le L\cdot A\cdot\left[1+(B+1)+(B+1)^2+\cdots+(B+1)^{L-1}\right] \\ &= L\cdot A\cdot\frac{(B+1)^L-1}{B} \\ &\le \frac{2L^2\max\{b,h\}(1+\bar{D}M)^L}{M}\cdot\frac{1}{\sqrt{T}}\ . \end{aligned}$$

4.4 Extension and Comments

Suppose now that the per-unit overage and underage cost in each period of the cycle depends on seasonality. Thus, the expected single-period cost is

$$Q_r(y) \quad = \quad h_r \cdot E[y - D_{1,r}]^+ + b_r \cdot E[D_{1,r} - y]^+ \ .$$

In such case, the results in this section continue to hold. In particular, Theorem 8 holds with $\max\{\max_r b_r, \ \max_r h_r\}$.

Although the average expected cost per cycle under the AIM-Seasonal algorithm converges to the optimal cost at the rate of $1/\sqrt{T}$, the convergence rate in Theorem 8 depends exponentially on the length of the cycle L. For short to moderate cycle lengths, we believe that our algorithm should perform well. Finding an alternative algorithm whose convergence rate does not depend exponentially on L, however, remains an open question.

5 Experiments

In Sections 2-4, we have introduced a suite of non-parametric algorithms that adaptively determine the inventory level in each period based on previously observed sales data. In this section, we report detailed experimental results that validate the performance of our algorithms for stationary inventory systems. For the ease of exposition, we focus our discussion on the case of perishable inventory. In Section 5.1, we briefly describe four inventory policies that we have tested, and show their performance in a representative problem. These policies are tested further in Section 5.2. We discuss the case of non-perishable inventory in Section 5.3.

5.1 Perishable Inventory: Performance

In this section, we compare the performance of four inventory policies for the stationary system with perishable inventory discussed in Section 2. These policies are listed below. Note that the first two policies assume more information than observed sales, and they are not implementable in practice unless either true demand distribution or historical uncensored demand data is available. These policies only serve as benchmarks.

- Optimal Policy: Assume that the demand distribution is available *a priori*. We set the inventory level after ordering to the newsvendor fractile of true demand in each period.
- Uncensored Demand Data: Assume that while demand distribution is not available, we observe the realization of uncensored demand in each period. We set the inventory level to the newsvendor fractile of the empirical distribution based on demand realizations from all pervious periods.
- AIM-Perishable: The AIM-Perishable algorithm is described in Section 2. We use the following step size: $\epsilon_t = \bar{D}/\left\{\max\{b,h\}\sqrt{t}\right\}$ for all t, where $\bar{D}$ denotes the maximum demand in any one period.
- CAVE: The Concave, Adaptive Value Estimation (CAVE) algorithm has been applied to the newsvendor problem. See Godfrey and Powell (2001) for details.

Between the two adaptive policies for censored demand information, only AIM-Perishable has a provable guarantee of convergence rate.

In this section, we conduct our experiment assuming that the demand distribution in each period is a discrete uniform distribution between 0 and 100. The overage and underage costs are $h = 20$ and $b = 80$, respectively. For this problem, the optimal order quantity $y^{NV} = 80$. The experiment is replicated on 200 randomly generated problem instances and the time horizon is 500 periods, unless otherwise specified.

Figure 1 compares the running average costs of the four policies mentioned above. The average for period t is taken over 200 problem instances as well as over all previous periods $\{1, \ldots, t-1\}$. At the beginning, the inventory level is initialized to 20. After 500 periods, the average cost of the AIM-Perishable algorithm is within 6% of the Optimal benchmark cost. We note that CAVE also performs quite well.

Figure 2 shows the scatter plot of the difference between the optimal cost and the running average cost of the AIM-Perishable algorithm over 5000 periods. This plot is drawn in a log-log scale. We observe a linear relationship, showing that

$$\log\left[\frac{1}{200}\sum_{j=1}^{200}\frac{1}{t}\sum_{t'=1}^{t}Q^j(y_{t'}^j) - Q(y^{NV})\right] \approx -0.5093\log(t) + 6.9908$$

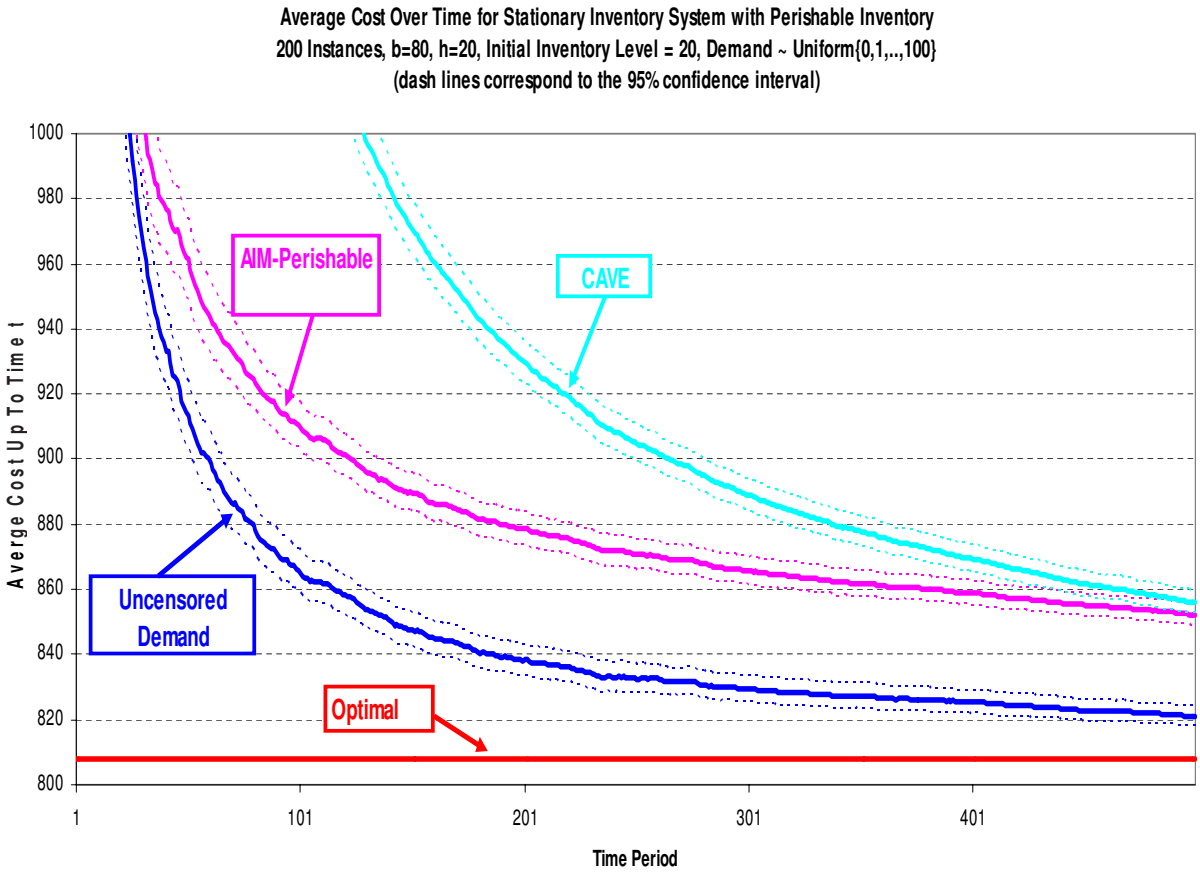


Figure 1: The average cost over time for 200 randomly generated problem instances of the inventory systems with a perishable product with $b = 80$ and $h = 20$ and initial inventory level of 20. The demand distribution is assumed to be a discrete uniform distribution over the set of integers from 0 to 100.

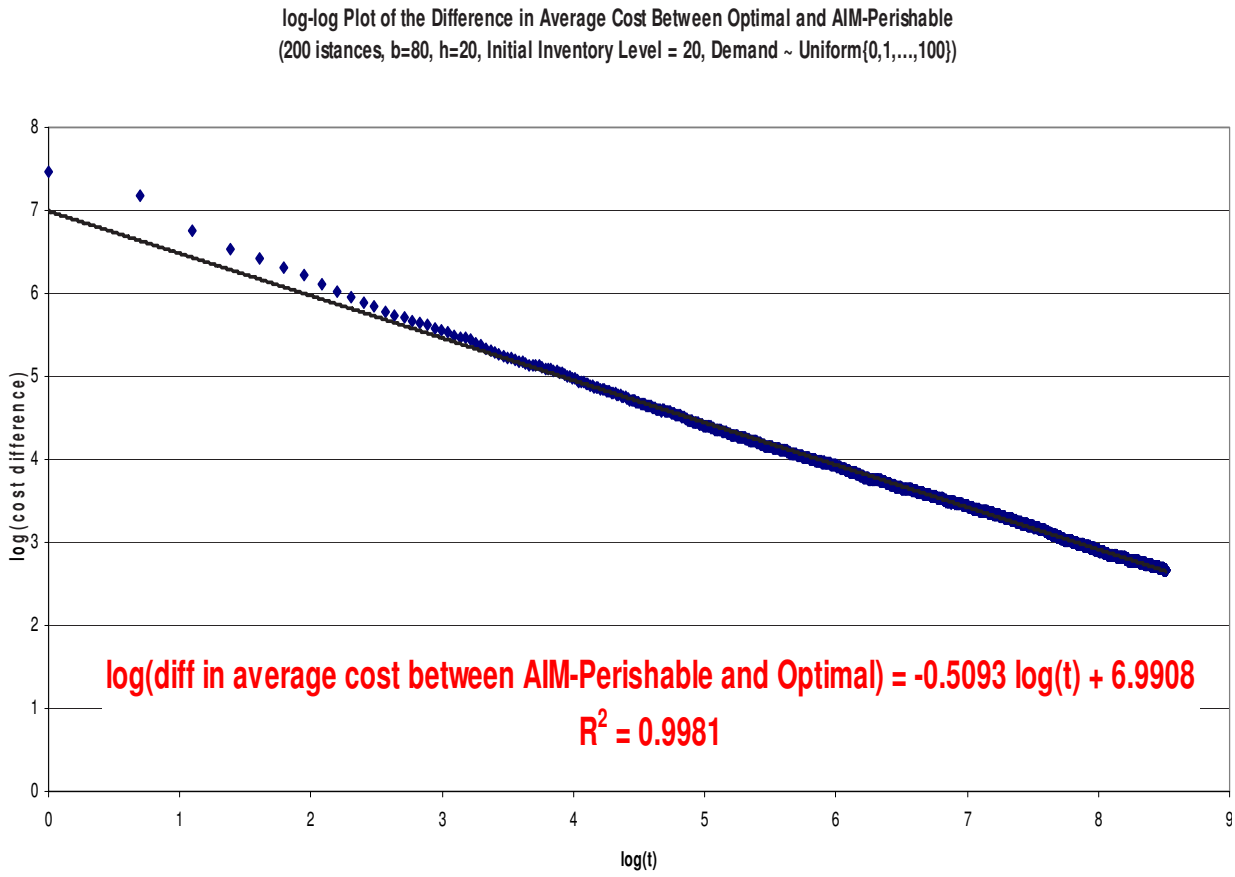


Figure 2: The log-log plot of the difference between the optimal cost and the average cost over time under the AIM-Perishable algorithm. The cost is averaged over 200 problem instances, where each problem instance has $b = 80$ and $h = 20$ and the initial inventory level is set at 20. The demand in each period is uniformly distributed over the set of integers from 0 to 100. Dash lines correspond to the 95% confidence interval.

where $j = 1, \ldots, 200$ is the index of problem instance, $y_{t'}^j$ denotes the order-up-to level generated by AIM-Perishable in period t' for the j^{th} problem instance, and $Q^j(y_{t'}^j)$ denotes the corresponding overage and underage cost. It follows

$$\frac{1}{200}\sum_{j=1}^{200}\frac{1}{t}\sum_{t'=1}^{t} Q^j(y_{t'}^j) - Q(y^{NV}) \approx \frac{\exp(6.9908)}{t^{0.5093}},$$

which is consistent with the convergence guarantee of $O(t^{-1/2})$ in Theorem 1.

5.2 Perishable Inventory: Sensitivity Analysis

In this section, we explore the how average costs of inventory policies change as we vary the initial inventory level, the demand distribution, and the overage and underage costs in each period.

Figure 3 shows the average cost of all four inventory policies during the first 100 periods for various initial starting inventory levels. For each starting inventory level, we compute the average cost during the first 100 periods over 200 problem instances. The demand in each period is a discrete uniform distribution over the set of integers from 0 to 100. We use $b = 80$ and $h = 20$. The optimal newsvendor quantity is $y^{NV} = 80$. This graph shows that the average cost of AIM-Perishable is robust against changes in initial starting inventory levels. In contrast, CAVE is quite sensitive to the initial inventory level.

Figure 4 reports the running average costs of all six inventory policies when demand and cost parameters vary. Subfigures (a) and (b) use a Gaussian demand distribution with mean 80 and standard deviation 20, while (c) and (d) use a Poisson distribution with mean 80.[1] Subfigures (a) and (c) use the same cost parameters as in Section 5.1 ($b = 80$ and $h = 20$), while (b) and (d) use the equal overage and underage costs ($b = 50$ and $h = 50$). The initial inventory level is 20 in all figures. We note that the AIM-Perishable performs well when the critical fractile $b/(b+h)$ is high – possibly because there is less loss of information due to censoring when inventory levels are high. CAVE also performs well.

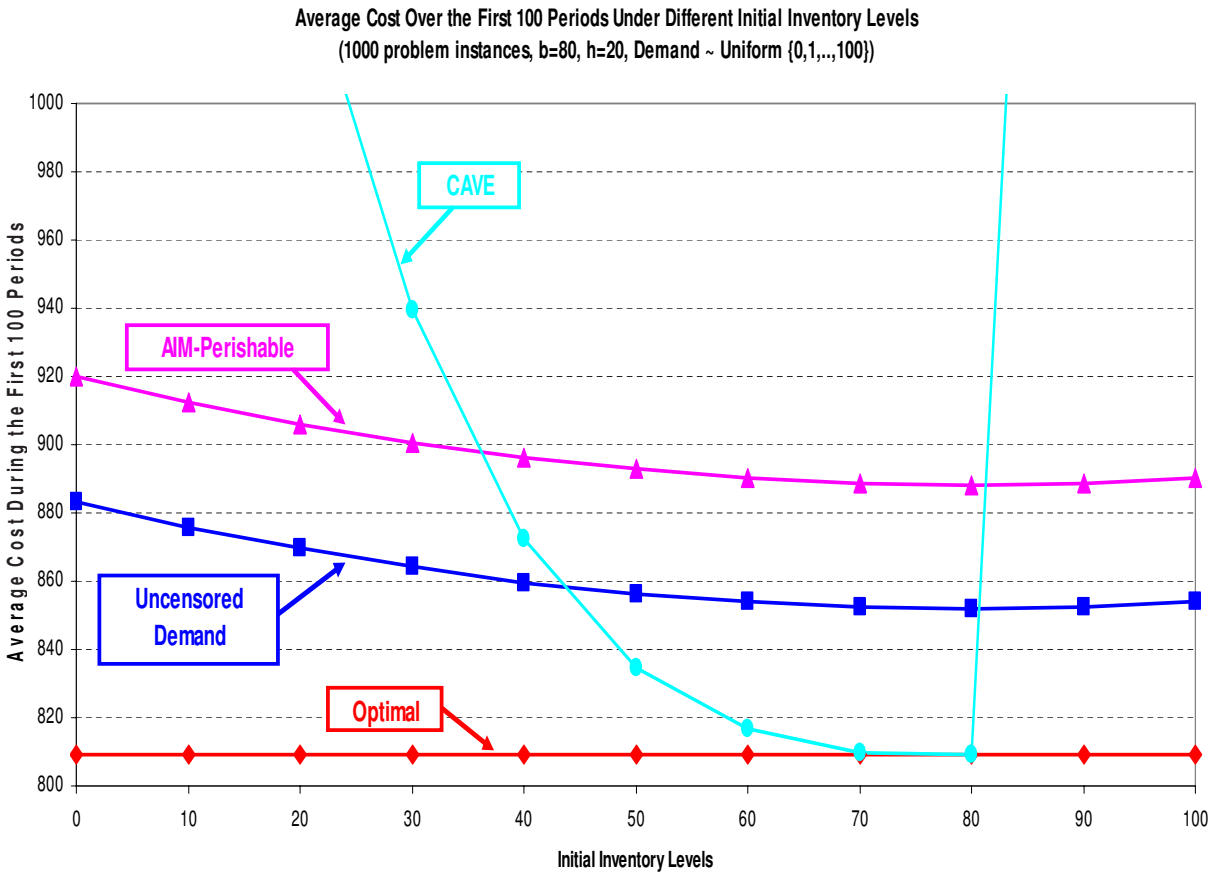
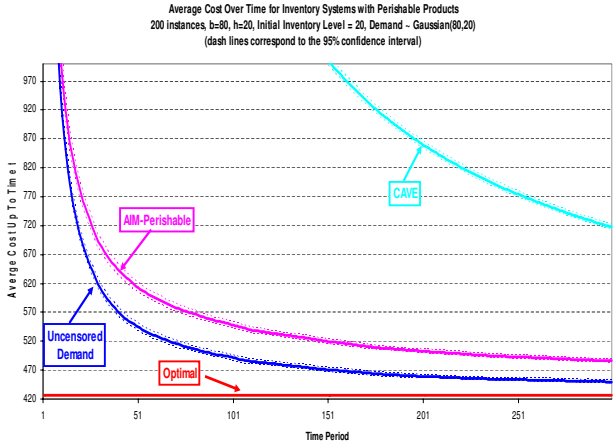
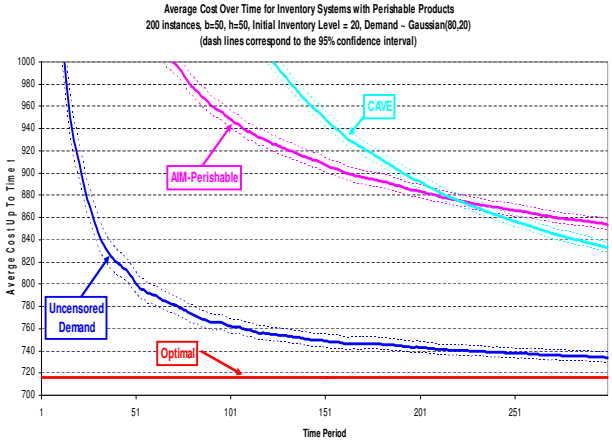


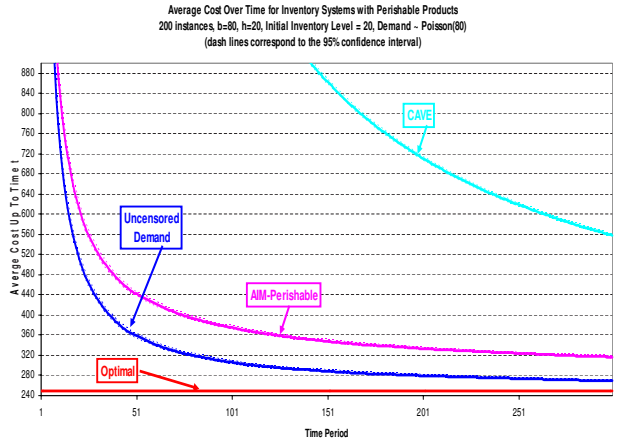
Figure 3: The average cost during the first 100 periods under various inventory policies for a set of initial starting inventory levels $\{0, 10, 20, \ldots, 100\}$. For each initial inventory level, the cost is averaged over 200 problem instances. Each problem instance has $b = 80$ and $h = 20$ and the demand in each period is uniformly distributed over the set of integers from 0 to 100.



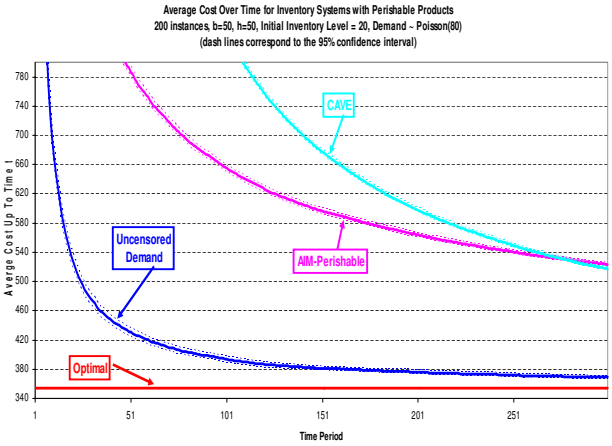
(a) $b = 80,\ h = 20,\ Normal(80, 20)$



(b) $b = 50,\ h = 50,\ Normal(80, 20)$



(c) $b = 80,\ h = 20,\ Poisson(80)$



(d) $b = 50,\ h = 50,\ Poisson(80)$

Figure 4: The running average cost under specified demand distributions and overage and underage cost parameters.

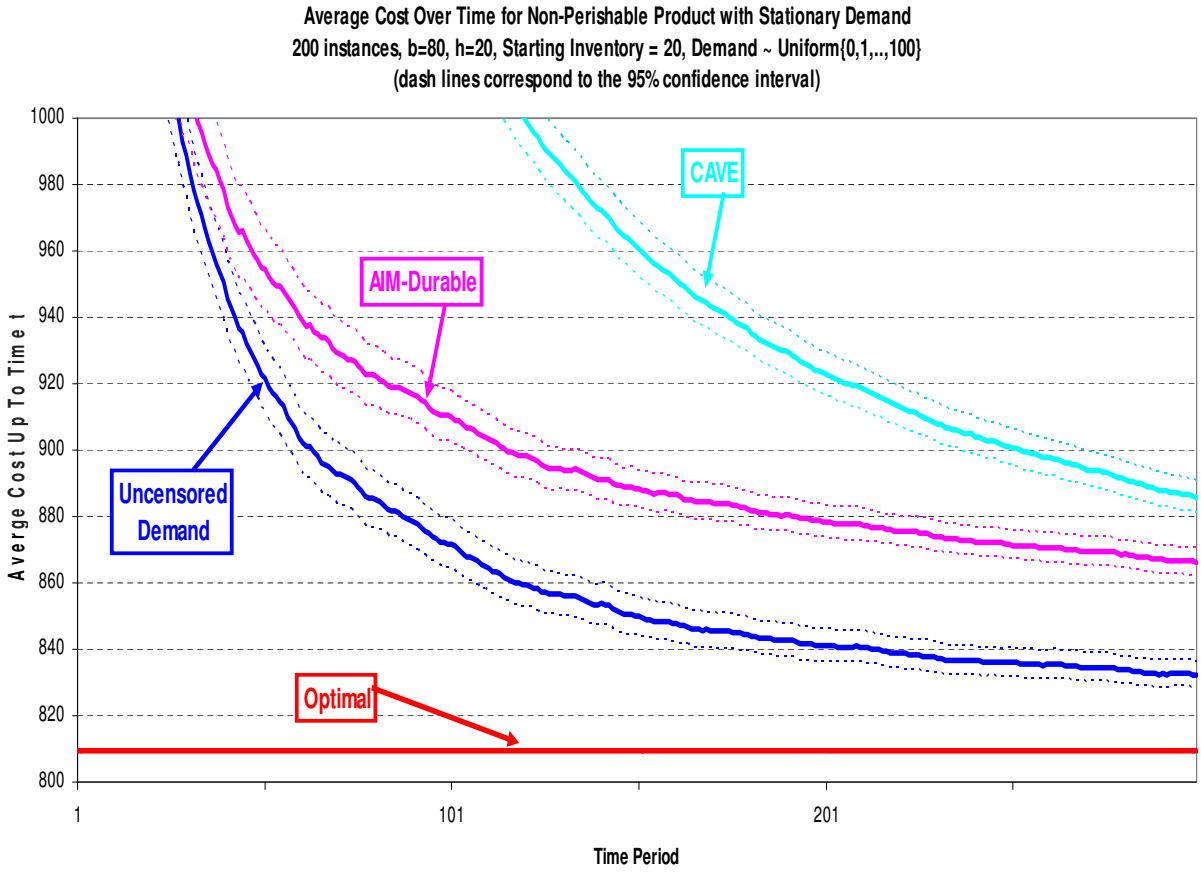


Figure 5: The average cost over 200 problem instances under the stationary demand with non-perishable products. For each instance, the demand in each period is uniformly distributed over the set of integers from 0 to 100 with $b = 80$ and $h = 20$. The initial starting level is set to 20. Dash lines correspond to the 95% confidence interval.

5.3 Non-Perishable Inventory

We have conducted additional experiments for the stationary system with non-perishable inventory, as discussed in Section 3. We compare the same set of inventory policies as in the perishable inventory case with the following clarifying explanation. The CAVE algorithm of Godfrey and Powell (2001) is defined for perishable inventory only; in the non-perishable setting, we use the output of the CAVE algorithm as the target inventory level. The AIM-Durable algorithm here refers to the algorithm in Section 3.

Figure 5 shows the average running costs of six inventory policies for non-perishable inventory. The parameters are the same as those used in Section 5.1. The outcome is very similar to the perishable case (Figure 1), and our AIM-Durable algorithm continues to perform well.

Acknowledgement

We would like to thank Jack Muckstadt for insightful suggestions and discussions on the stochastic inventory theory, and Karl Sigman and Gennady Samorodnitsky for stimulating discussions on the stochastic storage process. We also thank Retsef Levi for a careful reading of the paper and for many helpful comments and suggestions that greatly improved the presentation of the paper. We also thank Huseyin Topaloglu for sharing with us an implementation of the CAVE algorithm, which was used in the numerical computation in Section 5. To generate random variables in our experiments, we use the *Stochastic Simulation in Java* package developed by Pierre L'Ecuyer at the University of Montreal.

[1]For both of these distributions, if the realized demand exceeds 100, we round it down to 100.

A Proof of Lemma 2

We claim

$$\sum_{t=1}^{T} E\left[\Phi(w_t) - \Phi(w^*)\right] \;\leq\; \Gamma_1 + \Gamma_2 \;,$$

where

$$\begin{aligned}
\Gamma_1 &= \sum_{t=1}^{T} \left\{ \frac{E\left\|w_t - w^*\right\|^2}{2\epsilon_t} - \frac{E\left\|w_{t+1} - w^*\right\|^2}{2\epsilon_t} + \frac{\epsilon_t}{2} E\left\|H(w_t)\right\|^2 \right\}, \qquad \text{and} \\
\Gamma_2 &= diam(S) \cdot E\left[\sum_{t=1}^{T} \left\|\delta(w_t)\right\|\right] \;.
\end{aligned}$$

To prove this claim, let $\langle \cdot, \cdot \rangle$ denote the inner product in $\mathbb{R}^n$. Since the projection operation $P_S(\cdot)$ does not increase the distance between two points (i.e., non-expansive), we have for any $t \geq 1$,

$$\begin{aligned}
E\left\|w_{t+1} - w^*\right\|^2 &= E\left\|P_S\left(w_t - \epsilon_t \cdot H(w_t) - w^*\right)\right\|^2 \\
&\leq E\left\|w_t - \epsilon_t \cdot H(w_t) - w^*\right\|^2 \\
&= E\left\|(w_t - w^*) - \epsilon_t \cdot H(w_t)\right\|^2 \\
&= E\left\|w_t - w^*\right\|^2 + \epsilon_t^2 E\left\|H(w_t)\right\|^2 - 2\epsilon_t \cdot E\left[\langle H(w_t),\; w_t - w^* \rangle\right] \;.
\end{aligned}$$

By conditioning $E\left[\langle H(w_t),\; w_t - w^* \rangle\right]$ with respect to w_t and taking an expectation,

$$\begin{aligned}
E\left[\langle H(w_t),\; w_t - w^* \rangle\right] &= E\left[E\left[\langle H(w_t),\; w_t - w^* \rangle \;\middle|\; w_t\right]\right] \\
&= E\left[\langle E\left[H(w_t) \mid w_t\right],\; w_t - w^* \rangle\right] \\
&= E\left[\langle g(w_t) + \delta(w_t),\; w_t - w^* \rangle\right] \\
&= E\left[\langle g(w_t),\; w_t - w^* \rangle\right] + E\left[\langle \delta(w_t),\; w_t - w^* \rangle\right] \;,
\end{aligned}$$

where the second last equality follows from the definition of $\delta(\cdot)$ in the statement of this lemma. Therefore,

$$
\begin{aligned}
& E\left[\langle g(w_t),\ w_t - w^*\rangle\right] \\
&= E\left[\langle H(w_t),\ w_t - w^*\rangle\right] - E\left[\langle \delta(w_t),\ w_t - w^*\rangle\right] \\
&\leq \frac{E\left\|w_t - w^*\right\|^2 - E\left\|w_{t+1} - w^*\right\|^2 + \epsilon_t^2 E\left\|H(w_t)\right\|^2}{2\epsilon_t} - E\left[\langle \delta(w_t),\ w_t - w^*\rangle\right] \\
&\leq \left\{\frac{E\left\|w_t - w^*\right\|^2}{2\epsilon_t} - \frac{E\left\|w_{t+1} - w^*\right\|^2}{2\epsilon_t} + \frac{\epsilon_t}{2} E\left\|H(w_t)\right\|^2\right\} + E\left[\left\|\delta(w_t)\right\| \left\|w_t - w^*\right\|\right] \ , \\
&\leq \left\{\frac{E\left\|w_t - w^*\right\|^2}{2\epsilon_t} - \frac{E\left\|w_{t+1} - w^*\right\|^2}{2\epsilon_t} + \frac{\epsilon_t}{2} E\left\|H(w_t)\right\|^2\right\} + \left(diam(S) \cdot E\left[\left\|\delta(w_t)\right\|\right]\right) \ ,
\end{aligned}
$$

where the second inequality follows from the Cauchy-Schwartz inequality. The last inequality follows from the fact that $w^* \in S$ and $w_t \in S$ for all t. Since the subgradient $g(z)$ defines the supporting hyperplane of the convex function Φ at z, it follows that $\langle g(w_t),\ w_t - w^*\rangle$ is an upper bound on $\Phi(w_t) - \Phi(w^*)$. Therefore, by summing the above inequality over t, we complete the proof of the claim.

Now, it suffices to prove $\Gamma_1 \leq (\gamma + 1/\gamma)\ diam(S)\ \bar{B}\ \sqrt{T}$. From the definition of Γ_1,

$$
\begin{aligned}
\Gamma_1 &= \sum_{t=1}^{T} \left\{\frac{E\left\|w_t - w^*\right\|^2}{2\epsilon_t} - \frac{E\left\|w_{t+1} - w^*\right\|^2}{2\epsilon_t} + \frac{\epsilon_t}{2} E\left\|H(w_t)\right\|^2\right\} \\
&\leq \frac{E\left\|w_1 - w^*\right\|^2}{2\epsilon_1} + \frac{1}{2}\sum_{t=1}^{T}\left[\frac{1}{\epsilon_{t+1}} - \frac{1}{\epsilon_t}\right] E\left\|w_{t+1} - w^*\right\|^2 + \frac{\bar{B}^2}{2}\sum_{t=1}^{T}\epsilon_t \\
&\leq \frac{diam(S)^2}{2}\left\{\frac{1}{\epsilon_1} + \sum_{t=1}^{T}\left[\frac{1}{\epsilon_{t+1}} - \frac{1}{\epsilon_t}\right]\right\} + \frac{\bar{B}^2}{2}\sum_{t=1}^{T}\epsilon_t \\
&= \frac{diam(S)^2}{2\epsilon_{T+1}} + \frac{\bar{B}^2}{2}\sum_{t=1}^{T}\epsilon_t \\
&= \frac{diam(S)\ \bar{B}}{2\gamma}\ \sqrt{T+1}\ +\ \frac{diam(S)\ \bar{B}\ \gamma}{2}\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \ ,
\end{aligned}
$$

where the last equality follows from the definition of ϵ_t. Now, from the fact that $\sqrt{T+1} \leq 2\sqrt{T}$ and $\sum_{t=1}^{T} 1/\sqrt{t} \leq \int_0^T t^{-1/2} dt = 2\sqrt{T}$, we obtain the desired result.

B Proof of Lemma 4

Let $w^* = \arg\min\{\Phi(w) \mid w \in [0, \bar{S}]\}$. For each $z \in [0, \bar{S}]$, let $\phi(z)$ be a subgradient of Φ at z, such that $\phi(z) = E\left[H(z) \mid z\right]$ when z is an integer. We let $\tilde{\phi}$ be a piecewise-linear approximation of ϕ defined by

$$\tilde{\phi}(z) = (z - \lfloor z \rfloor) \cdot \phi\left(\lceil z \rceil\right) + (1 - z + \lfloor z \rfloor) \cdot \phi\left(\lfloor z \rfloor\right) \ .$$

In period t, the algorithm produces fractional $\hat{w}_t$ and integral $\bar{w}_t$, from which the random variable $H(\bar{w}_t)$ is constructed. It follows from the definition of $\bar{w}_t$ that $E[H(\bar{w}_t)|\hat{w}_t] = \tilde{\phi}(\hat{w}_t)$.

We let $\tilde{\Phi}$ be an anti-derivative of $\tilde{\phi}$, i.e., $\tilde{\Phi}'(z) = \tilde{\phi}(z)$ for any z. Since $\tilde{\phi}$ is a nondecreasing function, $\tilde{\Phi}$ is convex. Let $w^\circ \in [0, \bar{S}]$ be a minimizer of $\tilde{\Phi}$. Since $E[H(\bar{w}_t)|\hat{w}_t] = \tilde{\phi}(\hat{w}_t)$ holds, the sequence of points $\hat{w}_1, \hat{w}_2, \ldots$ may be viewed as an output of the algorithm in Lemma 2 for minimizing a convex function $\tilde{\Phi}$. Therefore,

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(\tilde{\Phi}(\hat{w}_t) - \tilde{\Phi}(w^\circ))\right] \ \leq \ \frac{2\bar{S}\bar{B}}{\sqrt{T}} \ .$$

Suppose the following claim holds:

$$\max\{\Phi(\lceil w \rceil) - \Phi(w^*),\ \Phi(\lfloor w \rfloor) - \Phi(w^*)\} \ \leq \ 2\left(\tilde{\Phi}(w) - \tilde{\Phi}(w^\circ)\right) + 2\bar{B} \ .$$

Then, since $\bar{w}_t$ is either $\lceil \hat{w}_t \rceil$ or $\lfloor \hat{w}_t \rfloor$, it follows

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T}(\Phi(\bar{w}_t) - \Phi(w^*))\right] \ &\leq \ E\left[\frac{1}{T}\sum_{t=1}^{T}\max\{\Phi(\lceil \hat{w}_t \rceil) - \Phi(w^*),\ \Phi(\lfloor \hat{w}_t \rfloor) - \Phi(w^*)\}\right] \\ &\leq \ 2 \cdot E\left[\frac{1}{T}\sum_{t=1}^{T}\left(\tilde{\Phi}(\hat{w}_t) - \tilde{\Phi}(w^\circ)\right)\right] + 2\bar{B} \ \leq \ \frac{4\bar{S}\bar{B}}{\sqrt{T}} + 2\bar{B} \ , \end{aligned}$$

proving the statement of Lemma 4.

Now it remains to prove the above claim. Consider the case where $\lfloor w^* \rfloor \leq w \leq \lceil w^* \rceil$. Then, $|\lceil w \rceil - w^*| \leq 1$ and $|\lfloor w \rfloor - w^*| \leq 1$. From $|\phi| \leq \bar{B}$, we have $\Phi(\lceil w \rceil) - \Phi(w^*) \leq \bar{B}$ and $\Phi(\lfloor w \rfloor) - \Phi(w^*) \leq \bar{B}$, which imply the above claim.

In the case of $w \geq \lceil w^* \rceil$, a similar argument implies that both $\Phi(\lceil w^* \rceil) - \Phi(w^*)$ and $\Phi(\lceil w \rceil) - \Phi(\lfloor w \rfloor)$ are bounded above by $\bar{B}$. We introduce another function $\Psi : [0, \bar{S}] \to \Re$ defined as an

anti-derivative of ψ, where

$$\psi(w) \;=\; \begin{cases} \phi(\lceil w \rceil), & \text{if } w > w^*; \\ 0, & \text{if } w = w^*; \\ \phi(\lfloor w \rfloor), & \text{if } w < w^*. \end{cases}$$

From the convexity of Φ, Ψ is also convex with the same minimizer w^*. Since Φ is continuously differentiable except at finitely many points, we have

$$\Phi(w) - \Phi(w^*) \;=\; \int_{z=w^*}^{w} \phi(z) \;\leq\; \int_{z=w^*}^{w} \psi(z) \;=\; \Psi(w) - \Psi(w^*) \;.$$

Furthermore, for any integer $i \geq \lceil w^* \rceil$, since $\phi(i)$ is nonnegative, we have

$$\phi(i+1) \;=\; 2 \cdot \frac{\phi(i+1)}{2} \;\leq\; 2 \cdot \frac{\phi(i+1) + \phi(i)}{2} \;=\; 2 \int_{i}^{i+1} \tilde{\phi}(\xi) d\xi$$

where the last inequality follows from the fact that $\tilde{\phi}$ is linear between i and $i+1$ with $\tilde{\phi}(i) = \phi(i)$ and $\tilde{\phi}(i+1) = \phi(i+1)$. Therefore,

$$\begin{aligned} \Psi(\lfloor w \rfloor) - \Psi(\lceil w^* \rceil) \;&=\; \int_{\lceil w^* \rceil}^{\lfloor w \rfloor} \phi(\lceil \xi \rceil) d\xi \;=\; \sum_{i=\lceil w^* \rceil}^{\lfloor w \rfloor - 1} \phi(i+1) \\ &\leq\; 2 \sum_{i=\lceil w^* \rceil}^{\lfloor w \rfloor - 1} \int_{i}^{i+1} \tilde{\phi}(\xi) d\xi \;=\; 2 \int_{\lceil w^* \rceil}^{\lfloor w \rfloor} \tilde{\phi}(\xi) d\xi \;=\; 2[\tilde{\Phi}(\lfloor w \rfloor) - \tilde{\Phi}(\lceil w^* \rceil)] \\ &\leq\; 2[\tilde{\Phi}(w) - \tilde{\Phi}(w^\circ)] \;, \end{aligned}$$

where the last inequality follows because w° is between $\lceil w^* \rceil$ and $\lfloor w^* \rfloor$, and is a minimizer of a convex function $\tilde{\Phi}$(w), which is nondecreasing for $w \geq \lceil w^* \rceil$. In summary, the claim follows from

$$\begin{aligned} &\max\{\Phi(\lceil w \rceil) - \Phi(w^*),\; \Phi(\lfloor w \rfloor) - \Phi(w^*)\} \\ &=\; \Phi\left(\lceil w \rceil\right) - \Phi\left(w^*\right) \\ &\leq\; \Psi(\lceil w \rceil) - \Psi(w^*) \\ &=\; [\Psi(\lceil w^* \rceil) - \Psi(w^*)] \;+\; [\Psi(\lfloor w \rfloor) - \Psi(\lceil w^* \rceil)] \;+\; [\Psi(\lceil w \rceil) - \Psi(\lfloor w \rfloor)] \\ &\leq\; \bar{B} + 2[\tilde{\Phi}(w) - \tilde{\Phi}(w^\circ)] + \bar{B} \;. \end{aligned}$$

We prove the case of $w \leq \lfloor w^* \rfloor$ similarly.

C Proof of Lemma 6

The proof of Lemma 6 makes use of Lemma 11. In a bounded random walk with a drift, hitting the boundary defines a regeneration. For any given time period t, the length of the renewal interval containing t is bounded above in the following result.

Lemma 11. *Let $D_1, D_2, \ldots$ be independent and identically distributed nonnegative random variables such that for all t, $D_t \leq \bar{D}$ with probability one and $E[D_t] > 1$. Consider a random walk $(W_t \mid t \geq 1)$ defined by*

$$W_{t+1} \quad = \quad [W_t + 1 - D_t]^+$$

with an initial condition $W_0 = 0$. Define random variables τ_i by $\tau_i = \inf\{t > \tau_{i-1} \mid W_t = 0\}$ for each $i \geq 1$, where $\tau_0 = 0$. Let $J_i = \{s : \tau_{i-1} < s \leq \tau_i\}$. For each $t \geq 1$, let $i(t)$ denote i such that J_i contains t. Then,

$$E\left[\left|J_{i(t)}\right|\right] \quad \leq \quad \frac{2\alpha}{(1-\alpha)^2}$$

holds where $\alpha = \exp\left\{-2 \cdot \left(E[D_1] - 1\right)^2 / \bar{D}\right\}$.

Proof. Since the collection of J_i's are disjoint and partition the natural numbers, $i(\cdot)$ is well-defined. We first claim the following result:

$$E[\left|J_{i(t)}\right|] \quad \leq \quad E[|J_1|^2] \quad = \quad \sum_{r=1}^{\infty} r^2 \mathcal{P}\{|J_1| = r\} \ ,$$

for each $t \geq 1$. The intuition behind this claim is that a longer interval is more likely to contain a given period t than a shorter interval.

To prove the above claim, consider the following recursive equation defined by conditioning on the time of the first renewal:

$$\begin{aligned} E[\left|J_{i(t)}\right|] \quad &= \quad \sum_{s=1}^{t-1} E\left[\left|J_{i(t)}\right| \cdot \mathbf{1}\left[|J_1| = s\right]\right] + E\left[|J_1| \cdot \mathbf{1}\left[|J_1| \geq t\right]\right] \\ &= \quad \sum_{s=1}^{t-1} E\left[\left|J_{i(t)}\right| \cdot \mathbf{1}\left[|J_1| = s\right]\right] + E\left[|J_1| \cdot \mathbf{1}\left[|J_1| \geq t\right]\right] \\ &= \quad \sum_{s=1}^{t-1} E[|J_{i(t-s)}|] \cdot \mathcal{P}\{|J_1| = s\} + E\left[|J_1| \cdot \mathbf{1}\left[|J_1| \geq t\right]\right] \ , \end{aligned}$$

where the last equality follows from the observation that $(W_t : t \geq 1)$ is a renewal process with a regeneration point at 0. It follows that for all $t \geq 1$

$$E[|J_{i(t)}|] \;\leq\; \max_{1\leq s\leq t-1} E[|J_{i(t-s)}|] + E\left[|J_1| \cdot \mathbf{1}\left[|J_1| \geq t\right]\right] \;.$$

By iteratively applying the above recursion, we have

$$\begin{aligned} E[|J_{i(t)}|] \;&\leq\; \sum_{s=1}^{t} E\left[|J_1| \cdot \mathbf{1}\left[|J_1| \geq s\right]\right] \;=\; \sum_{s=1}^{t}\sum_{r=s}^{\infty} r\mathcal{P}\left\{|J_1| = r\right\} \\ &\leq\; \sum_{s=1}^{\infty}\sum_{r=s}^{\infty} r\mathcal{P}\left\{|J_1| = r\right\} \;=\; \sum_{r=1}^{\infty}\sum_{s=1}^{r} r\mathcal{P}\left\{|J_1| = r\right\} \;=\; \sum_{r=1}^{\infty} r^2\mathcal{P}\left\{|J_1| = r\right\} \;, \end{aligned}$$

completing the proof of the claim.

Now, from the above claim,

$$\begin{aligned} E[|J_{i(t)}|] \;&\leq\; \sum_{r=1}^{\infty} r^2\mathcal{P}\left\{|J_1| = r\right\} \;\leq\; \sum_{r=1}^{\infty}\left(2\sum_{\ell=1}^{r}\ell\right)\mathcal{P}\left\{|J_1| = r\right\} \\ &=\; 2\sum_{\ell=1}^{\infty}\ell\sum_{\ell=r}^{\infty}\mathcal{P}\left\{|J_1| = r\right\} \;=\; 2\sum_{\ell=1}^{\infty}\ell\cdot\mathcal{P}\left\{|J_1| \geq \ell\right\} \;. \end{aligned}$$

We need to establish an upper bound on $\mathcal{P}\left\{|J_1| \geq \ell\right\}$. The event $|J_1| \geq \ell$ occurs if and only if the cumulative sum $\sum_{s=1}^{\ell'}(1 - D_s)$ up to each $\ell' = 1, \ldots, \ell$ remains positive. Thus,

$$\mathcal{P}\left\{|J_1| \geq \ell\right\} \;\leq\; \mathcal{P}\left\{\sum_{s=1}^{\ell}(1 - D_s) \geq 0\right\} \;=\; \mathcal{P}\left\{\sum_{s=1}^{\ell}(E[D_s] - D_s) \geq \ell\cdot(E[D_1] - 1)\right\} \;.$$

We give an upper bound on this probability by using the following inequality due to Hoeffding (1963). For a sequence $(U_s \mid s \geq 1)$ of independent random variables with mean 0 and $a_s \leq U_s \leq b_s$ for each s,

$$\mathcal{P}\left\{\sum_{s=1}^{\ell} U_s \geq \eta\right\} \;\leq\; \exp\left\{\frac{-2\eta^2}{\sum_{s=1}^{\ell}(b_s - a_s)^2}\right\}$$

holds for any $\ell \geq 1$ and $\eta > 0$. Since $(E[D_s] - D_s \mid s \geq 1)$ is a sequence of identical and independently distributed random variables with mean 0, and its support is contained in the interval of length $\bar{D}$, it follows

$$\mathcal{P}\left\{\sum_{s=1}^{\ell}(E[D_s] - D_s) \geq \ell\cdot(E[D_1] - 1)\right\} \;\leq\; \exp\left\{\frac{-2\cdot(\ell\cdot(E[D_1] - 1))^2}{\ell\cdot\bar{D}}\right\} \;=\; \alpha^{\ell} \;.$$

Therefore, it follows

$$E[|J_{i(t)}|] \;\leq\; 2\sum_{\ell=1}^{\infty} \ell \cdot \mathcal{P}\{|J_1| \geq \ell\} \;\leq\; 2\sum_{\ell=1}^{\infty} \ell\alpha^{\ell} \;=\; \frac{2\alpha}{(1-\alpha)^2} \;,$$

and we complete the proof of Lemma 11. □

Consider the auxiliary process $(W_t \mid t \geq 1)$ defined in the statement of Lemma 11. Note that the sample path of W_t is a renewal process, with regeneration point at 0. Let τ_i', J_i, and $i(t)$ be defined as in Lemma 11. Clearly, $Z_t \leq W_t$ holds almost surely for each $t \geq 0$.

For each sample path of Z_t, we introduce another stochastic process $(V_t \mid t \geq 1)$, where V_t corresponds to the cumulative inflow in the renewal cycle containing t, without accounting for its outflow. For each $t \geq 1$, define

$$V_t \;=\; \sum_{t'=1}^{t} \frac{1}{\sqrt{t'}} \cdot \mathbf{1}\left[t' \in J_{i(t)}\right] \;=\; \sum_{t'=\tau_{i(t)-1}+1}^{t} \frac{1}{\sqrt{t'}} \;.$$

We claim that for all t, $Z_t \leq V_t$. This result follows from the fact that when $Z_t > 0$, then $\tau_{i-1} < t < \tau_i$ for some i. Since $W_{\tau_{i-1}} = 0$, we have $Z_{\tau_{i-1}} = 0$, and therefore

$$Z_t \leq \sum_{\tau_{i-1} < t' \leq t} \frac{1}{\sqrt{t'}} = V_t.$$

Thus, for the proof of Lemma 6, it suffices to establish

$$\sum_{t=1}^{T} E\left[V_t\right] \;\leq\; \frac{4\alpha\sqrt{T}}{(1-\alpha)^2} \;.$$

For any fixed T,

$$\begin{aligned}\sum_{t=1}^{T} V_t \;&=\; \sum_{t=1}^{T}\sum_{s=1}^{t} \frac{1}{\sqrt{s}} \cdot \mathbf{1}\left[s \in J_{i(t)}\right] \;\leq\; \sum_{t=1}^{T}\sum_{s=1}^{T} \frac{1}{\sqrt{s}} \cdot \mathbf{1}\left[s \in J_{i(t)}\right] \\ &=\; \sum_{s=1}^{T} \frac{1}{\sqrt{s}} \sum_{t=1}^{T} \mathbf{1}\left[s \in J_{i(t)}\right] \;\leq\; \sum_{s=1}^{T} \frac{1}{\sqrt{s}} \left|J_{i(s)}\right| \;.\end{aligned}$$

Therefore,

$$E\left[\sum_{t=1}^{T} V_t\right] \;\leq\; \sum_{s=1}^{T} \frac{1}{\sqrt{s}}\, E\left[|J_{i(s)}|\right] \;\leq\; \sum_{s=1}^{T} \frac{1}{\sqrt{s}} \cdot \frac{2\alpha}{(1-\alpha)^2} \;\leq\; \frac{4\alpha\sqrt{T}}{(1-\alpha)^2} \;,$$

where the second inequality follows from Lemma 11, and the final inequality follows from the fact $\sum_{t=1}^{T} 1/\sqrt{t} \leq 2\sqrt{T}$ (as in the proof of Lemma 2).

D Proof of Lemma 9

To prove Lemma 9, we first establish the following results.

Proposition 12. *Let w^* be a minimizer of any convex function $f : \Re \to \Re$. For any $w, \bar{w} \in \Re$, we have $0 \leq f(\max\{\bar{w}, w\}) - f(\max\{\bar{w}, w^*\}) \leq f(w) - f(w^*)$.*

Proof. Consider all possible orderings among w^*, w and $\bar{w}$. Use the fact that $f(w)$ is weakly decreasing if $w < w^*$, and weakly increasing if $w > w^*$. □

Proposition 13. *Let $R_1^*, \ldots, R_L^*$ denote the optimal order-up-to levels. For any $r = 1, \ldots, L$, and order-up-to levels $R_r, \ldots, R_L$, we have, for each $l \in \{r, \ldots, L\}$,*

$$\begin{aligned} 0 &\leq U_r\left(R_r \mid R_{r+1}, \ldots, R_{l-1}, R_l, R_{l+1}^*, \ldots, R_L^*\right) - U_r\left(R_r \mid R_{r+1}, \ldots, R_{l-1}, R_l^*, R_{l+1}^*, \ldots, R_L^*\right) \\ &\leq C_l\left(R_l \middle| R_{l+1}^*, R_{l+2}^*, \ldots, R_L^*\right) - C_l\left(R_l^* \middle| R_{l+1}^*, R_{l+2}^*, \ldots, R_L^*\right) \ . \end{aligned}$$

Proof. We compare two sets of order-up-to policies from period r to L defined by the following target inventory levels: $\left(R_r, \ldots, R_{l-1}, R_l, R_{l+1}^*, \ldots, R_L^*\right)$, and $\left(R_r, \ldots, R_{l-1}, R_l^*, R_{l+1}^*, \ldots, R_L^*\right)$. Since the order-up-to levels are the same for all periods prior to l, the costs incurred by both policies during these periods are the same. Let Z be the random variable representing the inventory level at the beginning of period l, assuming $y_r = R_r$, i.e.,

$$Z = \max_{i \in \{r, \ldots, l-1\}} \left\{R_i - D_{[i,l-1]}\right\}.$$

From the definition of U_r,

$$\begin{aligned} & U_r\left(R_r \mid R_{r+1}, \ldots, R_l, R_{l+1}^*, \ldots, R_L^*\right) - U_r\left(R_r \mid R_{r+1}, \ldots, R_{l-1}, R_l^*, \ldots, R_L^*\right) \\ &= E_Z\left[U_l\left(\max\{Z, R_l\} \mid R_{l+1}^*, \ldots, R_L^*\right) - U_l\left(\max\{Z, R_l^*\} \mid R_{l+1}^*, \ldots, R_L^*\right)\right] \\ &= E_Z\left[C_l\left(\max\{Z, R_l\} \mid R_{l+1}^*, \ldots, R_L^*\right) - C_l\left(\max\{Z, R_l^*\} \mid R_{l+1}^*, \ldots, R_L^*\right)\right] \ , \end{aligned}$$

where the second equality follows from the definition of C_l. Now, since $C_l\left(\cdot \mid R_{l+1}^*, \ldots, R_L^*\right)$ is a convex function which achieves its minimum at R_l^* (by Lemma 7), Proposition 12 implies

$$\begin{aligned} 0 &\leq C_l\left(\max\{z, R_l\} \mid R_{l+1}^*, \ldots, R_L^*\right) - C_l\left(\max\{z, R_l^*\} \mid R_l^*, \ldots, R_L^*\right) \\ &\leq C_l\left(R_l \mid R_{l+1}^*, \ldots, R_L^*\right) - C_l\left(R_l^* \mid R_l^*, \ldots, R_L^*\right) \end{aligned}$$

for any realized value of z. By taking an expectation with respect to Z on both sides, we obtain the required result. □

Here is the proof of Lemma 9. If $r = L$, the lemma is trivially true. We proceed by assuming $r < L$. We express $C_r\left(R_r \middle| R_{r+1}, \ldots, R_L\right) - C_r\left(R_r \middle| R^*_{r+1}, \ldots, R^*_L\right)$ as the following telescoping sum

$$C_r\left(R_r \middle| R_{r+1}, \ldots, R_L\right) - C_r\left(R_r \middle| R^*_{r+1}, \ldots, R^*_L\right) \;=\; \sum_{l=r+1}^{L} \tilde{\Delta}_{r,l} \;,$$

where each $\tilde{\Delta}_{r,l}$, for $l = r+1, \ldots, L$, is defined by

$$\tilde{\Delta}_{r,l} \;=\; C_r\left(R_r \middle| R_{r+1}, \ldots, R_{l-1}, R_l, R^*_{l+1}, \ldots, R^*_L\right) - C_r\left(R_r \middle| R_{r+1}, \ldots, R_{l-1}, R^*_l, R^*_{l+1}, \ldots, R^*_L\right) \;.$$

From the definition of C_r, it follows

$$\begin{aligned}\tilde{\Delta}_{r,l} \;=\; & \left\{U_r\left(R_{r+1}, \ldots, R_{l-1}, R_l, R^*_{l+1}, \ldots, R^*_L\right) - U_r\left(R_r, \ldots, R_{l-1}, R^*_l, R^*_{l+1}, \ldots, R^*_L\right)\right\} \\ & - \left\{U_{r+1}\left(R_{r+1}, \ldots, R_{l-1}, R_l, R^*_{l+1}, \ldots, R^*_L\right) - U_{r+1}\left(R_{r+1}, \ldots, R_{l-1}, R^*_l, R^*_{l+1}, \ldots, R^*_L\right)\right\} \;.\end{aligned}$$

Applying Proposition 13 to U_r, the difference in the first set of braces is bounded above by $C_l\left(R_l \middle| R^*_{l+1}, R^*_{l+2}, \ldots, R^*_L\right) - C_l\left(R^*_l \middle| R^*_{l+1}, R^*_{l+2}, \ldots, R^*_L\right)$. Applying Proposition 13 to U_{r+1}, the difference in the second set of braces is nonnegative. Thus, for any $r \leq l$, we have

$$\tilde{\Delta}_{r,l} \leq C_l\left(R_l \middle| R^*_{l+1}, R^*_{l+2}, \ldots, R^*_L\right) - C_l\left(R^*_l \middle| R^*_{l+1}, R^*_{l+2}, \ldots, R^*_L\right) ,$$

which implies that

$$\begin{aligned}& C_r\left(R_r \middle| R_{r+1}, \ldots, R_L\right) - C_r\left(R_r \middle| R^*_{r+1}, \ldots, R^*_L\right) \\ & \quad = \sum_{l=r+1}^{L} \tilde{\Delta}_{r,l} \leq \sum_{l=r+1}^{L} C_l\left(R_l \middle| R^*_{l+1}, R^*_{l+2}, \ldots, R^*_L\right) - C_l\left(R^*_l \middle| R^*_{l+1}, R^*_{l+2}, \ldots, R^*_L\right) ,\end{aligned}$$

which is the desired result.

E Proof of Lemma 10

The result of Lemma 10 holds trivially if $r = L$. Using backward induction, we suppose the result holds for $r+1$, and prove the result for r. From the definition of C_r and U_r,

$$\begin{aligned} C'_r(R_r \mid R_{r+1}, \ldots, R_L) &= U'_r(R_r \mid R_{r+1}, \ldots, R_L) \\ &= Q'_r(R_r) + E_{D_r}\left[U'_{r+1}(R_r - D_r \mid R_{r+2}, \ldots, R_L) \cdot \mathbf{1}(R_r - D_r > R_{r+1})\right] \\ &= Q'_r(R_r) + E_{D_r}\left[C'_{r+1}(R_r - D_r \mid R_{r+2}, \ldots, R_L) \cdot \mathbf{1}(R_r - D_r > R_{r+1})\right] \\ &= Q'_r(R_r) + \int_{R_{r+1}}^{\bar{D}} C'_{r+1}(w \mid R_{r+2}, \ldots, R_L) f_W(w) dw \ , \end{aligned}$$

where we let $W = R_r - D_r$ and $f_W(\cdot)$ denotes the probability density of W. A similar expression holds for $C'_r(R_r \mid R^*_{r+1}, \ldots, R^*_L)$.

We consider the case of $R^*_{r+1} \leq R_{r+1}$. (The case of $R^*_{r+1} > R_{r+1}$ can be argued similarly.) Then,

$$\begin{aligned} & C'_r(R_r \mid R^*_{r+1}, \ldots, R^*_L) - C'_r(R_r \mid R_{r+1}, \ldots, R_L) \\ &= \int_{R^*_{r+1}}^{R_{r+1}} C'_{r+1}(w \mid R^*_{r+2}, \ldots, R^*_L) f_W(w) dw \\ &\quad + \int_{R_{r+1}}^{\bar{D}} \left\{C'_{r+1}(w \mid R^*_{r+2}, \ldots, R^*_L) - C'_{r+1}(w \mid R_{r+2}, \ldots, R_L)\right\} f_W(w) dw \ . \end{aligned}$$

Since $C_{r+1}(\cdot \mid R^*_{r+2}, \ldots, R^*_L)$ is convex with the minimum at R^*_{r+1} (by Lemma 7), the first integral in the equation above is nonnegative, and is bounded above by

$$M \cdot \left\{C_{r+1}(R_{r+1} \mid R^*_{r+2}, \ldots, R^*_L) - C_{r+1}(R^*_{r+1} \mid R^*_{r+2}, \ldots, R^*_L)\right\} \ .$$

For the second integral, we apply the induction hypothesis to the integrand. Thus, the integral is nonnegative, and is bounded above by

$$\begin{aligned} & \int_{R_{r+1}}^{\bar{D}} \left\{ M \sum_{r'=r+2}^{L} \left\{C_{r'}(R_{r'} \mid R^*_{r'+1}, \ldots, R^*_L) - C_{r'}(R_{r'} \mid R^*_{r'+1}, \ldots, R^*_L)\right\} \right\} f_W(w) dw \\ &= \left\{ M \sum_{r'=r+2}^{L} \left\{C_{r'}(R_{r'} \mid R^*_{r'+1}, \ldots, R^*_L) - C_{r'}(R_{r'} \mid R^*_{r'+1}, \ldots, R^*_L)\right\} \right\} \int_{R_{r+1}}^{\bar{D}} f_W(w) dw \\ &\leq M \sum_{r'=r+2}^{L} \left\{C_{r'}(R_{r'} \mid R^*_{r'+1}, \ldots, R^*_L) - C_{r'}(R_{r'} \mid R^*_{r'+1}, \ldots, R^*_L)\right\} \ . \end{aligned}$$

Therefore, we complete the induction step for r.

References

Auer, P., N. Cesa-Bianchi, and P. Fisher. 2002. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning* 47:235–256.

Azoury, K. S. 1985. Bayes Solution to Dynamic Inventory Models Under Unknown Demand Distribution. *Management Science* 31 (9): 1150–1160.

Bookbinder, J. H., and A. E. Lordahl. 1989. Estimation of Inventory Reorder Level Using the Bootstrap Statistical Procedure. *IEE Transactions* 21:302–312.

Braden, D. J., and M. Freimer. 1991. Information Dynamics of Censored Observations. *Management Science* 37:1390–1404.

Burnetas, A. N., and C. E. Smith. 2000. Adaptive Ordering and Pricing For Perishable Products. *Operations Research* 48 (3): 436–443.

Chang, H. S., M. C. Fu, J. Hu, and S. I. Marcus. 2005. An Adaptive Sampling Algorithm for Solving Markov Decision Processes. *Operations Research* 53 (1): 126–139.

Chang, S. H., and D. E. Fyffe. 1971. Estimation of Forecast Errors for Seasonal Style-Goods Sales. *Management Science* 18 (2): B89–B96.

Chen, F. 2000. Optimal Policies for Multi-Echelon Inventory Problems with Batch Ordering. *Operations Research* 48 (3): 376–389.

Chen, L., and E. Plambeck. 2005. Dynamic Inventory Management with Learning about Demand Distribution and Substitution Probability. *Working Paper*.

Chu, L. Y., J. G. Shanthikumar, and Z.-J. M. Shen. 2005. Solving Operational Statistics via a Bayesian Analysis. *Working Paper*.

Conrad, S. A. 1976. Sales Data and the Estimation of Demand. *Operations Research Quarterly* 27 (1): 123–127.

Ding, X., M. L. Puterman, and A. Bisi. 2002. The Censored Newsvendor and the Optimal Acquisition of Information. *Operations Research* 50 (3): 517–527.

Flaxman, A. D., A. T. Kalai, and H. B. McMahan. 2004. Online Convex Optimization In the Bandit Setting: Gradient Descent Without a Gradient. *Working Paper*.

Gallego, G., and I. Moon. 1993. The Distribution Free Newboy Problem: Review and Extensions. *Journal of the Operations Research Society* 44 (8): 825–834.

Gallego, G., and L. B. Toktay. 2004. All-or-Nothing Ordering Under a Capacity Constraint. *Operations Research* 52 (6): 1001–1002.

Godfrey, G. A., and W. B. Powell. 2001. An Adaptive, Distribution-Free Algorithm for the Newsvendor Problem with Censored Demands, with Applications to Inventory and Distribution. *Management Science* 47:1101–1112.

Harpaz, G., W. Y. Lee, and R. L. Winkler. 1982. Optimal Output Decisions of a Competitive Firm. *Management Science* 28:589–602.

Hoeffding, W. 1963. Probability Inequalities for Sums of Bounded Random Variables. *Journal of American Statistical Association* 58:13–30.

Iglehart, D., and S. Karlin. 1962. Optimal Policy for Dynamic INventory Process with Nonstationary Stochastic Demands. In *Studies in Applied Probability and Management Science*, ed. K. Arrow, S. Karlin, and H. Scarf. Stanford University Press.

Iglehart, D. L. 1964. The Dynamic Inventory Problem with Unknown Demand Distribution. *Management Science* 10 (3): 429–440.

Jagannathan, R. 1977. Minimax Procedure for a Class of Linear Programs Under Uncertainty. *Operations Research* 25 (1): 173–177.

Janakiraman, G., and J. A. Muckstadt. 2004. Inventory Control in Directed Networks: A Note on Linear Costs. *Operations Research* 52 (3): 491–495.

Karlin, S. 1960. Dynamic Inventory Policy with Varying Stochastic Demands. *Management Science* 6 (3): 231–258.

Karlin, S., and H. Scarf. 1958. Inventory Models of the Arrow-Harris-Marschak Type with Time Lag. In *Studies in the Mathematical Theory of Inventory and Production*, ed. K. Arrow, S. Karlin, and H. Scarf. Stanford University Press.

Kleinberg, R. 2004. Nearly Tight Bounds for the Continuum-Armed Bandit Problem. *Advances in Neural Information Processing Systems*.

Lai, T., and H. Robbins. 1985. Asymptotically Efficient Adaptive Allocation Rules. *Advances in Applied Mathematics* 6:4–22.

Lariviere, M. A., and E. L. Porteus. 1999. Stalking Information: Bayesian Inventory Management with Unobserved Lost Sales. *Management Science* 45 (3): 346–363.

Levi, R., R. Roundy, and D. B. Shmoys. 2005. Computing Provably Near-Optimal Sample-Based Policies for Stochastic Inventory Control Models. *Working Paper*.

Liyanage, L. H., and J. G. Shanthikumar. 2005. A Practical Inventory Control Policy Using Operational Statistics. *Operations Research Letters* 33:341–348.

Lovejoy, W. S. 1990. Myopic Policies for Some Inventory Models with Uncertain Demand Distributions. *Management Science* 36 (6): 724–738.

Lu, X., J.-S. Song, and K. Zhu. 2004. Dynamic Inventory Planning for Perishable Products with Censored Demand Data. *Working Paper*.

Lu, X., J.-S. Song, and K. Zhu. 2005. Inventory Control with Unobservable Lost Sales and Bayesian Updates. *Working Paper*.

Murray, G. R., and E. A. Silver. 1966. A Bayesian Analysis of the Style Goods Inventory Problem. *Management Science* 12 (11): 785–797.

Nahmias, S. 1994. Demand Estimation in Lost Sales Inventory Systems. *Naval Research Logistics* 41:739–757.

Perakis, G., and G. Roels. 2005. Regret in the newsvendor model with partial information. *Operations Research*. Forthcoming.

Powell, W., A. Ruszczynski, and H. Topaloglu. 2004. Learning Algorithms for Separable Approximations of Discrete Stochastic Optimization Problems. *Mathematics of Operations Research* 29 (4): 814–836.

Scarf, H. 1958. A Min-Max Solution of an Inventory Problem. In *Studies in the Mathematical Theory of Inventory and Production*, ed. k. Arrow, S. Karlin, and H. Scarf, 201–209. Stanford University Press.

Scarf, H. 1960. Some Remarks on Bayes Solutions to the Inventory Problem. *Naval Research Logistics Quarterly* 7:591–596.

Scarf, H. E. 1959. Bayes Solution to the Statistical Inventory Problem. *Annals of Mathematical Statistics* 30 (2): 490–508.

Song, J.-S., and P. Zipkin. 1993. Inventory Control in a Fluctuating Demand Environment. *Operations Research* 41:351–370.

Veinott, A. 1965a. The Optimal Inventory Policy for Batch Ordering. *Operations Research* 13:424–432.

Veinott, A. 1965b. Optimal Policy for a Multi-Product, Dynamic, Nonstationary Inventory Problem. *Management Science* 12:206–222.

Veinott, A., and H. Wagner. 1965. Computing Optimal (s, S) Inventory Policies. *Management Science* 1 (5): 525–552.

Zinkevich, M. 2003. Online Convex Programming and Generalizaed Infinitesimal Gradient Ascent. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*. Washington, DC.